Converting Game Theoretical 2x2 Games To Their N-Person Counterparts

Daan Wanningen

Supervised by:

dr. Gerard A. W. Vreeswijk

April 5, 2018

Abstract

In this thesis five archetypical *2x2* games and their linearly converted *N*-person variants are discussed. These games are: the Prisoner's Dilemma; Chicken; Battle of the Sexes; Hero; and Stag hunt. First, the scenarios of these games and their variations are elaborated upon to serve as an illustration. Second, the underlying social phenomena and the strategic pressures reflecting these have been discussed for 2x2 games. It is here argued the games fall into one of three dilemmas: (1) social dilemma; (2) coordination dilemma; (3) trust dilemma. Furthermore, the choice is made to linearly convert these *2x2* games to their *N*-person counterparts. In doing so, functions are adapted to allow for a relatively trivial conversion from the *2x2* games to their *N*-person variants. As such, the strategic pressures for *N*-person counterparts are discussed and compared to their *2x2* variants in light of their respective underlying social phenomena. Consecutively, the essential properties of both the *2x2* games and their *N*-person counterparts are identified with respect to the strategic pressures they represent. In turn, this work forms an abstract framework which may serve as a foundation for an implementation within an interactive environment for educational or experimental purposes. As such, the considerations to bear in mind with regard to the games themselves, their conversion, and their properties may be used to allow for meaningful analysis in multi-agent settings.

Keywords: Game theory; N-person Games; Abstract framework

Contents

1. Introduction

The broad field of game theory has been around for decades, the first *2x2* game being introduced by von Neumann and Morgenstern (1953). Since then the field has enjoyed copious amounts of research from various approaches and with useful results for several other fields (*e.g. economics*: Scarf, 1967; *sociology*: Dawes, 1980; *computer science (CS)*: Papadimitriou & Roughgarden, 2004; *atificial intelligence (AI)*: Parsons & Wooldridge, 2002; and *biology*: Hammerstein & Selten, 1994). While game theory has brought forth many interesting games, the main focus of the field has been to describe and evaluate the games themselves. As such, abstract frameworks and taxonomies for such games have been defined for the class of *2x2* games in general, and inherent properties of those games have been linked to their respective scenarios. This serves as the foundation of the uses in other fields, since it carefully defines specific properties (*e.g.* the availability of a dominant strategy) a game may or may not have, that give rise to the interesting pressures illustrating a certain scenario (*e.g.* individual versus group gain in social dilemmas). Furthermore, research regarding *N*-person games has a wide range of focus, including – but not limited to – different types of games, the inclusion of costs and benefits within games, the formation of coalitions within games, different game formats, and individual behaviours in games. However, frameworks regarding the conversion of the classical *2x2* games to their respective *N*-person variants have been scarcely researched.

While the research regarding *N*-person games has yielded interesting results, a proper foundation as represented by a framework is still lacking. Such a framework would allow for meaningful comparison between experiments regarding similar games and it forms a good basis for current and future research. For example, the usefulness of coalitions could be examined in light of different game types, based on the underlying mechanisms in these types. Moreover, research regarding the choices in different game formats may be argued for through the underlying mechanisms. Furthermore, such a framework may be useful in other settings. For example the internet, where *N*-person game theoretical scenarios are commonplace (Shoham, 2008). There are abundant exemplar applications on the internet where multiple entities interact that may have differing information and interests (*e.g.* games of routing, networking-formation games, and peer-to-peer networks) (Shoham, 2008). This calls for analysis and design of systems where such interactions are well-defined. Game theory is by far the most developed theory for this and a framework – as proposed here – defines such a system that allows for meaningful analysis. It may be no surprise then, that game theory features routinely in the conferences and journals of AI, electronic commerce, and in networking and other areas of CS (Shoham, 2008).

The conversion of *2x2* games to multiple ($N > 2$) players is less straightforward then it may seem and deserves intricate attention. In literature, *N*-person variants of certain classical *2x2* games are proposed that have different properties than their underlying *2x2* games. This is of course inevitable to some extent since the addition of multiple players requires some adaptations at the very least, but it is important to recognise these differences and define them precisely. In addition, while some games are introduced under the same name, they may differ significantly because of individually chosen adaptations. This is not a problem in itself but it should be well-documented and it may be beneficial for future research to subdivide games such that each is consistently used within literature to allow for meaningful comparison. Moreover, some *N*-person games are named differently while their underlying *2x2* game is similar, which may cause confusion.

For example the *2x2* Snowdrift (SD) game is often referred to as a game with the same payoff matrix (and as such, the same underlying mechanism) as the game of Chicken. Although the payoffs for the SD game are modelled as costs and benefits rather than predetermined static payoffs (as is classically modelled in Chicken), these costs and benefits are constrained such that the resulting *2x2* payoff preferences are the same (Santos, Pinheiro, Santos & Pacheco, 2012). However, when converting to an *N*-person scenario the SD game often uses a certain threshold below which the lowest payoff is granted for every player, whereas *N*-person Chicken is often modelled as a continuous linear transformation of the underlying *2x2* game (Meux, 1973). A linear transformation of a *2x2* game refers to a transformation of the payoffs, where the cooperation and defection payoffs are defined as linear functions of the number of cooperators or number of players choosing the same strategy. The same problem arises for the *N*-PD, where it is sometimes transformed linearly and sometimes a threshold is used. This problem is represented by the class of Public Goods games. This class refers to a common characteristic of *N*-person games, the sharing of either costs or benefits or both, representing the sharing of scarce goods. The payoffs in such games are thus reflected by costs and benefits. As such, usually either cooperation or defection payoffs are modelled as a linear function of the amount of cooperation, but often a certain threshold is used below which no good is produced and nobody receives any payoff (*e.g.* Pacheco, Santos, Souza & Skyrms, 2009; Ledyard, 1997). However, literature remains unclear whether to model PGGs with the existence of a threshold depicting the success or failure of cooperation, or whether either the costs or the benefits or both are a function of the amount of cooperation. In addition, when both the cooperation and defection payoffs of the underlying *2x2* game are indeed linearly transformed, differences among the starting points of the function and choice made regarding the steepness of the slope remain inconsistent in literature and are sometimes chosen arbitrarily. These inconsistencies may not be problematic for the papers in question, when the constraints are well-defined and their implications are understood. However, it is problematic and confusing when an attempt is made to compare different researches and their implications in a meaningful manner.

Furthermore, it seems like the classical approach to define *N*-person games has remained scarcely researched. The research that has been conducted regarding a linear conversion of the classical *2x2* games and their inherent properties, has – to the knowledge of this author – only been done regarding a specific game, the Prisoner's Dilemma (Hamburger, 1973). However, such research is exactly what allows for the (consistent) frameworks and their implementations relevant for the fields of AI and CS, and more specifically multi-agent learning (MAL) (Shoham, 2008). When multiple agents learn concurrently one cannot distinguish between learning and teaching, and the question of "optimal" learning as well as the more general notion of an "optimal policy" are no longer well defined. As such, the very questions of single-agent learning change when moving to multi-agent learning. To support this change of approach, a well-defined framework that allows for meaningful analysis of different kinds of abstract scenarios is certainly desired.

Here it is proposed to linearly convert some classical *2x2* games to their *N*-person variants and that the properties, constraints and interactions between them may serve as an abstract framework and a foundation for an implementation and future work. The games used to achieve this are the five archetypical games: the Prisoner's Dilemma (PD); the game of Chicken; the Battle of the Sexes game (BoS); the Hero game; and finally, the Stag Hunt dilemma (SH). These games are referred to as archetypical because of the distinct pressures each applies to the so-called natural (*i.e.* minimax) outcome. Within the abstract framework (a) the underlying social phenomena of these games are defined, and (b) the properties and constraints necessary to ascertain them for these games are discussed. The aim of this effort is to gain a deeper understanding regarding the games themselves, and the effects and necessary requirements when converting *2x2* games to their respective *N*-person variants in general. In turn, these insights may be used to explicate the considerations to bear in mind when analysing game theoretic scenarios and the strategies relevant in them, for both human and non-human agents.

In this thesis three social phenomena are described that underlie the five archetypical games used: (1) the social dilemma; (2) the coordination dilemma; and (3) the trust dilemma (*e.g.* Rapoport, 1967; Dawes, 1980; Liebrand, 1983; Snidal, 1985; Zhao, Szilagy & Szidarovszky, 2008). Furthermore, some relatively simple functions are defined that allow for the linear conversion of the archetypical games in such a manner that the strategic pressures representing the underlying social phenomena still closely resemble their *2x2* counterparts. Moreover, these dilemmas and the games that depict them are defined in accordance to the properties as brought forth by Hamburger (1973). As such, certain properties are shown to hold for both the *2x2* versions and their *N*-person variants, whereas others do not hold. In addition, properties are identified that ensure the essential characteristics of the games in question and their underlying social phenomena. The linear transformations allow the games to be illustrated by graphs, which are shown to be a straightforward manner of modelling the games, such that the underlying social phenomena and their related properties may be clearly explained. As such, the abstract framework posed in this thesis may serve as a foundation for an implementation of these archetypical games and their *N*-person variants. In turn, this implementation may then be used in educational or experimental settings to explain the archetypical games, their underlying social phenomena and properties, and the effects of an *N*-person conversion thereof, in a straightforward and clear manner. In addition, this framework may be used as fundamental argumentation to devise strategies to be used in these games by either human or computer agents. As such, the framework may also be used to illustrate the dynamic facets of multi-agent learning.

In the next chapter the scenarios of each of the *2x2* games elaborated upon in this thesis will be briefly discussed. In *Chapter 3* an in-depth explanation of the underlying social phenomena as brought forth by the strategic pressures the *2x2* games impose on an individual will follow. In *Chapter 4* this will be done with regard to the *N*-person conversion of the respective *2x2* games. In *Chapter 5* the inherent properties of the games will be defined and discussed in the light of their respective social phenomena. In the final chapter an example manner of implementation will be given for clarification, the concepts introduced in chapters *3* and *4* and their implications will be summarised, and their relevance and relation to other work discussed.

2. Interesting games

This chapter serves as an illustration of the games that will be used in chapters 4 and 5 to explore their underlying pressures and respective properties, in both the *2x2* and *N*-person form. The games that will be discussed are the PD, the game of Chicken, the BoS game, the Hero game, and the SH game. The classic scenarios of the *2x2* games will be elaborated upon, after which some variants of their respective *N*-person scenarios are discussed.

2.1. Prisoner's Dilemma

The PD is perhaps the most well-known, researched and accepted game in the field of game theory. The scenario of the *2x2* variant is classically depicted as described in Luce and Raiffa (1957):

> *"[A] district attorney who lacks sufficient evidence for conviction separates two suspects and informs each that he has two alternatives: to confess to the crime the police are sure the pair has committed or not to confess. If neither confesses, the prisoners are informed that they will receive minor punishment. If both confess, a sentence less than the most severe one will be recommended for both. If one confesses and the other does not, the one who confesses will get off almost scot-free, but the other will receive the most severe sentence possible." (Scodel and Minas, 1960, pp. 133)*

It is noteworthy that although the PD is an oversimplified scenario, it is in fact a description of a situation seen often in everyday life. Specifically, a person must choose between self-interest, risking a worse overall outcome, but yielding the highest individual potential outcome, or to opt 'for the greater good' maximising the overall outcome, but without the potential of having a higher individual benefit. This is represented by the four outcomes for each individual playing this game, the outcomes where: both cooperate (maximising the social welfare) represented by a; one cooperates and one defects (where the cooperator is the sucker and the defector is tempted) represented by b and c respectively, and; both defect (minimising social welfare) represented by d. The payoff preference ordering for an individual being represented here by $c > a > d > b$. As such, its *N*-person variant should reflect this dilemma between self- and group interest.

The *N*-PD is often modelled as a linearly transformed variant of its underlying *2x2* game, with both the defection and the cooperation payoffs as a linear function of the number of cooperators. In addition, the slopes of the functions have the same steepness, *i.e.* not only are they linear, they are parallel to each other. In other words, the payoffs for an individual in the *N*-PD are calculated by the multiplication of the *2x2* outcomes by the increment in payoffs for each respective strategy. As such, cooperators are receiving $b + (a - b) \times n$ and defectors are receiving $d + (c - d) \times n$. This ensures that an individual always (*i.e.* at every amount of cooperation) has the same strategic pressures as represented by the payoff preferences.

The (unscrupulous) Diner's Dilemma is a variant of the *N*-person PD. In this variant the same strategic pressures as in the normal *N*-PD are applied, but the payoffs are modelled as costs and benefits an individual may have. The situation is exemplified by the following scenario: Several individuals go out to dine, with the prior agreement that the bill is split equally. The diners then each have two choices; (1) to order the expensive dish (e.g. lobster), or (2) the inexpensive dish (e.g. hamburger) (O'Donovan, Jones, Marusich, Teng, Gonzalez & Höllerer, 2013). The assumption here being that the expensive dish yields more value for a given individual than the inexpensive dish and that everyone orders simultaneously (making sure a given individual doesn't know the others' strategies). The individual valuation is then calculated through dividing the value of the individual dish by the overall cost of everyone. This variant is most similar to the original PD when several other constraints are met on top of the two mentioned above. First, after taking into account the costs of each dish, the value of the expensive dish is lower than the inexpensive dish. Second, a given individual prefers the expensive meal *only* when sufficient others prefer the inexpensive meal, while the benefit of the expensive meal is relatively greater, which in turn results in a relatively greater payoff obtained. This is possible because of the relative reduction in costs when more others order the inexpensive meal, since benefits are individual and costs are averaged out across all. In other words, the marginal benefit exceeds the marginal cost of the consumption taken (Gneezy, Haruvy & Yafe, 2004). These constraints make sure that the individual outcome for both strategies are a positive function of the amount of 'cooperation' (i.e. ordering the inexpensive dish) and that ordering the expensive dish yields a relatively better outcome than ordering the inexpensive dish for any amount of cooperation. The function for cooperators then becomes $a - \frac{C}{N}$, while for defectors it becomes $b - \frac{C}{N}$, where a and b represent the benefits of the inexpensive and expensive dishes respectively, C the total costs of the dinner and N the total number of diners.

As such, whenever a single round is played, there is a single Nash Equilibrium where everyone orders the expensive dish (and is worse off than if everyone had opted for the inexpensive dish) – which is the same representation as the *2x2* game. However, whenever more rounds are played cooperation may emerge to some extent, because of the balance between the costs and benefits as modulated by the number of players opting for one dish or the other (O'Donovan et al., 2013).

2.2. Chicken

There is a class of public goods games that is currently inconsistently defined within literature. Within this class there are several games that are used interchangeably; they claim to have the same underlying mechanism, but are modelled differently per research, which is not necessarily bound to the names of these games (*e.g.* Meux, 1973; Taylor & Ward, 1982; Bornstein, Budescu & Zamir, 1997; Santos et al., 2012). These are the following games: the Chicken game; the Snowdrift (SD) game; and the Hawk-Dove (HD) game.

The *2*-player Chicken game is portrayed by the following scenario. Two drivers race toward each other and are on a collision course. Both drivers have two options, either to dodge or to continue driving in the same fashion – which will lead to a collision if neither attempts to dodge. As such, there are three possible outcomes: (1) one of the two drivers dodges; (2) both dodge, resulting in 'loss of face' for the one(s) that dodged (and the other winning the dare), but where neither of them is harmed; or (3) neither dodge, resulting in no 'loss of face' for either but a catastrophic accident. It is trivial to see that although cases (1) and (2) may yield lower value for either driver (or higher for the winning driver), the third case is the lowest value.

The SD game is illustrated by the following scenario: "When a snowdrift traps several individuals, work is required to shovel off the snow and let everyone get home, an effort which is shared among those who participate" (Santos, et al., 2012). In this scenario, a threshold may be used that needs to be achieved to avoid a disastrous outcome, after which collective costs may be divided among cooperators but not defectors, and benefits among both. In this case, a single cooperator may be able to avoid the outcome on its own, but at great cost, whereas the cost may have been divided between several cooperators. In the *2*-person case, this yields the same payoff structure as the game of Chicken (Doebeli & Hauert, 2005). This is relatively straightforward to see when the benefit of getting home is defined as b and the cost of shovelling as c, where $b > b - c > 0$. Each of the individuals who cooperate and shovel then get $b - \frac{c}{n}$ where n is the number of cooperators. If both defect, no one gets anything, yielding a payoff of 0. If one cooperates and the other defects, the cooperator gets $b - \frac{c}{n}$ while the defectors get b. Now the structure is $b > b - \frac{c}{2} > b - \frac{c}{1} > 0$, for the *2*-person game, which is the same structure as the game of Chicken (Souza, Pacheco & Santos, 2009).

Furthermore, the HD game was conceived by Smith (1980) as a way to explain an animal contest problem in the fields of biology and evolutionary game theory. It depicts the following scenario:

> *"It is a contest over a shareable resource. The contestants can be either a Hawk or a Dove. These are not two separate species of bird; they are two subtypes of one species with two different types of strategy (two different morphs). […] The strategy of the Hawk (a fighter strategy) is to first display aggression, then escalate into a fight until he either wins or is injured. The strategy of the Dove (fight avoider) is to first display aggression but if faced with major escalation by an opponent to run for safety. If not faced with this level of escalation the Dove will attempt to share the resource."*

Clearly, it depicts a similar scenario as both the Chicken and the SD games. However, it is modelled in the exact same manner as the SD game mentioned above – resulting in equivalence, with costs and benefits that yield payoffs in the same structure as Chicken (Smith, 1980). As such, the HD game will not further be discussed. Rather, the differences between the *N*-person variants of Chicken and SD are discussed.

The *N*-person variant of Chicken may be modelled in such a manner that there is not a single disastrous outcome, but rather a continuum of outcomes (from disastrous to advantageous) – depending on the relative number of cooperators and defectors. As such, the same strategic pressures are applied to all players at any amount of cooperation, albeit diluted by the total number of players involved in playing the game (Hamburger, 1973). This is perhaps strategically closest to resembling its *2*-person variant – since the strategic pressures applied to each individual remain similar – and is usually referred to as *N*-person Chicken. However, the game of Chicken as represented by its original *2*-person scenario is no longer apt, since crashing into one another while others steer away would still intuitively yield a worse outcome for that particular individual than steering away themselves, whereas it is now modelled as the better option. Instead, it models a scenario like the *N*-PD, but where the value of defecting increases at a faster pace than the value of cooperating, as cooperation increases. This is to model the fact that defecting weighs heavier than cooperating, as implied by the 'disastrous outcome'. Nevertheless, this does still reflect the abstract idea behind Chicken. Each individual of a group prefers a certain strategy individually, resulting in disaster as more individuals attempt to enforce their own preferred strategy, whereas there is 'room' for such individuals when others 'give in' to those strategies.

The contrast with what is usually called the *N*-person variant of the Snowdrift scenario now becomes clear. Recall that in the SD game collective costs may be divided among cooperators but not defectors, and benefits are divided among both. This results in a static pressure to defect and an increasing pressure to cooperate, as cooperation increases – while defection remains relatively more profitable than cooperation, at any amount of cooperation (if there is any). While this applies the same pressures as the *2*-person game of Chicken, it is now dependant on a threshold of a single cooperator, after which the pressure for defectors becomes static. As such, the different amounts of cooperation apply different strategic pressures. Furthermore, consider that the payoff for defecting at any amount of cooperation (other than $n = 0$) is a better outcome than the cooperation payoff, while the defection payoff does not increase as cooperation increases. In comparison with the linearly transformed *N*-person Chicken, the *N*-person SD game somewhat simplifies the scenario.

To conclude, while both Chicken and Snowdrift games model the same dilemma in a *2*-person scenario, the *N*-person conversion of both scenarios result in different games. While either scenario may be used as an analogy for conflicts involving bilateral threats, the *N*-person Chicken variant intuitively more closely resembles a real-world generalisation. In for example military or political confrontations – that may be disastrous to all parties involved – it is more likely that it is a continuous slope toward disaster when one or more parties fail to cooperate, as opposed to requiring only a single cooperative party to avoid disaster (Bornstein, et al., 1997).

2.3. Battle of the Sexes

The BoS game is intuitively different from the PD and Chicken games, since the pressure here is not linked to the amount of cooperation. It is usually depicted by the following scenario:

> *"A man and his wife want to spend an evening together, so they try to choose an entertainment of common interest. The husband (Player 1) prefers a football game, while the wife (Player 2) prefers a ballet performance. However, both of them prefer going out together to going to either entertainment alone." (Zhao, et. al, 2008, pp. 3670)*

As may be observed in the story, the main pressure is to coordinate the strategies such that each individual has a different strategy than the other. As such, labelling the strategies of each individual as cooperation or defection is no longer intuitively correct. Rather, it is an arbitrary strategy that yields a preferred outcome depending on others' strategies. For the sake of consistency the strategies will still be labelled as such throughout this paper, but it is noteworthy that it is indeed not intuitive considering the scenario. In fact, the cooperative strategy resembles the preferred strategy for an individual player (e.g. going to the football game as a husband), whereas the defective strategy resembles giving in to the other's preferred strategy (e.g. going to the ballet as a husband). In addition, while the classical BoS game may not model any difference in preferences between uncoordinated behaviour (i.e. the husband going to a different place than the wife and vice versa), this is modelled in the strictly ordinal variant discussed here. As such, when both opt for their own preferred strategy (i.e. the husband going to football and the wife going to ballet), it results in a better outcome than if both give in to the other's preference (i.e. husband going to the ballet and wife going to football). This is of course an intuitive manner of modelling, even though it could be argued that the outcome where both defect is nonsensical in this scenario.

The *N*-person variant of this game can still be linearly transformed, and it is often modelled as such. However, there are still some differences between individual research papers. For example Szilagyi and Somogyi (2008) transform the game linearly such that one strategy increases linearly and the other decreases linearly as a function of the amount of cooperation. This has the advantage of being modelled in the same manner as the PD and Chicken games, but it is no longer intuitively correct. Whereas the highest and second-highest payoffs are at the outcome where the spread is even between cooperators and defectors (the 'coordination point') in the *2x2* game, the highest payoff for one strategy is at the least amount of the other strategy and vice versa for this *N*-person variant.

Another example stems from Zhao et al. (2008), and models the payoffs as functions of individual players choosing the same strategy (where the payoff decreases as this amount increases). They opt to model individuals as having a preference for either one of the strategies and define cooperation and defection as adhering to one's individual preference or not. However, this complicates the game by having to define two distinct types of players, with each still having two strategies available, resulting in four effective outcome pairs. As will be discussed in the next chapters, it can also be modelled as inherent property – under the assumption that each player is the same and prefers one strategy over the other when close to the coordination point, and vice versa when farther away from the coordination point. As such, it still resembles the coordination dilemma and the preference for one strategy or another, but it is no longer linked to an individual player. This may be a slight deviation from the original scenario, but still applies the underlying pressures (but with a different preference relation over the strategies rather than over the players).

2.4. Hero

The Hero game is the only game out of the five games discussed in this thesis that has not enjoyed intricate attention. This is perhaps due to the similarity to the BoS game. It does deserve attention however, since the underlying pressure is relatable to real-world scenarios and distinct from the BoS game nonetheless. The game of Hero does not have a classical scenario in literature, but its *2x2* variant could be depicted as follows:

Two neighbours want to go to a bar to drink beers or to a club to dance. However, since they resent each other, each only enjoys going to either the bar or the club when the other is not present. Furthermore, since both prefer drinking to dancing, going to the bar is always relatively more appreciated than going to the club.

It should be clear from this example that the main pressure is to coordinate their strategies such that they are not in the same place, at the same time. The other pressure is going to the bar, since it always is relatively better than going to the club (but is still worse to be at the bar together than being at the club alone). As such, it is still a coordination dilemma (similar to BoS), but with one strategy being dominant over another (similar to the PD). The hero of the story is the individual that choses to go to the club, granting the other the best, and him- or herself the second best outcome.

The *N*-person conversion may be done in a similar fashion as the BoS game, where both defection and cooperation are parallel linear functions of the number of players choosing the same strategy.

2.5. Stag Hunt

The Stag Hunt Dilemma is a story that became a game. The story told by Rousseau in *A Discourse On Inequality*, is as follows:

"If it was a matter of hunting deer, everyone well realized that he must remain faithful to his post; but if a hare happened to pass within reach of one of them, we cannot doubt that he would have gone off in pursuit of it without scruple." (Skyrms, 2004, pp. 1)

As opposed to the *N*-PD, in this scenario there is no conflict between individual and mutual benefit, there is however, conflict between the latter and individual risk. Stag hunt is about cooperation, in the sense that both options – cooperation and defection, or stag and hare respectively - are Nash Equilibria in the classical *2*-person version (Skyrms, 2004). However, as is perhaps intuitive, to hunt a stag successfully is more rewarding, but both hunters need to participate, whereas a hare may be successfully hunted alone, but is less rewarding. The idea is that of a social contract, where the choice lies between two goods, but the better option requires cooperation (Pacheco, et al., 2009). As such, there are three different payoffs: successful stag hunt s; successful hare hunt h; and unsuccessful stag hunt c. The corresponding payoff ranking then is $s > h > c$. Since hare hunts do not require cooperation it is often modelled such that it yields the same value, regardless of whether another 'helps' with the hunt (*i.e.* $s > h_1 = h_2 > c$, where h_1 and h_2 represent hunting for hare alone or together, respectively). However, it may also be modelled such that the value of hare is more rewarding when hunting alone than when hunting together – to illustrate that the reward is shared when both hunt. In that case the corresponding payoff relations become: $s > h_1 > h_2 > c$.

When this scenario is modelled as an *N*-person dilemma however, there are additional modelling choices to be made. For example Pacheco et al. (2009) modelled the *N*-person SH game with a threshold for participants below which no public good is produced, *i.e.* the stag is not successfully hunted. Moreover, they do not make the assumption that total participation gives each individual the highest payoff. Rather, Pacheco et al. (2009) model it such that even if a given individual defects, when the threshold is reached, that individual will still receive the benefits of cooperation without having the extra cost as produced by the effort of hunting stag. Indeed, they model costs as well – creating the incentive to free-ride. While this is of course interesting, it may also sound familiar – when the threshold is reached, the payoff functions are indeed exactly the same as in the *N*-PD. As such, it does not resemble the scenario as depicted above anymore.

However, when modelled with a strict preference for each outcome, it does yield a game different from the *N*-PD. In fact, when the payoffs are transformed linearly the respective cooperative and defective functions may even be modelled to cross one another. As such, it yields an interesting pressure that may intuitively correspond best to the *2x2* scenario. Specifically, the main pressure still being to cooperate (since both functions increase as an amount of cooperation), but where the cooperative function rises faster whilst it starts lower than the defection function. As such it resembles a risk versus reward-like game where taking the risk as a group ultimately yields the best outcome – as represented in the scenario of Rousseau.

3. Underlying strategic pressures of games

To clarify the usefulness of game-theoretical properties and their related constraints, it is important to elaborate on the underlying dilemmas of various games – both for *2*-person games and *N*-person versions thereof. In the following chapter the strategic pressures on an individual player in the five archetypical games will be discussed. First, a discussion will follow regarding the general taxonomy of *2x2* games, to give a well-defined foundation for the five archetypical games discussed here. Second, the pressures for the *2*-person versions of the games will be considered, after which these pressures will be considered when regarding the conversion to *N*-person versions thereof.

3.1. General taxonomy

What is here described as an abstract framework is nothing new per se, it has been attempted successfully in many ways for *2x2* (strictly) ordinal games and even for some *N*-person scenario's (*e.g.* Rapoport & Guyer, 1966; Rapoport, 1967; Hamburger, 1974; Komorita, 1976; Dawes, 1980; Liebrand, 1983; Kilgour & Fraser, 1988; Wang & Yang, 2003). Perhaps one of the most prominent examples for two-person scenarios is Kilgour and Fraser's (1988) work "A Taxonomy for all ordinal *2x2* games". In their work, they develop a taxonomy for the 726 *2x2* ordinal games including an ordering for these games through their inherent properties. This amount of games is not randomly chosen, rather it follows from several (minimal) constraints on the class of *2x2* ordinal games. While there are only 78 strict ordinal games (Rapoport & Guyer, 1966), Kilgour and Fraser (1988) extended this by relaxing the strictness constraint. This means that while the format is still normal form, and there are still two players with each two strategies, their respective payoffs are not necessarily strictly ordered. A strict ordering of preference being represented by the numbers $1 < 2 < 3 < 4$, for each of the four outcomes of an individual player. As such the payoff orderings for each individual player become A-H, as described in *Table 1*.

Table 1.

The different preference orderings available in a normal form 2x2 ordinal game.

Class	Ordering	Description	# of possibilities
A:	1, 1, 1, 1	Indifferent among all 4 outcomes	1
B:	1, 1, 1, 2	Indifferent among three least preferred	4
C:	1, 1, 2, 2	Indifferent between two least and two most	6
D:	1, 1, 2, 3	Indifferent between two least preferred	12
E:	1, 2, 2, 2	Indifferent among three most preferred	4
F:	1, 2, 2, 3	Indifferent between two middle preferences	12
G:	1, 2, 3, 3	Indifferent between two most preferred	12
H:	1, 2, 3, 4	Strict preference for each outcome	24

From "A taxonomy of all ordinal 2x2 games", Kilgour D.M. & Fraser N.M., 1988, Theory and Decision, 24(2), 99-117.

In other words, *75* possibilities for each player resulting in $75 \times 75 = 5625$ different possibilities when including asymmetric games. However – to eliminate strategically equivalent games – both the players, and the rows and columns may be swapped. This results in equivalence between some of the games, reducing the number to *726* strategically distinct games. These games are then ordered (as subsets in which a game may or may not fall) by Fraser and Kilgour (1988) on four inherent properties: symmetricity; ordinality; complete opposition; and, no-conflict. Symmetricity is defined here as both players having the same set of strategies available to them with the same set of outcomes for each strategy. Ordinality regards the preference ordering of the payoffs (as shown in *Table 1*). Strict ordinality is then defined as both players having the same strict preference ordering over the outcomes (*i.e.* both players have a preference ordering of class *H*), mutual ordinality as both players having the same (either strict or non-strict) preference ordering over the outcomes (*e.g.* both players have a preference ordering of class *F*), and non-mutual ordinality as each player having different preference orderings over the outcomes (*e.g.* player 1 having a preference ordering of class *D*, whereas player 2 has a preference ordering of class *G*). Furthermore, complete opposition is defined as players having absolutely no common interest regarding each outcome. The last property is no-conflict, which is defined by Kilgour & Fraser (1988) as both players having an available strategy which results in the best outcome for each player. These properties and their subclasses of games are summed up in *Table 2*.

Table 2.

The 2x2 ordinal game property statistics.

	SY				$\overline{SY}$				
	NC		$\overline{NC}$		NC		$\overline{NC}$		
	CO	$\overline{CO}$	CO	$\overline{CO}$	CO	$\overline{CO}$	CO	$\overline{CO}$	Row totals
SO	0	6	0	4	0	15	3	50	78
MO$\overline{SO}$	1	4	0	2	0	10	8	19	44
$\overline{MO}$	0	9	0	4	0	206	0	385	604
Column totals	1	19	0	10	0	231	11	454	726

Legend:
- SO: Strict ordinal
- MO$\overline{SO}$: Mutually ordinal but not strict ordinal
- $\overline{MO}$: Not mutually ordinal
- SY: Symmetric
- $\overline{SY}$: Asymmetric
- NC: No-conflict
- $\overline{NC}$: Not no-conflict
- CO: Complete opposition
- $\overline{CO}$: No complete opposition

Note. The 2x2 ordinal game property statistics. Showing the numbers of games that are no conflict or of complete opposition (or the remaining case, mixed motive), both symmetric and asymmetric, in the strict ordinal, mutually ordinal and not mutually ordinal categories. From "A taxonomy of all ordinal 2x2 games", Kilgour D.M. & Fraser N.M., 1988, Theory and Decision, 24(2), 99-117.

As may be observed in the table, most of the games are asymmetrical and two-thirds of the symmetrical ones are no-conflict games, emphasising the influence of the constraints that may be modelled. In addition, some constraints – like the not no-conflict constraint – are useful to have, because they may exclude games which are trivial (*e.g.* in no-conflict scenarios both players have the same most-preferred outcome, resulting in the strategy that leads to it to always be chosen by both rational players). However, as will be discussed below, while some constraints are useful for *2x2* games, they may not be so for *N*-person variants thereof – which is why the SH game is included in this thesis.

3.2. Strategic pressures in 2-person scenarios

3.2.1. Four archetypes

Rapoport (1967) defines a 'mere' four games with the constraints of strict ordinality, symmetricity and not no-conflict, for *2x2* games. The addition of the fifth game is through the relaxation of the not no-conflict constraint. This game is the SH game, and will be elaborated upon separately after this section. The four games Rapoport (1967) defines represent four distinct types of strategic pressure they put on each player and are labelled “exploiter”, “leader”, “hero” and “martyr” and are shown in matrices *1-4*. These pressures are seen as the motivations of a player to shift from the so-called natural (*i.e.* minimax) outcome of the game, three of which are motivated through a greater payoff for the shifting player. When shifting: an exploiter is rewarded while the other player is punished; a leader rewards both, but himself more than the other; a hero rewards both but the other more than himself; and finally, a martyr punishes him- or herself while rewarding the other, but if both shift both would gain. These pressures are represented by the first four of the five games elaborated upon in the previous chapter, as shown in the matrices on the next page. The “Prisoner’s dilemma” represents the martyr type (*Matrix 1*), the game of “Chicken” represents the exploiter type (*Matrix 2*), and a form of “Battle of the Sexes” the leader type (*Matrix 3*). The hero type has no well-researched representative game so far (*Matrix 4*), but is here simply referred to as the Hero game.

	C_2	D_2
C_1	*3, 3*	*1, 4*
D_1	*4, 1*	*2, 2*

Matrix 1

	C_2	D_2
C	*3, 3*	*2, 4*
D	*4, 2*	*1, 1*

Matrix 2

	C_2	D_2
C_1	*2, 2*	*3, 4*
D_1	*4, 3*	*1, 1*

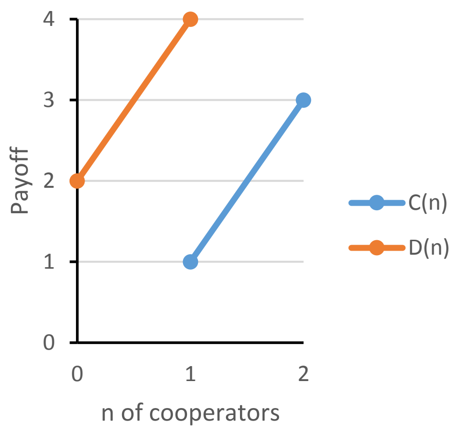
Matrix 3

	C_2	D_2
C_1	*2, 2*	*4, 3*
D_1	*3, 4*	*1, 1*

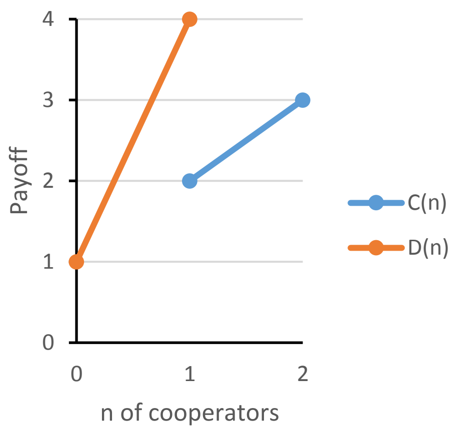
Matrix 4

The payoffs in the matrices need not be exactly the numerical values they are here. Rather, they represent orderings between the payoffs and as such, pressures for an individual player to choose a specific strategy.

It is important to have these strategic pressures and as such, constraints that enable them. For example Dawes (1980) argues social dilemmas are defined by two simple properties: “(a) each individual receives a higher payoff for a socially defecting choice (e.g. having additional children, using all the energy available, polluting his or her neighbors) than for a socially cooperative choice, no matter what the other individuals in society do, but (b) all individuals are better off if all cooperate than if all defect.” In other words, the defection payoff should always be higher than the cooperative one for each individual player, regardless of the strategies of others, and the ‘social welfare’ (*i.e.* average payoff) is lowest when nobody shows cooperative behaviour. However, Liebrand (1983) argues that the requirement of a dominant strategy for each individual is not crucial for a social dilemma. Rather, he argues that the first constraint of Dawes (1980) should be reformulated as follows: each individual has a strategy that yields the best payoff in at least one configuration of strategy choices and that has a negative impact on the interests of the others involved. This relaxation of the constraint allows for a games such as Chicken, BoS, and SH. It is noteworthy that this formulation allows for the conversion from *2*-person to *N*-person scenarios, since it abstracts away from individual pressures. It should be clear that the games representing a social dilemma are the PD and the Chicken games (matrices *1* and *2*). Of course the requirement of a dominant strategy should be relaxed to allow for Chicken, but when transforming Chicken linearly, it models a similar choice between individual gain and social gain as the PD. In fact, the PD and Chicken games only differ in the severity of punishment between being the 'sucker' (*i.e.* cooperating when the other defects) and being the 'egocentric' (*i.e.* both defecting). As a consequence, the exploiter and martyr types are quite similar in the essence of the dilemma, while they differ in the pressure they put on an individual to defect (the PD pressuring more toward defection). Of course there are different pressures from the minimax outcome of the game as put forward by the different types, but – as will be discussed in the next section – especially the essence of the dilemma is important when converting to *N*-person versions of the *2x2* games. This essence is shown clearly in graphs *1* and *2* by the gain in payoff for every strategy when cooperation increases, while defecting yields a higher individual payoff relative to the cooperative payoff, at any amount of cooperation.



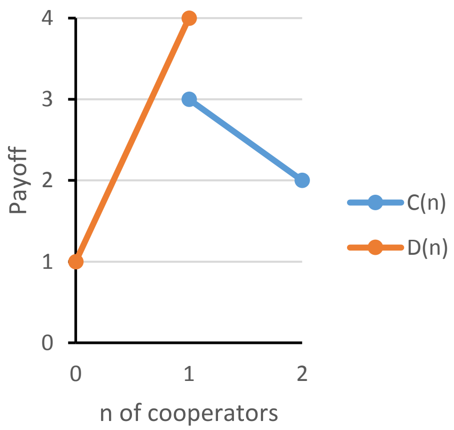
Graph 1. The 2x2 Prisoner's Dilemma as represented in Matrix 1. With the respective cooperative and defective payoffs modelled as a function of the number of cooperators.



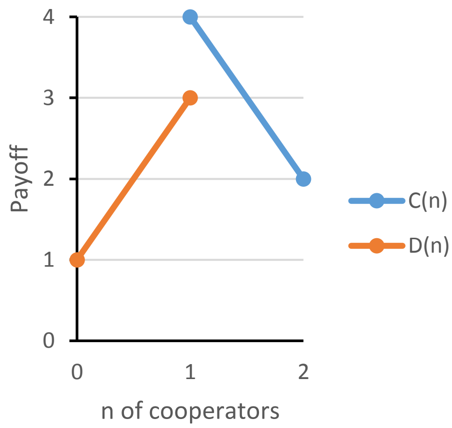
Graph 2. The 2x2 game of Chicken as represented in Matrix 2. Modelled in the same manner as Graph 1.

In contrast, matrices *3* and *4* do not have a dilemma like the PD and Chicken games. Rather, they model a form of coordination dilemma, where the social welfare is highest when –

perhaps somewhat confusingly – the strategies are coordinated such that they are uncoordinated (*i.e.* if one cooperates, the other should defect, or the other way around). Another pressure is that the defector in *Matrix 3* (or the cooperator in *Matrix 4*) receives a relatively higher payoff, but only when coordination succeeds. As such, there are two similar pressures in these *2x2* games, formulated here to allow for *N*-person conversion: (a) all individuals are better off if the strategies are coordinated, such that the distribution between cooperative and defective choices is spread evenly; and (b) one strategy (defecting in *Matrix 3*, or cooperating in *Matrix 4*) is relatively more rewarding than the other when coordination succeeds. A third pressure differs for matrices *3* and *4*: (c) each individual is worse off choosing the more rewarding strategy if coordination fails in *Matrix 3* (as opposed to both players choosing the other strategy), while in *Matrix 4* each individual is better off choosing the more rewarding strategy. Again, the essence of the games is seen clearly in graphs *3* and *4* below – where both payoffs increase toward equal distribution of cooperators and defectors. The Hero game represented by *Graph 4*, has the same essence as the BoS game, but with a slightly different pressure. This difference – reflected by (c) – is trivially seen in the graphs. The BoS game has a higher baseline payoff for cooperation, but a higher maximum payoff for defection, reflecting reward versus risk respectively. However, in Hero, cooperation yields less risk and the best reward, thus lacking a balancing pressure between cooperation and defection. As such, it effectively has the same pressure as the PD game – with cooperation as a dominant strategy rather than defection, and an emphasis on equal distribution of cooperators and defectors to achieve maximum social welfare rather than an emphasis on cooperation.



Graph 3. The strictly ordinal 2x2 Battle of the Sexes as represented in Matrix 3. Modelled in the same manner as Graph 1.



Graph 4. The 2x2 game of Hero as represented in Matrix 4. Modelled in the same manner as Graph 1.

However, here it is argued that relaxing some constraints does not necessarily depreciate these *2x2* games in regard to representing social dilemmas, as will be elaborated upon below.

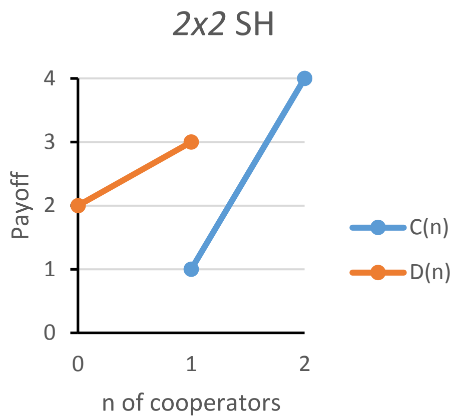
3.2.2. Fifth archetype

When the not no-conflict constraint is relaxed, it allows for a game Liebrand (1983) calls the 'Trust game', but is more commonly referred to in literature as a strictly ordinal variant of the 'Stag Hunt' dilemma (Skyrm, 2004; Brams & Kilgour, 2009) – as depicted in *Matrix 5* below.

	C_2	D_2
C_1	*4, 4*	*1, 3*
D_1	*3, 1*	*2, 2*

Matrix 5

It is clear that here there is indeed no conflict between players, since the best strategy for each is to cooperate. While this game still yields interesting strategic pressures, it is no longer a social dilemma nor a coordination dilemma. Rather, it is a form of coordination game, but where there is still a dilemma in the BoS game, it is absent here. Instead, the trustworthiness of your opponent is the only pressure being applied – failing to trust in your opponent results in a deficient outcome, where the cooperating player is punished more relative to the defecting player. When individual payoff is less important than the payoff relative to the other player (i.e. winning is more important than absolute gains) it does become a dilemma, but involves changing the utility of individual payoffs and is not inherent to the game itself. The *2x2* game – without this extra interpretation – results in a trivial game if each player is considered rational. However, the fact that – like the BoS game – one strategy is relatively high-risk high-reward, versus the other relatively being low-risk low-reward, does yield interesting properties for an *N*-person variant of this game, as will be discussed in the next section. This effect is reflected by the deficient minimax outcome, where – again like the leader type of BoS – the unilaterally shifting player may grant both a better outcome, but themselves more than the other. In contrast, the other player may then unilaterally shift to grant both the best outcome, which represents the trust dilemma. Furthermore, this effect may already be seen in *Graph 5* (on the next page) - where both cooperative and defective payoffs increase as the amount of cooperation increases, but the cooperative payoff more than the defective, while the defection payoff has a relatively higher baseline.



Graph 5. The strictly ordinal 2x2 Stag Hunt dilemma as represented in Matrix 5. Modelled in the same manner as Graph 1.

Clearly, there are many possible manners and reasons to consider constraints and it is important to formulate them precisely. Of course with less, different or more constraints, either more or less properties may become available. However, these five types of pressure represent some of the most well-studied games and perhaps illustrate the most interesting social scenarios.

4. Strategic pressures in N-person scenarios

4.1. General considerations

The conversion from 2 to *N*-person games has differences in strategic pressures on individual players. According to Dawes (1980), three strategic differences may be observed between the *2*-person Prisoner's Dilemma and the *N*-person version thereof, respectively: (1) the specificity of harm; (2) the (lack of) possibility of anonymity; (3) diluted pressure on the outcome. While Dawes (1980) focused on the PD game only, all of these differences apply to any conversion from a *2x2* game to an *N*-person version thereof. The differences will now be elaborated upon in the light of each of the five archetypical games.

The first regards the difference that the threat an arbitrary individual may pose, is spread out across the amount of participants in an *N*-person game. This difference is relevant for the PD, Chicken, BoS and Hero games. For the SH game it is not relevant. This is because in the former games a social dilemma is depicted in one way or another. For the PD and Chicken games individual gains are opposed to social gains. For the BoS and Hero games there is a focus on coordination (and thus social gains), but one strategy is still better than the other, resulting in the player choosing the suboptimal strategy to be able to punish the other players (and their selves) by not coordinating. For the SH game however, the more players cooperate the greater each individual payoff becomes – to the point where cooperating is the optimal solution for each individual. As such, punishing comes irrelevant.

The second stems from the anonymity *N*-person games grant. In a *2x2* scenario, given the payoff matrix, an individual can deduce what the other player's strategy must have been. All else being equal, this cannot be deduced in an *N*-person scenario, where an individual theoretically can only deduce how many of the total chose a certain strategy (which may prove to be hard in practice, as *N* increases). Of course this difference is relevant for each of the games discussed, although it is somewhat more relevant for games where individuals are able to punish other players for obvious reasons.

The third difference stems from the first, the control an individual has over the outcome of the game is decreased with each increment of the number of players. Of course with the addition of players the pressure one individual is able to exert on the outcome decreases, since each individual needs to have some form of pressure. Because of this, the influence one individual has on the outcome is divided by the number of total players *N*. This is relevant for all *N*-person games discussed. Specifically for the SH game this is relevant, since it is a form of trust game. This trust now no longer regards a single individual, but rather a group, diluting the risk an individual takes by cooperating.

The aforementioned considerations portray that strategic differences are imposed by the addition of players – regardless of the type of game involved. However, these differences do seem to generalise better to real-world scenario's, which is of course desired. Nevertheless, when converting games to *N*-person versions of these games, there are other differences to consider. For example, how to transform the payoffs to keep the same payoff structure.

4.2. Linear transformation to N-person games

As mentioned in *Chapter 2*, transforming the payoffs while keeping the same payoff structure may be done in several manners, such as having a certain cut-off point where a given outcome in the scenario results, or a continuous transformation. In literature this latter transformation is the general manner to convert *2x2* games to their *N*-person versions for most games (Hamburger, 1973), and – to keep the pressures as consistent as possible – this is often done in a linear fashion. Hamburger (1973) introduces the notion of 'compound games' to refer to this generalisation of *2x2* games to their respective *N*-person variants. In compound games, each player in an *N*-person game is seen as playing several *2*-person games simultaneously. In other words, the *2x2* game underlies the *N*-person game and each player in an *N*-person game is seen as playing several *2*-person games simultaneously. The payoffs for each individual are simply calculated by the increments of the payoffs for each strategy of the *2x2* matrix, multiplied by the amount of co-operators. The *2x2* matrices then serve as the four starting points of the *N*-person counterparts. This is also why the previous *2x2* graphs are formatted as such, since these form a clear picture of the conversion to their *N*-person scenarios. These *N*-person variants should look familiar, since they indeed include the same payoffs as the *2x2* matrices and graphs, but linearly transformed accordingly. This is done using the following formulas:

$$C(n) = b + (a - b)(n - 1) \quad \text{and} \quad D(n) = d + (c - d)n \,,$$

where n refers to the number of cooperators, and the other letters in the formulas refer to the standardised *2x2* matrix, as shown on the next page in *Matrix 6*.

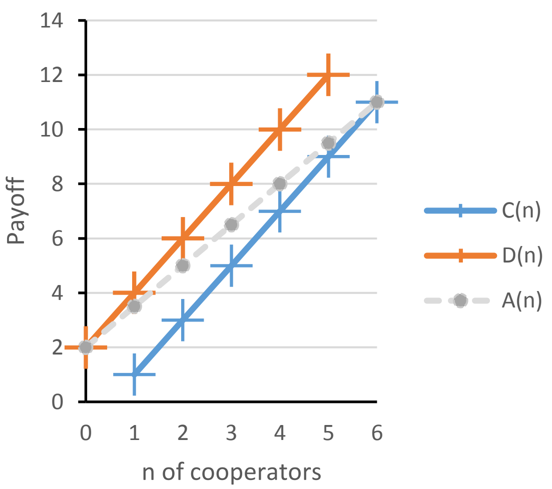
	C	D
C	*a, a*	*b, c*
D	*c, b*	*d, d*

Matrix 6

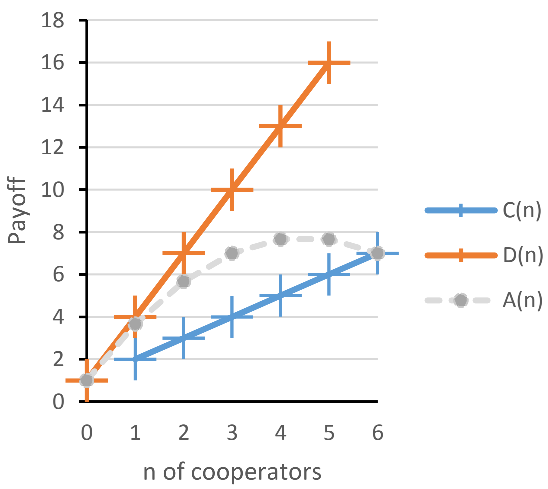
Each of the games can now be linearly transformed from their *2x2* format to their *N*-person variant, using their *2x2* matrices as baseline. However – as will be shown in the following graphs – not every archetypical game yields the intuitively expected scenario when transformed as such. Furthermore, these graphs all assume a *6*-person game for consistency, but may be any arbitrary number of players.

4.2.1. Transformation of social dilemma games

The first two archetypical games are the PD and Chicken games. These games do intuitively correspond to their linear transformations. The *N*-person PD and Chicken games are shown below in graphs *6* and *7* respectively.



Graph 6. 6-Person Prisoner's Dilemma. Here C(n) and D(n) are the payoff slopes of cooperators and defectors, respectively. A(n) is the average payoff.

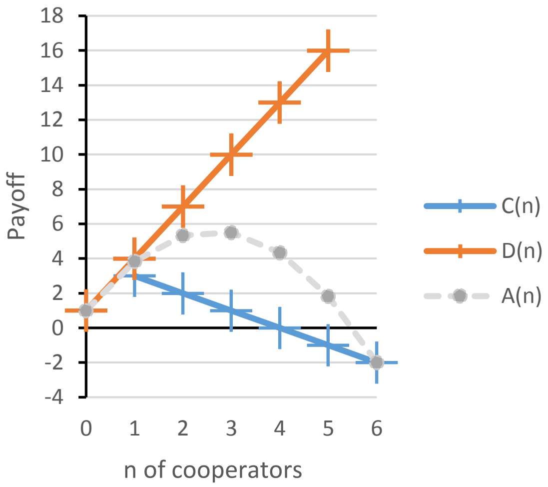


Graph 7. 6-Person Chicken.

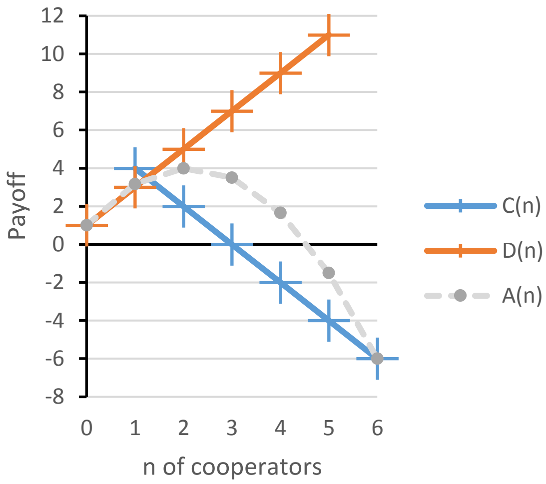
Both graphs 'look' the same as their *2x2* counterparts. They are also intuitively correct when regarding their respective social scenarios. Given that Chicken being modelled as cars driving toward each other would be modelled differently, the social dilemma as previously mentioned by Dawes (1980) is certainly modelled as such.

4.2.2. Transformation of coordination dilemma games

For the BoS game and the Hero game, the *N*-person graphs when linearly transformed with their *2x2* matrices as baseline are shown below.

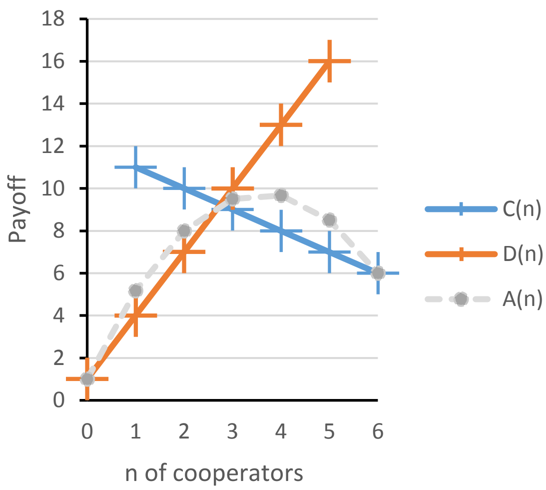


Graph 8. 6-Person Battle of the Sexes when positively linearly transformed (i.e. the same way as the PD and Chicken games).

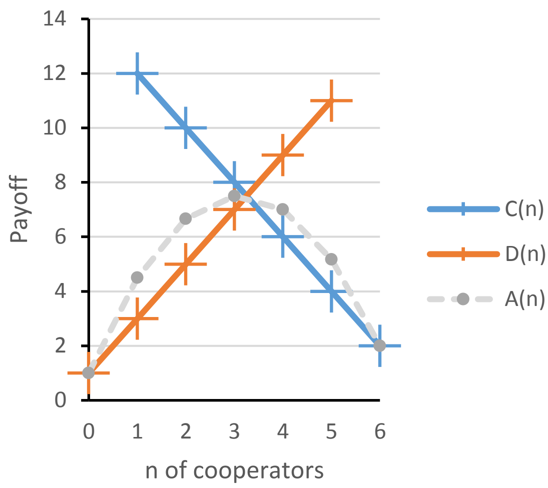


Graph 9. 6-Person game of the underlying 2x2 game of Graph 4, when positively linearly transformed (the same way as the PD and Chicken games).

It should immediately be clear that this transformation results in an inadequate representation of the underlying *2x2* game. Although the maximum social welfare – as represented by the average payoff – is still at the median number of players for the BoS game, it is no longer so for the Hero game. In addition, the payoffs at that amount of cooperation are far apart rather than as close as possible for both games. Furthermore, the pressure toward coordination is outweighed by a pressure toward defection, especially as *n* increases. In fact, the defection payoff is always higher for the BoS game and almost always for the Hero game. However, by modelling the starting points differently, the games may represent their *2x2* counterparts more adequately, as shown in graphs *10* and *11* on the next page.



Graph 10. 6-Person Battle of the Sexes when positively linearly transformed, but with different starting points: d = 7 and c = 10 as defection points, and b = 9 and a = 8 as cooperation points, as represented by Matrix 6.



Graph 11. 6-Person game of the underlying 2x2 game in Graph 4, when positively linearly transformed but with different starting points: d = 5 and c = 7 as defection points, and b = 8 and a = 6 as cooperation points, as represented by Matrix 6.

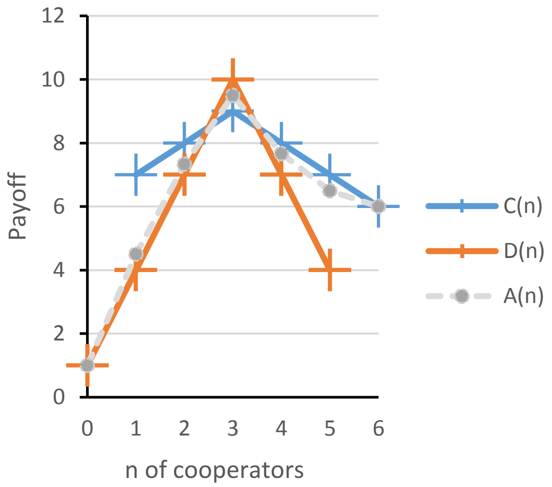
Here the same functions are used but with a crossover point at the median amount of cooperation. In other words, the point of maximum coordination – as identified by the median $(\frac{N}{2})$ of total outcome pairs – now becomes the baseline of the function. An outcome pair being defined here as the outcomes for both cooperation *C(n)* and defection *D(n)*, at a given amount of cooperation *n* – where for either cooperation or defection the outcome may be empty (*i.e.* *C(n)* = *Ø* when all defect, or *D(n)* = *Ø* when all cooperate). In order to describe the transformation from the underlying *2x2* games formally, the aforementioned functions are adjusted slightly for BoS and Hero:

$$C(n) = y + (a - b)(n - \tfrac{N}{2}) \qquad \textit{and} \qquad D(n) = x + (c - d)(n - \tfrac{N}{2})\,,$$

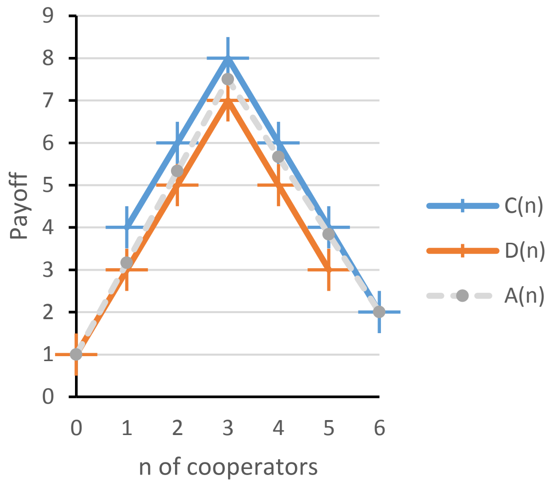
$$\textit{where:} \qquad x = d + (|c - d|)\frac{N}{2} \qquad \textit{and} \qquad y = x + (b - c)\,,$$

where *x* and *y* represent the new baseline of each function. As such, the first part of the functions represent the baseline conversions of the *2x2* game, whereas the second part still resembles the linear increments for both payoffs at each number of cooperators. These functions generalise to any number of players and keep the crossover point at the median number of players. These functions may be applied to linearly transform not only the BoS and Hero games, but the PD as well. It is noteworthy however that an even number of players is assumed here to establish the median.

While both games now more closely resemble their *2x2* counterparts, maximum payoff – regardless of cooperative or defective strategies – is no longer at the maximum amount of coordination, which is the main pressure in the *2x2* game. While this results for both cases in a strategically interesting game, it does not intuitively yield the same pressures as its *2x2* counterpart. However, when another relatively more complicated manner of modelling is used, both games resemble their *2x2* counterparts as described in the previous section as closely as possible (as shown in *Graph 12* and *Graph 13* below).



Graph 12. 6-Person Battle of the Sexes when both positively and negatively linearly transformed from the coordination point (i.e. differently from the PD and Chicken games).



Graph 13. 6-Person game of Hero, when both positively and negatively linearly transformed (unlike the PD and Chicken games) from the coordination point

In these graphs the 'coordination point' is used as baseline as well, but where the function is linearly negative when moving away from this point, it is inversed to a positive linear function toward the point. Described more formally, the same formulas introduced on the previous page may be used, but with a condition:

$$Iff: n \leq \frac{1}{2}N,$$

$$C(n) = y - (a-b)(n - \tfrac{N}{2}) \quad \text{and} \quad D(n) = x - (c-d)(n - \tfrac{N}{2}),$$

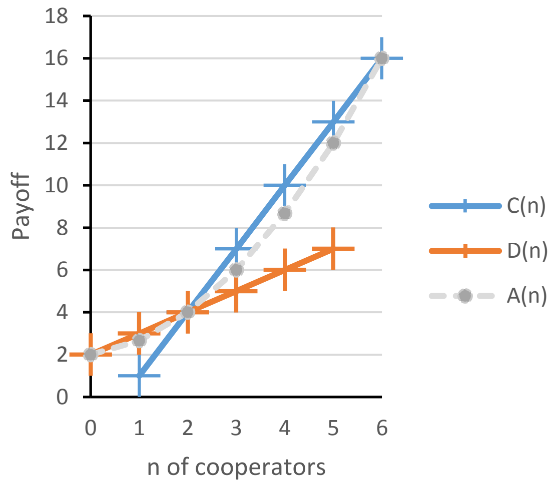
$$\textit{else:}$$

$$C(n) = y + (a-b)(n - \tfrac{N}{2}) \quad \text{and} \quad D(n) = x + (c-d)(n - \tfrac{N}{2}),$$

The only difference between the two former and the two latter functions is that the interval between payoffs is inversed to a negative one in the former two. As such, both of these games represent their strategic pressures as discussed for their *2x2* counterparts as adequate as possible. It does come at a cost however, which is that neither payoff is a simple linear function anymore.

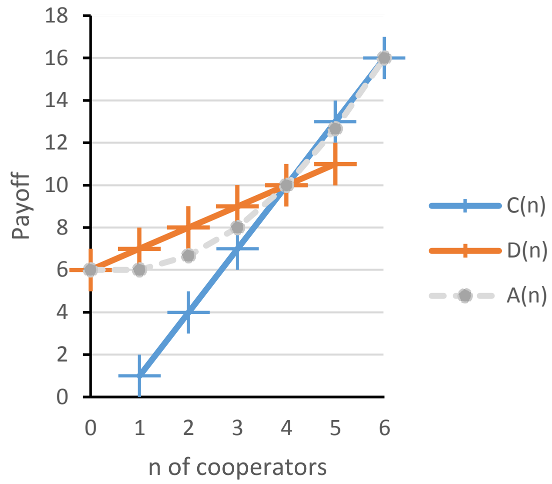
4.2.3. Transformation of trust dilemma game

For the SH game, a similar problem arises when converting the game to its *N*-person variant. Although both the coordination and the defection payoffs are positive functions of the amount of cooperation – a choice remains to be made regarding the strategic pressures at each amount of cooperation. When the *2x2* matrix is simply linearly transformed (in the same manner as the PD and Chicken games) it results in a game as illustrated by *Graph 14* below.



Graph 14. 6-Person Stag Hunt dilemma, when positively linearly transformed from its respective 2x2 matrix (i.e. the same way as the PD and Chicken games).

While this graph does resemble the same kind of pressure as its *2x2* counterpart, it should be noted that by transforming it as such, the point at which cooperation becomes relatively more profitable than defection will always be at $n = 3$, regardless of the number of players. This is of course an unwanted property since the cooperation payoff increases at a faster pace than the defection payoff and as such the game would gradually become more trivial as the number of players increases. To see why, take for example an SH game played by 15 individuals, the cooperation payoff would be at least as good or better than the defection payoff at every point other than $n < 3$, resulting in 14 out of 16 total outcome pairs where the payoff for cooperating is equal or greater than the defection payoff. For such a game it is trivial to see that cooperating is by far the better strategy. Now take the same game played by 4 individuals, in this case 3 out of 5 outcome pairs yield an equal or greater payoff for the cooperative strategy (where 1 outcome pair is indeed equal). Thus, it is important to decide on a certain point where cooperation becomes more beneficial than defection and – similar to the BoS and Hero games – this point should scale with the number of players. To intuitively resemble the *2x2* format this point should fall at $(\frac{N}{2})$, since in that case the median of the outcome pairs yields greater payoffs for defectors, whereas the rest yields an equal or greater payoff to the cooperators. As such, it most closely resembles its *2x2* counterpart, because – as seen in *Graph 5* – the median amount of cooperation ($n = 1$) is the last point where the defective strategy yields a greater payoff than the cooperative one. To reflect the same cross-over point, it is modelled in *Graph 15* at $n = 3$, as shown below.



Graph 15. 6-Person Stag Hunt dilemma, when positively linearly transformed to most closely resemble its 2x2 matrix (i.e. the same way as in graphs 10 and 11)

As may be seen in the graph, the slope of both the defection and the cooperation functions remains the same as the *2x2* slope (*i.e.* both payoff slopes have the same intervals as their respective *2x2* slopes). Hence it is in essence the same linear transformation as before, but with a different starting point, chosen here such that it scales with the number of players at $(\frac{N}{2})$. In fact, the same formulas as the ones for the BoS and Hero games may be used (but with a slight adjustment since the cooperation function increments faster than the defection function):

$$C(n) = x + (a - b)(n - \tfrac{N}{2}) \quad \textit{and} \quad D(n) = y + (c - d)(n - \tfrac{N}{2}) \ ,$$

where:

$$x = b + (|a - b|)(\tfrac{N}{2} - 1) \quad \textit{and} \quad y = x + (c - b) \ ,$$

The interesting part however, is that this *N*-person variant of the *2x2* SH game is intuitively less trivial than its *2x2* counterpart. This may be explained through Dawes' (1980) dilution of pressure on the outcome as mentioned in the previous section. The risk versus reward pressures of the game are now spread out across several individuals rather than one, making it harder to gauge what strategy a given individual may choose. Of course in the sense of a rational player the game remains as trivial as before, but humans rarely are strictly rational.

4.2.4. Conclusion on linear transformations of N-person games

To conclude, while there is a general consensus regarding the linear transformation of *2x2* games to their respective *N*-person counterparts, there still are some choices to be made regarding their exact representation. The manner in which these implementations are represented are of vital importance for the resulting underlying social phenomena of these games. While it does not matter which starting point is chosen for the PD game (since it is separable, a constraint elaborated upon in the next chapter), it does for every other game.

In contrast, while it does matter for the Chicken game, it is to the knowledge of this author most often represented as it is here. This is probably since it resembles the strategic pressures for a social dilemma as put forward by Dawes (1980) and Liebrand (1983). As for the other games, research is limited and choices regarding the structure are usually done in an arbitrary but well-argued fashion (to fit the respective research) without a general consensus.

For the BoS game it is generally agreed upon to linearly transform it, but it is still modelled differently in different papers. Starting points differ and whether the payoffs are defined as a function of cooperators or as a function of individuals playing the same strategy differ among researchers (Szilagyi & Somogyi, 2008; Zhao, et al., 2008).

The consensus regarding the SH game is generally to model a certain threshold. For example Skyrms (2004), and Pacheco et al. (2009), modelled the SH game with a cut-off point *M*. Below this cut-off defectors simply receive the lowest payoff and cooperators receive the lowest payoff minus a fixed cost, resulting in two horizontal payoff slopes (where defection is still relatively better than cooperation). Above the cut-off point, the payoff slopes are modelled as they are here for the PD game, where both increase as a function of the amount of cooperation but defectors outcomes are always better than their respective cooperative outcomes. This results in a drastically different game then the one depicted in this paper, but does resemble the classic SH scenario of an either successful or unsuccessful hunt for stag, with the assumption of being able to free-ride when success is achieved. However, it is here argued that how the game is modelled here is distinct from the others presented and does yield interesting pressures.

5. Inherent properties of games

As indicated in the previous chapter, defining the *N*-person variants of the respective *2x2* games creates more dynamics to such games, yielding some significant differences. Therefore, it is important to explore the dynamics of this extension and the constraints necessary to provide the same interesting properties and pressures *2x2* games may yield. In this chapter a more in-depth discussion on the inherent properties of each of the games will follow. First, the inherent properties for their classical *2x2* variants are discussed and how these translate to their respective *N*-person variants. Second, some new properties are introduced that are specifically relevant to the *N*-person variants.

5.1. Inherent properties for 2-person scenarios

In a comprehensive paper, Hamburger (1973) has explored the differences between properties of a *2x2* Prisoner's Dilemma and *N*-person versions thereof in an attempt to provide a taxonomy. Some of these differences are relevant for any *N*-person symmetric version of Rapoport and Guyer's (1966) original *78* different ordinal *2x2* games, whereas others are PD specific. However, the properties are generalizable to any game and as such serve as a useful template to discuss the other games. Therefore, the properties will be summarised for the PD game, after which the other games will be elaborated upon in light of these properties. The quintessential properties of a *2x2* PD game according to Hamburger (1973) are as follows:

- *2*-person;
- 2-alternative;
- ordinally symmetric (*i.e.* symmetrical strictly ordinal payoff preferences);
- each player has a dominant strategy (*i.e.* a strategy that yields a better outcome regardless of the other player's strategy);
- each player has a most-threatening strategy (*i.e.* a strategy a player has that is least preferred to the other player);
- dominant and most-threatening strategies coincide;
- strongly stable equilibrium (*i.e.* neither player receives maximum payoff, but if either deviates unilaterally from the equilibrium it causes the other to receive maximum payoff);
- deficient equilibrium.

However, when converting the *2x2* games to their *N*-person variants some of these properties are either lost, changed, or weakened. Specifically, the first two are of course not necessary, but the second is retained here. In addition, ordinal symmetry holds for any of the *2x2* games discussed here, but is changed for their *N*-person variants. It is no longer ordinal, but may still be imposed on an interval level, as will be discussed in the next section. Furthermore, the property of a strongly stable equilibrium needs to be weakened to allow for *N*-person variants, since it refers to the points right next to the equilibrium. It should be clear that the *N*-PD (or any other *N*-person game) cannot allow for a strongly stable equilibrium as defined here because it would imply that a unilateral deviation from it would cause all other players to receive maximum payoff, which is too strong. Rather, an outcome may be called strongly stable if shifting unilaterally would benefit all others while it hurts the shifting player. As such, a stable outcome may be defined as the unilateral shift of a player being harmful to both but equally or more so for the shifting player. Subsequently, a weakly stable outcome may be defined as the unilateral shift of a player being harmful to both, but more so for the others than for the shifting player. Finally, an unstable outcome may then be defined as the unilateral shift of a player to be beneficial for either themselves alone, or for both themselves and the others.

5.1.1. Inherent properties of PD

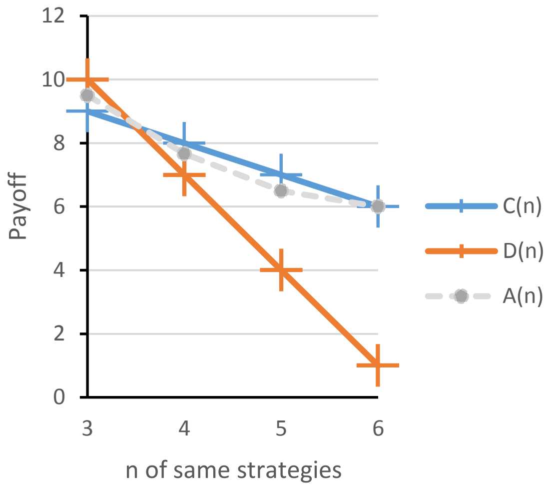
With respect to the properties mentioned above, the essential properties for a PD scenario regardless of the number of players are that the dominant (and most-threatening) strategies result in a deficient outcome. For the *2x2* game these two properties imply all the others (Hamburger, 1973). For the *N*-person variant with the changed definitions of stability the property of a strongly stable equilibrium may be included as well.

5.1.2. Inherent properties of Chicken

In the game of Chicken players do not have a dominant strategy available, but do have a most-threatening strategy. The outcome when playing the most-threatening strategy is deficient. However, the minimax outcome is unstable, since shifting to the other strategy at that point results in a better individual outcome for the shifting player (and a worse outcome for the other). Recall that this pressure represents the exploiter type as defined by Rapoport (1967). These properties are defined such that they allow for *N*-person conversion. In other words, the properties apply to both the *2x2* and the *N*-person variant of Chicken. As such, the essential properties for the game of Chicken are the availability of a most-threatening strategy which results in a deficient outcome, and an unstable natural (*i.e.* minimax) outcome.

5.1.3. Inherent properties of BoS

Individuals playing the *2x2* BoS game have no dominant strategy available, but do have a most-threatening strategy. Similar to Chicken, this strategy is not rational, since it is also most-threatening to themselves. However, another threatening strategy is available. This 'most-threatening' strategy is to not-coordinate, which reflects the scenario behind the game – a focus on coordination rather than a specific strategy. In other words, this is no conventional most-threatening strategy that worsens the outcome of the other regardless of their strategy, but it is a threatening 'counter'-strategy that depends on the other's strategy. Interestingly, the player that does not receive maximum payoff at the equilibrium has a threatening counter-strategy at that point: to shift unilaterally means both get the second lowest payoff. It is more threatening for the individual that received the maximum payoff, than it is for the individual that received the second-best payoff, since the former loses relatively more. While this strategy may not be useful for single-round games, in repeated games it becomes a useful concept. This is reflected by the *N*-person variant of the game, where the most-threatening strategy property no longer applies, since it is dependent on the other players' strategies. To see why, consider a player who cooperated in the previous round where two others cooperated, this player could have defected to grant all players a worse payoff. In contrast, a player who defected in the previous round where four of the others cooperated, would have granted all others a worse payoff when this player would have cooperated. As such, the most-threatening strategy an individual may pose is dependent on how many others chose the same strategy. If more than half of the total number of players chose that same strategy it is the most-threatening, otherwise the other strategy is most-threatening (as may be seen in *Graph 16* on the next page).



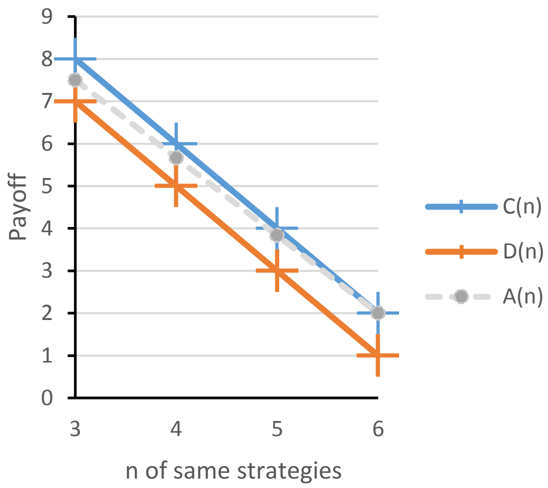
Graph 16. The 6-Person Battle of the Sexes as shown in Graph 12, but as a function of the same number of strategies played in a round. This same number of strategies is the maximum value of same strategies (*e.g.* when 2 cooperate and 4 defect, the value is 4).

It is noteworthy that the most-threatening strategy is still not rational for a player. However, it is still more threatening to the defective players, which reflects the counter-pressure to shift away from the equilibrium point.

Furthermore, there are two equilibria in the 2x2 game, which are both stable, but not strongly stable. In other words, if one player deviates unilaterally, both get a lower payoff. For the *N*-person variant there are more than two combinations to achieve the equilibrium, and as such it seems more apt to refer to it as the equilibrium point, rather than two separate equilibria. Moreover, recall this game represents the leader type, where – according to the minimax strategies – the natural outcome is the second lowest payoff. The first player to shift from that outcome receives maximum payoff and grants the other the second best payoff, 'leading' them both to a better outcome, but themselves more than the other. In the *N*-person variant the first player to shift does not grant the best and second best payoffs, but does yield the same tend, spread out across all players. In other words, the shifting player still leads everyone to a better outcome, but themselves (and others choosing the same strategy) more than the others (choosing the opposing strategy). As such, the essential properties for the BoS game are the availability of a threatening counter-strategy which results in a deficient outcome, a stable equilibrium point, and an unstable natural outcome.

5.1.4. Inherent properties of Hero

For the *2x2* Hero game a dominant strategy is available, but no singular most-threatening strategy. The equilibria in this game are – like the BoS game – still at the coordination point. In fact, it is similar to the BoS game with regard to the availability of a threatening counter-strategy and it has the same equilibrium point. The only difference is the availability of a dominant strategy for both players (rather than a most-threatening strategy), which increases the stability of the equilibrium point. This is because the threatening counter-strategy now punishes both evenly, still resulting in a worse outcome for the player using it. Like the BoS game, the dominant strategy in Hero does not remain dominant in its *N*-person variant. This is because a dominant strategy regards the difference in payoffs an individual may grant themselves. Recall a strategy is dominant when playing this strategy always yields the better outcome, regardless of the number of others. This is not the case in the *N*-person Hero game, since when half or more than half of the players are cooperating, the dominant strategy reverses. In other words, while cooperating is the dominant strategy at any amount of cooperation less than half of the total number of players, defection is dominant above these amounts of cooperation. Again, this represents the availability of a counter-strategy, where dominance is relative to the number of players playing the same strategy. It is noteworthy however, that the cooperative strategy still always yields a better outcome relative to the defective strategy at any amount of cooperation (as shown in *Graph 17* below).



Graph 17. The 6-Person Hero game as shown in Graph 13, but as a function of the same number of strategies played in a round. This same number of strategies is the maximum value of same strategies (*e.g.* when 2 cooperate and 4 defect, the value is 4).

In addition, the natural outcome of the game also stabilises to some extent, but is still unstable. This is because a unilaterally shifting player rewards both themselves and the others, but the players choosing the other strategy more than themselves (and players with the same strategy), which is why the game is called Hero (Rapoport, 1967). As such, the Hero game has the essential properties: the availability of a dominant counter-strategy; a stable equilibrium point; an unstable natural outcome; and, the dominant counter-strategies coincide with the equilibrium point.

5.1.5. Inherent properties of SH

For the 2x2 SH game no dominant strategy is available but there is a most-threatening strategy. There is also a stable equilibrium where both players cooperate and if either deviates unilaterally it results in a worse outcome for both, but relatively worse still for the opposing player. Moreover, this does not coincide with the outcome according to the most-threatening strategies – which grants both the second-lowest payoffs. In the *N*-person variant this is no longer the second lowest payoff, since the pressure is spread out across individuals. However, it is a deficient outcome, as any individual can unilaterally shift to grant all players a better outcome, but themselves more than the others. This is also what reflects the most-threatening strategy, since for each defector the cooperators lose more relative to the defectors. As such, the essential properties are the availability of a most-threatening strategy, an unstable natural outcome, and a stable equilibrium. The former two of these are the same properties as the Chicken game. This is no coincidence, since the functions for both cooperation and defection increment by the same number of cooperators (although cooperation increments faster in the SH game and defection in the Chicken game, the strategies may be swapped). However, the property of a stable equilibrium ensures the starting points of the functions are modelled in such a manner, that it forms a representation of the SH game.

5.1.6. Conclusion on inherent properties

To conclude, the essence for each of the dilemma games (as represented here) may be summarised by the following properties:

a. the social dilemma games have either a dominant strategy, a most-threatening strategy, or both, which result in a deficient outcome;
b. the coordination dilemmas have a threatening counter-strategy or a dominant counter-strategy, while only the dominant counter-strategy coincides with the equilibrium point, and;
c. the trust dilemma game has a most-threatening strategy which does not coincide with the stable equilibrium.

Furthermore, the availability of a dominant strategy has a stabilising effect, as shown by: (1) the strongly stable equilibrium in the PD as opposed to the unstable most-threatening and natural outcomes of Chicken; and (2) the weakly stable equilibrium point of BoS, in contrast to the stable equilibrium point in Hero which coincides with the dominant counter-strategy.

5.2. Additional inherent properties for N-person scenarios

Some other properties may be introduced that are of interest for the conversion to *N*-person games specifically. In this section each of these will be briefly explained, and their interrelationships and usefulness discussed. They are defined by Hamburger (1973) as:

- separability;
- interval symmetricity;
- common-interval symmetricity;
- existence of non-oscillatory solution;
- total payoff is not less with an additional cooperator;
- monotonicity of average-payoff function.

5.2.1. Separability

The property of separability requires players to have a dominant strategy and have their highest and lowest payoffs diagonally opposite. This means that both the cooperation and

	C	**D**
C	A_1, A_2	B_1, C_2
D	C_1, B_2	D_1, D_2

Matrix 7. Reference *2x2* matrix with strict ordinal preferences. For PD: $C_i > A_i > D_i > B_i$; Chicken: $C_i > A_i > B_i > D_i$; BoS: $C_i > B_i > A_i > D_i$; Hero: $B_i > C_i > A_i > D_i$; SH: $A_i > C_i > B_i > D_i$

defection payoffs increase with the number of cooperators, and that these increase with the same interval. As such, the constraint imposes parallel payoff functions between cooperative and defective behaviour. It is illustrated here in terms of *Matrix 7*, where separability is reflected by $A_i - C_i = B_i - D_i$.

The only game that has this property is the PD game. However, when the payoffs of cooperation and defection are defined as a function of the number of players choosing the same strategy and it is ensured that these increase with the same interval (*i.e.* $C_i - B_i = D_i - A_i$), the Hero game yields this property as well. The property of separability is an honourable mention since it ensures the pressures applied to cooperative and defective players are consistent for any amount of cooperation in an *N*-person game. Furthermore, the functions as defined for the BoS may be used for any separable game, but not for non-separable games (since the difference between the starting points of each function no longer matter when the functions are separable). The non-separable *N*-person games (Chicken, BoS and SH) have differing amounts of pressure for both cooperative and defective behaviour at different amounts of cooperation. This is not necessarily harmful, and in fact it models the underlying scenario more adequately for these games (*e.g.* pressure to defect becoming greater as more people cooperate in the *N*-person Chicken game). However, it should be mentioned that this reflects the choice of distributing the pressures applied in a *2x2* game evenly among individuals in an *N*-person version thereof (*i.e.* linearly transforming the payoffs).

5.2.2. Interval and common interval symmetry

This leads to the discussion of interval symmetry and common interval symmetry, since they are closely related. Furthermore, the latter imposes linearity on the payoff functions. Interval symmetry simply means that payoffs are spaced equally in order of preference, but not necessarily symmetrical between individuals, whereas common interval symmetry is symmetrical between individuals (*e.g.* $C_1 - A_1 : A_1 - D_1 : D_1 - B_1 :: C_2 - A_2 : A_2 - D_2 : D_2 - B_2$ in terms of *Matrix 7* with regard to the PD). It is noteworthy that common interval symmetry does not need to imply separability as an implicit property. This is because common interval symmetry imposes equal differences between ordered payoffs, whereas separability imposes an equal difference between the defection and cooperation payoffs for any n (where n may refer to the number of cooperators, or the number of individuals choosing the same strategy). Of course all of the games discussed here are common intervally symmetric (since all have linearly transformed payoffs). This linearity imposes an equal spread of the pressures across players, which perhaps most closely resembles real-world scenarios. In addition, it allows for the use of relatively simple functions to convert the *2x2* games to their *N*-person variants.

5.2.3. Monotonicity of the average payoff function

In combination with separability, common interval symmetry gives rise to a monotonic increase of the average-payoff function – which ensures total payoff is not less with an additional cooperator. The monotonic increase of the average payoff reflects the consistency of pressures in a game, but is not an essential property of *N*-person games. However, the property that total payoff is not less with an additional cooperator does reflect a noteworthy alteration of pressures within a game, as will be elaborated upon below.

5.2.4. Non-oscillatory solution and skewed pressures

As mentioned before, when the constraints of separability and common interval symmetry are met, linearity is imposed on the average payoff slope. In contrast, when a game is common intervally symmetric yet not separable, the average payoff slope is no longer linear. Although this may not be problematic, it is noteworthy since it applies different pressures at differing amounts of cooperation. For social dilemma games it causes the initial increment to be relatively fast after which it gradually increments slower to the point where it decreases, as the amount of cooperation increases. To see why, consider that the function of the average payoff in a game is calculated as follows:

$$A(n) = \frac{(C(n) \times n + D(n) \times (N - n))}{N},$$

where $C(n)$ is the cooperation payoff, $D(n)$ the defection payoff, n the number of cooperators and N the total number of players. Clearly, as the cooperation increases, the payoffs for both cooperating and defecting increase – but while the payoff for defectors increases relatively faster with each additional cooperator, there is also one less defector, mediating the increase. In turn, this marginally skews the tendency of an individual to choose a specific strategy as cooperation increases and gives rise to the possibility of coordinated oscillation.

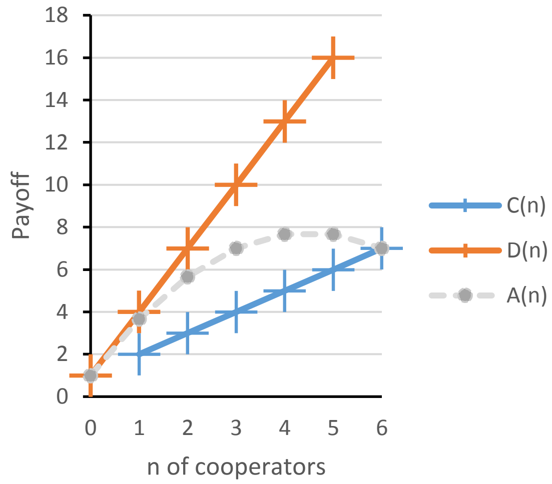
The property of a non-oscillatory solution ensures that it is not feasible for individuals to take turns in cooperating and defecting, such that the average outcome for each individual would be higher than the average when each individual is showing cooperative behaviour. In other words, $2A_i \geq B_i + C_i$ (as shown in *Matrix 7*) for *2x2* social dilemma games, PD and Chicken. For *N*-person social dilemmas it is somewhat more complicated to ensure this property, since the outcome for one individual is no longer dependent solely on one person, but rather on several others'. A non-oscillatory solution could then be rephrased to: "the total payoff ('social welfare') when full cooperation $C(n) = N$ is achieved should not be less than the total payoff of any other amount of cooperation $C(n)$"; or: "the total payoff should not be less with an additional cooperator".

It is noteworthy that this property only concerns social dilemma games, since this is the only dilemma that is represented by a mutually exclusive choice between individual or group gain. For BoS and Hero an oscillatory solution is always existent and even desired (since it represents the underlying dilemma). Furthermore, for the trust dilemma as represented by SH, it is always non-existent since cooperation ultimately yields the best outcome for all (*i.e.* no opposition between individual and group gain is present, only between the former and latter risk).

Furthermore, for N-person social dilemmas it is still improbable to achieve coordinated oscillation. The game of Chicken allows for oscillatory solutions but only at some specific amounts of cooperation when the difference between the defection payoff and cooperation payoff is greater than the total increase when there is an additional cooperator, *i.e.*:

$$D(n+1) - C(n+1) > \big(C(n+1) \times n + D(n+1) \times (N-n)\big) - \big(C(n) \times n + D(n) \times (N-n)\big)$$

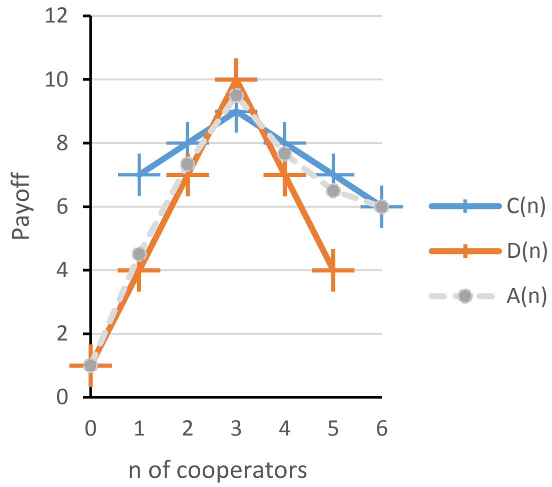
For example, in a game of six players – where an oscillatory solution exists for four cooperators – players would need to take turns in cooperating and defecting such that each round four cooperate and two defect (as shown in *Graph 18*, a reprint of *Graph 7*).



Graph 18. 6-Person game of Chicken, reprinted here for clarification. Noteworthy here is that the value of *A(n)* is the same for *n = 4* and *n = 5*, which are the highest average values.

This is of course improbable to achieve (when no communication is allowed), and even more so as the number of players increase. However, when considering mixed strategies – more often seen in agents than in humans – it is definitely a relevant pressure. For example, when every player plays a mixed strategy where each cooperates 95% of the time and defects 5% of the time, the average total payoff is greater than full cooperation. Nevertheless, this is not an oscillatory solution per se, since it is not a deterministic solution. Furthermore, this mixed-strategy pressure reflects the underlying dilemma exactly. As such, it is argued here that the property of a non-oscillatory solution is not necessary for the games considered in this thesis, while the inconsistency of the strategic pressures applied to each individual in non-separable games is noteworthy.

This inconsistency is the only difference between the PD and Chicken games as discussed here (and as such, desired). For the BoS game this inconsistency translates into a higher total payoff when moving away from the equilibrium point to the right hand side, as opposed to moving away from the equilibrium point to the left hand side (as shown in *Graph 19*, which is *Graph 12* reprinted).



Graph 19. 6-Person BoS game, reprinted here for clarification. Noteworthy here is that the value of *A(n)* is increments faster on the left hand side of the equilibrium point relative to the right hand side.

In other words – despite the fact that both payoffs have the same intervals when moving away from the point – the total payoff is relatively larger when there are more cooperators. This is not necessarily harmful, and it reflects the risk versus reward incorporated in the game. To see why, consider that the cooperative strategy yields lower potential but also a lower risk (*i.e.* higher average payoff). This relatively increases the total payoff as cooperation increases, since there are more low-risk players in that case. In fact, as it is modelled here, cooperation yields a better outcome than defection at any amount of cooperation except at the equilibrium point. This is because the difference between the baselines of both functions have the same difference as the underlying *2x2* variant. This may be modelled differently such that this difference also scales with the number of players, but it will model the preference ordering as seen in the underlying *2x2* variant differently, which is why it is here modelled as it is.

For the SH game the average payoff function increases relatively faster as cooperation increases, which is intended since it represents the underlying scenario precisely. Specifically, an increase in the tendency to take the risk, as the amount of risk-takers increases.

5.2.5. Conclusion on N-person specific properties

To conclude, several of these properties are useful to have, while only common interval symmetry is essential since it imposes linearity on the functions. Separability is useful since it further improves the consistency of the pressures, ensuring the game represents its underlying *2x2* variant as close as possible, as represented by the monotonic increase of the average payoff function. However, it is not a necessary property and without it, interesting games like BoS and Chicken may be included that model a form of risk versus reward type of pressure. The property of a non-oscillatory solution may be noteworthy for the game of Chicken but it is relatively uninteresting and certainly not necessary.

6. Final remarks and discussion

In this section an example of a manner of implementation for this framework is briefly discussed first. Second, the relevance of this work and other research related to this work is discussed. Third, an elaborate discussion of the work presented here follows. Last, a concluding summary is given of the work presented in this thesis.

6.1. Implementation considerations

To clarify what is meant with an implementation of this framework, this section is devoted to a brief illustration of a manner to achieve this. This illustration will use the example of an implementation as a mobile web-game since it is a straightforward manner of representation and useful for educational and experimental purposes. It is noteworthy however, that it is indeed an example and other manners of implementation may be used. Furthermore, choices made regarding the representation are not meant to be conclusive nor are they meant to be complete. Rather, they simply serve as an illustrative example.

In such a mobile web-game, multiple persons may join a virtual lobby with their smartphones to play one of the games mentioned in the framework. A game may then be hosted by an administrator, which should preferably be done on a computer with a large display (*e.g.* a video projector). Consequently, there are two displays available to individuals playing a game, a 'main' display – represented by the host – which is visible to all, and each individual's own smartphone display. It should be clear that the administrator plays the role of experimenter or tutor, and serves as host alone, not as a player. As such, the host may start and end games in general and rounds within each game. A round is here simply referring to the phenomenon where each player in the game chooses a strategy, which – after all players have chosen – results in a payoff for each respective individual. Furthermore, at the beginning of a game a host could have the option to show or hide different information (*e.g.* availability of different properties, play/payoff history, or total individual/group payoffs).

The main display may show the general information of the respective game, such as the payoff graphs as previously shown in chapters *3* and *4*, and different optional information. In addition, the properties of the respective games themselves may be shown for further clarification or illustration. Furthermore, the individual smartphone may be used in two manners: (1) as an input tool, allowing the individual in question to choose one strategy or another, or (2) as an additional information display. An example of the latter use may be showing the personal history of an individual for either strategies used, payoffs earned, or both.

6.2. Relevance

The *N*-person conversion of five archetypical *2x2* games has been examined to devise an abstract framework for these games. With this abstract framework an attempt has been made to (a) define the underlying social phenomena of these games, and (b) discuss the properties and constraints necessary to ascertain them for these games. This was done in an effort to gain a deeper understanding regarding the games themselves, gain insight into the effects and necessary requirements when converting *2x2* games to their respective *N*-person variants in general, and serve as foundational work to implement such a framework in an interactive environment for educational or experimental purposes.

It has been shown that all of the games discussed here may be transformed linearly and in accordance with their respective underlying social phenomena and strategic pressures, albeit spread out across individuals. Furthermore, this linear transformation may serve as a general manner to convert classical 2x2 games to their *N*-person variants, which allows for meaningful comparison between future research regarding *N*-person games. Furthermore, the functions proposed in this work may serve as tools to implement these archetypical games in an interactive game environment as described in the previous section. This may be done in a straightforward manner by using the classical 2x2 games and allows for an arbitrary number of players (although an even number is necessary). Furthermore, well-defined functions allow for careful analysis of MAL strategies (Shoham, Powers & Grenager, 2003). This is because in a MAL setting there is no longer an optimal strategy, but rather a strategy space that depends on others' strategies. The classical approach in game theory is to ask the question: "What is the best—most rational—thing an agent can do?" (Parsons & Wooldridge, 2002). In order to make such a choice, an agent must take into account the decisions of others, and must assume that they will act so as to optimise their own outcome. It is inevitable then, that – in order to reason about others' strategies – payoff functions should be well-defined.

Another point of relevance is the fact that the graphs may serve as useful illustrative tools in the implementation of an interactive environment. This allows for a relatively trivial representation, which may be used to depict the payoffs granted to each individual for each respective game. It is noteworthy that this also grants a consistent manner of illustration, since both the *2x2* and the *N*-person variants of any of the games discussed here may be represented by such a format. Moreover, in this manner it is no longer necessary to represent the games in complex multidimensional matrices in order to play these games. This is of course desired when this framework is implemented in an interactive game environment as discussed in the previous section.

The definition of the underlying social phenomena may be used to gain a deeper understanding with regard to the archetypical 2x2 games and the *N*-person conversions thereof. As such, it may serve as clarification on *N*-person conversion in general but also as a basis of argumentation for useful strategies of either humans or (AI) agents. For example, they may serve as basis of argumentation to move away from the classical analysis tools of game theory such as Nash equilibria and its variants (Shoham, 2008). In fact, it is already argued in this work that equilibria play a less important role in *N*-person scenarios. Rather, strategies like minimax and maximin are suggested to be more relevant in *N*-person scenarios as basic tools of analysis. Taking it even a step further, the vast majority of the effort on traditional AI problems is spent on topics such as designing a good heuristic function, searching, and planning (Shoham, 2008). Designing a heuristic function in particular, but also the planning of actions, may be served by a fundamental basis of argumentation as granted by the definitions of the underlying social phenomena.

Furthermore, both the general and *N*-person specific properties, and the usefulness of these may be clarified through the implementation of this framework. As such, it grants individuals playing the game a deeper understanding with regard to these properties. For example – in line with the previous paragraph – that the properties as defined for the *2x2* games (*e.g.* the availability of a dominant strategy) become less relevant in *N*-person variants thereof, since the pressures are spread out across all individuals. Furthermore, in a multi-agent setting one cannot separate learning from teaching because of this same spreading of pressures (Shoham, Powers & Grenager, 2007). In other words, both the usefulness and complexity of reasoning about others' actions increases as the number of players increase, because the individual payoffs become more dependent on others' actions. In turn, this increases the need for heuristics that save on computing time but still take into account the interactions between multiple players. The implementation of a framework as discussed in this work may shed light on such considerations by examining human behaviour in both educational and experimental settings. In addition, the properties still play an important prescriptive role, such that the pressures – albeit spread out across individuals – remain similar for *N*-person games and their *2x2* counterparts.

6.3. Related work

As mentioned in previous sections, the linear conversion of these games has been a choice, not a necessary requirement. There has been research regarding these (and other) classical 2x2 games, where a threshold is modelled as opposed to a full-scale linear conversion. For example Pachecco et al. (2009) modelled an *N*-person SH game with a threshold and costs and benefits an individual may have. Each cooperator then shares a fixed amount of costs and when a specific threshold is met all individuals share a fixed benefit. As such, it is assumed an individual is able to free-ride when this threshold is met, which in turn creates an *N*-PD like game. While this game is drastically different than the *N*-person SH game mentioned here, it yields interesting results and properties of its own. Furthermore, it would be interesting to research how the other games discussed here would behave when modelled with a threshold.

In addition research regarding another approach to model *N*-person games altogether has been conducted as well. For example Hamburger (1974) discusses the "take some" format. This format allows for a 2x2 matrix representation of *N*-person games, even when more than two strategies are available to players. This format is defined as: "each player (simultaneously) picks one number from a set of positive numbers. The sum, S, of the numbers picked is compared to a number L determined by the outcome of a random event independent of the players' choices. Players know the distribution of L, but not the outcome. If S < L, each player's payoff is the number he picked. Otherwise, everyone gets payoff 0". As such, it models a sort of threshold similar to the one used in Pachecco et al. (2009). While this format does not allow for games that do not have a most-threatening strategy, it is a useful manner of modelling since it allows for a relatively straightforward representation of *N*-person games. This is of course desired when implementing such a game in an interactive environment. However, as discussed in the previous section, it is argued here that the use of graphs may serve this purpose equally well. The addition of more than two strategies in an *N*-person game does make it complex to illustrate it in a graph (since it would add a third dimension to such a graph), while it does not matter for the take some format. As such, this format may be alluring for games with multiple players and multiple strategies.

Moreover, research exists with regard to games that yield a threshold as an inherent property. For example Whitehead (2008) discussed the "El Farol bar problem", where a scenario is described of a choice between going to the bar and staying at home. This game is typically used to depict a scenario with relatively large numbers of players (*e.g.* $N = 100$), in a straightforward manner. In this scenario going to the bar yields the best outcome if less than 60% of the other players go to the bar, otherwise it is the worst outcome. The outcome of staying at home is irrespective of others' strategies and yields the second best outcome. As such, the game may be modelled as a *2x2* matrix as well, regardless of the number of players. This may also serve as a general manner in which thresholds can be modelled, and may be used for the games described here. This might prove to be interesting for future research.

6.4. Discussion

As mentioned in the foregoing, the conversion to multiple players may be done in a linear fashion or by modelling a certain threshold. It has been argued here that the linear transformation of the games more closely resemble their real-world scenarios. Accordingly, the strategies that agents may use are dependent equally on each individual, rather than a singular threshold above or below which the dilemma fails or succeeds, which would trivialise the respective games. While a threshold may be interesting to look into as well, this work is limited to full-scale linear conversions. Such thresholds have not been included because it goes beyond the scope of this work. However, it may prove interesting for future work.

Furthermore, it has been shown that there remain important choices to be made with regard to the exact implementation of such a linear transformation. Specifically, the starting points of the functions for both strategies and the difference between them may be modelled in different manners, which may result in drastically different games, as represented by the underlying strategic pressures. In this work the functions for the social dilemma games are modelled such that their respective 2x2 matrices serve as starting points. For the other games the functions have been transformed such that the four points around the median number of players represent the four payoff preferences as modelled by their respective 2x2 matrices. In addition, the functions are modelled in such a manner that it has been ensured no negative payoffs may be achieved. An unintended shortcoming of these functions is that they require an even number of players for the coordination and trust dilemma games, since the baselines of these are modelled to be at the median number of players. A manner to circumvent this could be to always round the number down in such a case, but it would skew the pressures toward cooperation for the trust dilemma game and toward one counter-strategy in the coordination dilemma games. Another choice that has been made in this work concerns the difference between the baseline points. These are modelled to represent the preference ordering of the underlying 2x2 game. It could be argued however, that this difference should also scale with the number of players. This is because in for example the BoS game, defection is only better at one outcome pair ($n = \frac{N}{2}$), regardless of the number of players, as opposed to one out of three (roughly 66%) of total outcome pairs – which is the case in the 2x2 game. As such, the cooperative strategy is relatively lower in risk in the *N*-person variant than it is in the *2x2* game, and even more so as the total number of players increases. This is not necessarily harmful to the game, but it might be interesting to model it such that it scales with the total number of players.

Moreover, the essence for each of the dilemma games – as represented in this work – may be summarised by the following properties: (a) the social dilemma games have either a dominant strategy, a most-threatening strategy, or both, which result in a deficient outcome; (b) the coordination dilemmas have a threatening counter-strategy or a dominant counter-strategy, while only the dominant counter-strategy coincides with the equilibrium point, and; (c) the trust dilemma game has a most-threatening strategy which does not coincide with the stable equilibrium. In addition, the availability of a dominant (counter-) strategy has been linked to the stability of the equilibria for these games. As mentioned in the *Relevance* section, while these properties may serve as clear-cut definitions of pressures in 2x2 games, they are less relevant in linearly converted *N*-person scenarios. As such, it may be more fruitful to speak of generalised pressures in such games (*e.g.* the pressure to coordinate or cooperate, as a function of the number of same strategies chosen or the number of cooperators, respectively).

Furthermore, these generalised pressures may be used to come up with useful heuristics in MAL settings. These may, for example, generalise better than (Nash) equilibrium strategies since not every game has a dominant strategy available, while they do have well-defined functions. In addition, the use of generalised pressures for heuristics may help explain the gap often seen between (AI) agent and human rationale. For example, creating alliances to escape the Nash equilibrium in the PD is not trivial for agents, because agents have no control over others' behaviours and as such, are usually assumed to be strictly rational in game theoretical settings (Panait & Luke, 2005). However, humans rarely use strictly rational strategies in these scenarios. An area of MAL research under the heading of "bounded rationality" focuses on bridging this gap between agent and human rationale in games (Shoham, 2008). For example, a promising result is that – in the finitely repeated PD game –constant defection is no longer the only subgame-perfect equilibrium (*i.e.* strictly rational strategy) if the players are finite automata (*i.e.* rationally bounded agents) with sufficiently few states (Neyman, 1985). Such results could be used to argue that bounded agents more closely resemble their human counterparts, which often do not converge to a subgame perfect equilibrium indeed. In short, this framework may be used to serve as an explanatory foundation for the use of heuristics to bridge the gap between agent and human rationale, which is a desired and meaningful illustration for educational and experimental settings alike.

With regard to *N*-person specific properties, it has been shown that only common interval symmetry is essential since it imposes linearity on the functions. Separability is useful since it further improves the consistency of the pressures, but it is not a necessary property. In addition, the property of a non-oscillatory solution has been shown to be relatively uninteresting and certainly not necessary. The *N*-person specific properties discussed in this work are not complete, nor do they aim to be. Other interesting properties may be defined when for example a threshold is modelled. However, the properties discussed here are generalizable to any linearly converted *N*-person game and may form a foundation for future work.

6.3 Conclusion

In this thesis five archetypical *2x2* games and their linearly converted *N*-person variants have been discussed. First, the scenarios of these games and their variations have been elaborated upon to serve as an illustration. Second, the underlying social phenomena and the strategic pressures reflecting these have been discussed for *2x2* games. Furthermore, the choice was made to linearly convert these *2x2* games to their *N*-person counterparts. In doing so, functions have been adapted to allow for a relatively straightforward conversion from the *2x2* games to their *N*-person variants. As such, the strategic pressures for *N*-person counterparts have been discussed and compared to their *2x2* variants in light of their respective underlying social phenomena. Consecutively, the essential properties of both the *2x2* games and their *N*-person counterparts have been identified with respect to the strategic pressures they represent. In turn, this work forms an abstract framework which may serve as a foundation for an implementation within an interactive environment for educational or experimental purposes.

References

Arthur, W. B. (1994). Inductive reasoning and bounded rationality. *The American economic review*, *84*(2), 406-411.

Bishop, R. L. (1964). A Zeuthen-Hicks theory of bargaining. *Econometrica: Journal of the Econometric Society*, 410-417.

Bornstein, G., Budescu, D., & Zamir, S. (1997). Cooperation in intergroup, N-person, and two-person games of chicken. *Journal of Conflict Resolution*, *41*(3), 384-406.

Brams, S. J., & Kilgour, D. M. (2009). How democracy resolves conflict in difficult games. In *Games, Groups, and the Global Good* (pp. 229-241). Springer, Berlin, Heidelberg.

Challet, D., Marsili, M., & Ottino, G. (2004). Shedding light on el farol. *Physica A:Statistical Mechanics and Its Applications*, *332*, 469-482.

Dawes, R. M. (1980). Social dilemmas. *Annual review of psychology*, *31*(1), 169-193.

Doebeli, M., & Hauert, C. (2005). Models of cooperation based on the Prisoner's Dilemma and the Snowdrift game. *Ecology letters*, *8*(7), 748-766.

Friedman, J. W. (1983). On characterizing equilibrium points in two person strictly competitive games. *International Journal of Game Theory*, *12*(4), 245-247.

Gneezy, U., Haruvy, E., & Yafe, H. (2004). The inefficiency of splitting the bill. *The Economic Journal*, *114*(495), 265-280.

Guyer, M., & Rapoport, A. (1966). A taxonomy of 2 x 2 games. *General Systems*, *11*, 203-214.

Guyer, M., & Rapoport, A. (1970). Threat in a two-person game. *Journal of Experimental Social Psychology*, *6*(1), 11-25.

Hamburger, H. (1969). Separable games. *Systems Research and Behavioral Science, 14*(2), 121-132.

Hamburger, H. (1973). N-person Prisoner's Dilemma. *Journal of Mathematical Sociology*, *3*(1), 27-48.

Hamburger, H. (1974). Take Some—A format and family of games. *Systems Research and Behavioral Science*, *19*(1), 28-34.

Hammerstein, P., & Selten, R. (1994). Game theory and evolutionary biology. *Handbook of game theory with economic applications*, *2*, 929-993.

Hardin, R. (1971). Collective action as an agreeable n-prisoners' dilemma. *Systems Research and Behavioral Science*, *16*(5), 472-481.

Kilgour, D. M., & Fraser, N. M. (1988). A taxonomy of all ordinal 2× 2 games. *Theory and Decision*, *24*(2), 99-117.

Komorita, S. S. (1976). A model of the N-person dilemma-type game. *Journal of Experimental Social Psychology*, *12*(4), 357-373.

Ledyard, J. (1997). *Public goods: A survey of experimental research* (No. 509). David K. Levine.

Liebrand, W. B. (1983). A classification of social dilemma games. *Simulation & Games*, *14*(2), 123-138.

Meux, E. P. (1973). Concern for the common good in an N-person game. *Journal of Personality and Social Psychology*, *28*(3), 414.

Neyman, A. (1985). Bounded complexity justifies cooperation in the finitely repeated prisoners' dilemma. *Economics letters*, *19*(3), 227-229.

O'Donovan, J., Jones, R. E., Marusich, L. R., Teng, Y., Gonzalez, C., & Höllerer, T. (2013). A Model-based Evaluation of Trust and Situation Awareness in the Diner's Dilemma Game. In *Proceedings of the 22th Behavior Representation in Modeling & Simulation (BRIMS) Conference* (pp. 11-14).

Rapoport, A. (1967). Exploiter, Leader, Hero, and Martyr: the four archetypes of the 2×2 game. *Systems Research and Behavioral Science*, *12*(2), 81-84.

Pacheco, J. M., Santos, F. C., Souza, M. O., & Skyrms, B. (2009). Evolutionary dynamics of collective action in N-person stag hunt dilemmas. *Proceedings of the Royal Society of London B: Biological Sciences*, *276*(1655), 315-321.

Panait, L., & Luke, S. (2005). Cooperative multi-agent learning: The state of the art. *Autonomous agents and multi-agent systems*, *11*(3), 387-434.

Papadimitriou, C. H., & Roughgarden, T. (2005, January). Computing equilibria in multi-player games. In *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms* (pp. 82-91). Society for Industrial and Applied Mathematics.

Parsons, S., & Wooldridge, M. (2002). Game theory and decision theory in multi-agent systems. *Autonomous Agents and Multi-Agent Systems*, *5*(3), 243-254.

Santos, M. D., Pinheiro, F. L., Santos, F. C., & Pacheco, J. M. (2012). Dynamics of N-person snowdrift games in structured populations. *Journal of Theoretical Biology*, *315*, 81-86.

Scarf, H. E. (1967). The core of an N person game. *Econometrica: Journal of the Econometric Society*, 50-69.

Shoham, Y. (2008). Computer science and game theory. *Communications of the ACM*, *51*(8), 74 -79.

Shoham, Y., Powers, R., & Grenager, T. (2003). *Multi-agent reinforcement learning: a critical survey*. Technical report, Stanford University.

Shoham, Y., Powers, R., & Grenager, T. (2007). If multi-agent learning is the answer, what is the question?. *Artificial Intelligence*, *171*(7), 365-377.

Skyrms, B. (2004). *The stag hunt and the evolution of social structure*. Cambridge University Press.

Smith, J. M. (1980). Evolutionary *game* theory. In *Vito Volterra symposium on mathematical models in biology* (pp. 73-81). Springer, Berlin, Heidelberg.

Snidal, D. (1985). Coordination versus prisoners' dilemma: Implications for international cooperation and regimes. *American Political Science Review*, *79*(4), 923-942.

Souza, M. O., Pacheco, J. M., & Santos, F. C. (2009). Evolution of cooperation under N-person snowdrift games. *Journal of Theoretical Biology*, *260*(4), 581-588.

Szolnoki, A., & Perc, M. (2010). Reward and cooperation in the spatial public goods game. *EPL (Europhysics Letters)*, *92*(3), 38003.

Taylor, M., & Ward, H. (1982). CHICKENS, WHALES, AND LUMPY GOODS: ALTERNATIVE MODELS OF PUBLIC-GOODS PROVISION. *Political Studies*, *30*(3), 350-370.

Wang, X. H., & Yang, B. Z. (2003). Classification of 2X2 games and strategic business behavior. *The American Economist*, *47*(2), 78-85.

Weil, R. L. (1966). The N-person Prisoner's dilemma: Some theory and a computer-oriented approach. *Systems Research and Behavioral Science*, *11*(3), 227-234.

Whitehead, D. (2008). The El Farol bar problem revisited: Reinforcement learning in a potential game. *ESE discussion papers*, *186*.