Utrecht University

Thesis to complete the Complex Systems profile

A Response Strategy for the Battle of the Sexes Game with Intrinsic Correlations

Dated: November 18, 2018

Author:
K.J.H. Peters

Supervisors:
Prof. dr. ir. H.T.C. Stoof
Prof. dr. ir. V.W. Buskens

Abstract

In this Thesis we apply a recently proposed model by Correia *et al.*[1] to the Battle of the Sexes game. This model considers games with intrinsic correlations between the final possible outcomes, and lets players respond to these correlations with a response strategy. We studied all possible Nash equilibria in these response strategies, which we found to be 9 in total, including equilibria corresponding to the previously-known pure and mixed-strategy Nash equilibria. We finally computed which response-strategy profile leads to the optimal payoff for the players. We found that for a large part of the space of correlation probabilities the equilibrium corresponding to the pure-strategy Nash equilibrium is still optimal, since the average payoff of the other Nash equilibria is often less than one would obtain from consistently choosing one's least-preferred pure-strategy option.

We furthermore study this model when applied to a simple 1-dimensional network, i.e. a ring, by incorporating the exact solution of the 1-dimensional Ising model. We find that the external field that players experience is always the same logarithmic function of the correlation probabilities, but the way players react to this field differs for the various Nash equilibria. The interaction strength between the players, which is mostly of the ferromagnetic type, is a different function for dominant diagonal correlations and dominant off-diagonal correlations. However, the change between these two seems to be continuous, but we lack an analytical solution for part of the probability space of the correlations. We finally find a magnetization which has the same dependence on the correlation probabilities for all the Nash equalibria.

Acknowledgements

I would like to thank Prof. Dr. Henk Stoof and Prof. Dr. Vincent Buskens for letting me work on this interesting subject. Although it was only for 10 weeks, I really enjoyed working in this (interdisciplinary) field of research, and learned a lot about game theory in general and recent developments in the field. I'm also grateful for them being always willing to help and provide new insights to tackle this research.

I would furthermore like to thank Adriana Correia and Joris Broere for discussions during the research. Especially the weekly meetings with Henk and Adriana, and later Joris as well, were a great help to reach my final goals.

Contents

1 Introduction

Consider the following scenario. You planned a night out with a friend and agreed to go the movies. You know two movies would start at 10 pm, but you hadn't decided yet on which movie to see; you would decide this at the movie theater. Your friend arrives at 9:45, but you are running late yourself, because of a flat tire on your bike. Your phone is also out of charge, so you cannot reach each other. At 10 pm your friend decides to buy a ticket. He knows you would rather see a comedy, while he himself is more a fan of action movies.

You arrive a couple minutes after 10 pm and notice there are still tickets available for both movies. You would rather see a movie together then alone, so you have to guess which movie your friend went to. Since you only have money to buy one ticket, buying tickets for both movies is no option. Which movie do you pick?

The situation sketched is an example of what is known as the Battle of the Sexes in game theory; a game wherein two players are trying to coordinate on the same choice, but with no means of communication. Traditionally, the story behind this game is that of a man and a woman who would meet for an activity. However, they both cannot recall whether they would go to the movies or to the opera. The man prefers to go the movies, while the woman would like to attend the opera. Of course, they would rather go together than alone, hence the coordination problem. Now, assuming they cannot communicate each other's preferred choice, where should they go?

It quickly becomes clear that the Nash equilibria of this game are unfair, since one player is better off than the other. On the other hand, a mixed-strategy equilibrium is suboptimal since it results in an average payoff which is less than consistently going to one's least-preferred choice. One way to deal with such problems is by introducing a correlation device; a third party which sends information to the players about which choice they should make. Recently, Correia *et al.*[1] presented a model which makes these correlations part of the game itself, i.e. a game with intrinsic correlations between final outcomes. Players should then respond to these correlations to find desirable equilibra. It is this model that we will apply to the Battle of the Sexes and study extensively in this Thesis.

Broere *et al.*[2] furthermore showed, using simulations, that when the (uncorrelated) Battle of the Sexes game is played on a network, global behavior is strongly dependent on the network structure. They showed that networks with high clustering lead to uniform behavior within clusters, but heterogeneous between clusters. As a step towards analytical results for correlated games on networks, we will use methods from statistical physics to model the Battle of the Sexes on a simple 1-dimensional network, i.e. a ring.

The structure of this thesis is as follows. The theory in Section 2 elaborates on the Battle of the Sexes game itself, as well as introduces game theoretical concepts and the notation we will use. We will also formally introduce the notion of correlated games and present the model by Correia *et al.* which lets players respond to the intrinsic correlations to improve their payoff. Furthermore, we present the (bipartite) Ising model and its exact solution in 1 dimension, such that we can apply methods of statistical physics to extend our model to simple 1-dimensional networks.

The model for correlated games is then analyzed analytically in great detail in Sec. 3; we provide a discussion of all possible equilibria in the probability space of the correlations, as well as which response strategies yield the highest payoff, given certain correlations.

In Section 4 we connect the previously found results to the Ising model by renormalizing the probabilities each final state is reached. This provides us with a description more in line with statistical physics, such that we can extend our model to games on networks.

To conclude, we will summarize our most important results in Section 5 and put everything

1 \ 2	A	B
A	$1, S$	t, t
B	$0, 0$	$S, 1$

Table 1: Payoff table for two-player BoS, with $0 < t < S < 1$.

1 \ 2	A	B
A	$1, S$	$0, 0$
B	$0, 0$	$S, 1$

Table 2: Payoff table for two-player BoS, with $0 < S < 1$.

into context. We describe the problems this model solves and how it can be improved or extended for future work.

2 Theoretical background

In this section we will first specify the Battle of the Sexes (BoS) game and introduce the conventional terminology and notation of game theory. We will also elaborate on the notion of correlated games[1] and present a model that lets players respond to intrinsic correlations of a game. Next we will explain the Ising model, giving a brief (conceptual) introduction to spin angular momentum and providing the solution of the Ising model in 1 dimension. We adjust the model slightly to accommodate for distinguishable particles on a bipartite lattice, such that we can later make the connection to the BoS.

2.1 Game Theory: Battle of the Sexes

The situation sketched in the introduction is one of the problems game theory tries to solve. It can be tackled in various ways, some of which we will look into now.

2.1.1 Pure strategies

In general, we can label the two options the players have as A and B. Furthermore, one can define a payoff table, which tells exactly what each player would get after both players have made their choice. For the BoS, a general payoff table is given in Table 1. Note that since $0 < S < 1$, player 1 has the highest payoff when both players choose A, while player 2 receives the highest when both players choose B. Additionally, since $0 < t < S$, it is not only unfavorable to miscoordinate, i.e. both choosing something different, but it is even worse if both players pick their least-preferred option.

A simplified version is shown in Table 2. Here, there is no longer the distinction in payoff between choosing (A, B) or (B, A), removing this asymmetry from the payoff table. From now on, if we refer to the Battle of the Sexes' payoffs, it will always be this table we are referring to.

As can easily be seen in Table 2, if both players choose A, no one has an incentive to deviate from this choice, since if player 1 changes from A to B, his payoff will decrease from 1 to 0, and when player 2 changes from A to B, his wil decrease from S to 0. Such a set of choices is known as a *Nash equilibrium.* For the two-player BoS, both (A, A) and (B, B) are Nash equilibria.

Formally, a player i belongs to the set of players $1, 2, ..., N$ and the options player i has belong to the pure-strategy space S_i. We can then define the strategy profile $s = \{s_1, s_2, ..., s_I\}$, with $s_i \in S_i$ the *pure strategy* of player i. A pure-strategy profile is then a Nash equilibrium if no single player can benefit from a unilateral deviation in strategy, i.e. s^* is a Nash equilibrium[3] when for all players i

$$\forall s_i \in S_i : u^{(i)}(s_i^*, s_{-i}^*) \geq u^{(i)}(s_i, s_{-i}^*), \tag{1}$$

with $u^{(i)}(s_i, s_{-i})$ the payoff obtained by player i for strategy profile $s = \{s_i, s_{-i}\}$. Here we use $-i$ to indicate all players excluding player i, such that $s_{-i} = \{s_1, s_2, ..., s_{i-1}, s_{i+1}, ..., s_N\}$.

Note that in the Battle of the Sexes, although (A, A) and (B, B) are both Nash equilibria, due to the difference in preference of the players, there will always be one player better off than the other. As such, these equilibria can be regarded as unfair.

2.1.2 Mixed strategies

In a *mixed strategy* probabilities are assigned to each pure strategy. A pure strategy can in that sense be seen as a mixed strategy, wherein this pure strategy has probability 1 to be chosen, while all other pure strategies are chosen with probability 0.

Similar as with the pure strategy, a mixed-strategy profile σ^* is a Nash equilibrium if for all players i

$$\forall s_i \in S_i : u^{(i)}(\sigma_i^*, \sigma_{-i}^*) \geq u^{(i)}(s_i, \sigma_{-i}^*), \tag{2}$$

where

$$u^{(i)}(\sigma) = \sum_{s \in S} \left(\prod_{j=1}^{N} \sigma_j(s_j) \right) u^{(i)}(s)$$

is the expected payoff of player i when playing with strategy profile σ. It is called a mixed-strategy Nash equilibrium when at least 1 player, thus not necessarily all, uses a mixed strategy. Applying this to the Battle of the Sexes, the conditions for a Nash equilibrium are given by

$$\begin{cases} u^{(i)}(\sigma_i^*, \sigma_{-i}^*) \geq u^{(i)}(A, \sigma_{-i}^*) \\ u^{(i)}(\sigma_i^*, \sigma_{-i}^*) \geq u^{(i)}(B, \sigma_{-i}^*) \end{cases},$$

with $i \in \{1, 2\}$ in a two-player game.

Assuming player i plays μ with probability $P_\mu^{(i)}$, the payoffs received by player 1 when playing either A or B are respectively given by

$$\begin{aligned} u^{(1)}(A) &= P_A^{(2)} \cdot 1 + P_B^{(2)} \cdot 0 = P_A^{(2)}, \\ u^{(1)}(B) &= P_A^{(2)} \cdot 0 + P_B^{(2)} \cdot S = (1 - P_A^{(2)})S, \end{aligned}$$

where we used $P_A^{(i)} + P_B^{(i)} = 1$ in the second line. By requiring these payoffs to be equal, we find the probability $P_A^{(2)*}$ for which player 1 remains indifferent. This probability is given by

$$P_A^{(2)*} = \frac{S}{1 + S}. \tag{3}$$

Similarly, by requiring $u^{(2)}(A) = u^{(2)}(B)$, we find

$$P_A^{(1)*} = \frac{1}{1 + S}. \tag{4}$$

The probabilities $P_A^{(1)*}$ and $P_A^{(2)*}$ now form a mixed-strategy Nash equilibrium for the BoS.

At the mixed-strategy equilibrium, the probabilities to reach the final state $\mu\nu$ are given by $P_{\mu\nu}^* = P_\mu^{(1)*} P_\nu^{(2)*}$. By using $P_B^{(i)} = 1 - P_A^{(i)}$, we find the probability matrix

$$P^* = \begin{pmatrix} \frac{S}{(1+S)^2} & \frac{1}{(1+S)^2} \\ \frac{S^2}{(1+S)^2} & \frac{S}{(1+S)^2} \end{pmatrix},$$

which is asymmetric due to the different preferences of the players. Using this matrix, the average payoffs the players obtain are

$$\left\langle u^{(1)*}\right\rangle = \sum_{\mu\nu} P^{*}_{\mu\nu} u^{(1)}_{\mu\nu} = \frac{S}{1+S},$$
$$\left\langle u^{(2)*}\right\rangle = \sum_{\mu\nu} P^{*}_{\mu\nu} u^{(2)}_{\mu\nu} = \frac{S}{1+S}.$$

Hence both players obtain the same average payoff at the mixed-strategy Nash equilibrium. However, as $S/(1+S) < S$, the payoff both players receive is even less than choosing one's less-favored option consistently. This is due to the fact that players will miscoordinate with nonzero probability. We will now look into a way to avoid miscoordination by introducing correlations.

2.1.3 Correlated strategies

By considering the (expected) payoffs players obtain, one can see that the Battle of the Sexes presents an interesting case for game theory. On the one hand, the pure strategy Nash equilibria are unfair, in the sense that one player is always better off than the other. On the other hand, although fair, the mixed-strategy Nash equilibrium yields a suboptimal payoff for both players. One can avoid this problem by introducing a *correlation device*; a trusted third party, telling both agents what to choose by drawing final states with a certain probability.

The idea is that players might gain if they could build a correlation device that sent signals to both players. In our game of Battle of the Sexes, it would be beneficial if a third party would flip a coin to determine the outcome of the game. For example, if it's heads, the third party would inform both players to choose A, whereas if it's tails, the third party would inform them to choose B. This way, while still not being able to communicate with each other, players would always coordinate on the same choice, thus the expected payoff is higher than one would obtain at the mixed-strategy Nash equilibrium. In the symmetric case, both players would now receive a payoff of $1/2 + S/2 = (S+1)/2 > S$ on average.

Using formal game-theoretical language, a correlation device is identified by a triple[3] $(\Omega, \{H_i\}, p)$. Here, Ω is the (finite) space of all possible outcomes of the correlation device and p is a probability distribution on the state space Ω. The information player i has, regarding which $\omega \in \Omega$ is the outcome, is represented by the information partition H_i. Suppose the true state is ω, then player i has the information that it lies in $h_i(\omega)$. Here, the set $h_i(\omega) \in H_i$ consists of the states that player i deems as possible outcomes when the true state is ω. Note that $\omega \in h_i(\omega)$ as well, such that the true state is always part of the set of states seen as possible by player i.

Applying this formalism to our example of the Battle of the Sexes, we denote the outcomes of the correlation device by H for heads and T for tails. Then the state space is $\Omega = \{H, T\}$, the probability measure $p(\omega) = 1/2$ is constant, as both states are drawn with equal probability, and the information partition is $H_i = \{\{H\}, \{T\}\}$ for both players.

Now, given a correlation device $(\Omega, \{H_i\}, p)$, we need to define strategies that let the players base their play on the information provided to them by the correlation device. We define *correlated strategies* $\mathscr{s}_i$ as maps from Ω to the set of pure strategies S_i with $\mathscr{s}_i(\omega) = \mathscr{s}_i(\omega')$ if $\omega' \in h_i(\omega)$. The latter condition implies that if $\omega' \in h_i(\omega)$, then the actions of player i are the same for the states ω and ω'. Returning to the BoS game, we would have for both players the maximally correlated strategy $\mathscr{s}_i(H) = A$ and $\mathscr{s}_i(T) = B$, since we want both of them to choose A when the result is heads, and B when it is tails.

The correlated strategy profile $\mathscr{S}^* = \{\mathscr{S}_1^*, \mathscr{S}_2^*, ..., \mathscr{S}_N^*\}$ is a *correlated equilibrium* if for all players i

$$\forall \mathscr{S}_i : \sum_{\omega \in \Omega} p(\omega) u^{(i)}(\mathscr{S}_i^*(\omega), \mathscr{S}_{-i}^*(\omega)) \geq \sum_{\omega \in \Omega} p(\omega) u^{(i)}(\mathscr{S}_i(\omega), \mathscr{S}_{-i}^*(\omega)) \tag{5}$$

Hence, no single player should be able to benefit by changing $\mathscr{S}_i$ while all other players' correlated strategies are kept fixed, similar to what we have seen in the case of the pure and mixed-strategy Nash equilibria.

This condition is easily checked in our BoS game where a third party flips a coin to determine the final outcome of the game. Suppose player 1 is told to play A. Then, by construction, player 2 is also told to play A. Therefore, player 1 cannot benefit from playing B instead, as this would decrease his payoff from 1 to 0. Similarly, player 2 doesn't want to deviate, since this would decrease his payoff as well, from S to 0. The same arguments hold when player 1 receives the information to play B. Hence we have a correlated equilibrium.

Note that the correlated equilibrium discussed here is not unique, since there may be other probability distributions for which both players will always follow the correlation device. In fact, it is easy to check that all correlation devices that assign nonzero probability to (A, A) and/or (B, B), but zero probability to the other states, result in a correlated equilibrium. However, in general, a correlation device might also assign nonzero probabilities to unfavorable outcomes. In that case one should question whether it is beneficial to always obey the instructions of the correlation device. This will be discussed in great detail in Sec. 2.1.4.

2.1.4 Response strategies for correlated games

As opposed to the game with correlated strategies that was presented above, we will now make the correlation device part of the game itself in the following sense. Players no longer get to construct a correlation device they would always want to obey, but we now have a game with *intrinsic correlations* to which the players can respond. Changing the correlation device will define a new game, to which players might respond in a different way. This type of game will be referred to as a *correlated game*.[1] As a concrete example, one can think of a traffic light at an intersection; the traffic light is simply there and introduces correlations, but incoming traffic has no way to influence the traffic lights (the correlation device) and has to respond to the color of light shown (the correlations).

Consider the correlated Battle of the Sexes, i.e. the BoS game with a correlation device which assigns initial probabilities $p_{\mu\nu}$ to draw a state $\mu\nu$, with $\mu, \nu \in \{A, B\}$. In the traditional sense of a correlated equilibrium, players would always obey the instruction of the correlation device and turn to mixed-strategy equilibrium otherwise. Now, however, player i will respond to the initial correlations by either following the instruction to choose μ, with *response probability* $P_{F_\mu}^{(i)}$, or not follow this instruction with probability $P_{NF_\mu}^{(i)} = 1 - P_{F_\mu}^{(i)}$.

The decisions players make on whether or not to obey the instructions of the correlation device, represented by the response probabilities, are the new actions players take, whilst still not being able to communicate. These response probabilities affect the likelihood of each final outcome of the game. Hence, this leads to *renormalized probabilities* $p_{\mu\nu}^R$ given by

$$p_{\mu\nu}^R = \sum_{\mu'\nu'} P_{\mu \leftarrow \mu'}^{(1)} P_{\nu \leftarrow \nu'}^{(2)} p_{\mu'\nu'}, \tag{6}$$

with $P_{\mu \leftarrow \mu'}^{(i)}$ the probability that player i is told to play μ' by the correlation device, but plays μ instead. In terms of the response probabilities, these can be written as

$$P_{\mu \leftarrow \mu'}^{(i)} = \delta_{\mu\mu'} P_{F'_\mu}^{(i)} + (1 - \delta_{\mu\mu'}) P_{NF'_\mu}^{(i)}.$$

The expected payoff is now given by the payoffs averaged over the renormalized probabilities, i.e.

$$\left\langle u^{(i)} \right\rangle = \sum_{\mu\nu} u^{(i)}_{\mu\nu} p^R_{\mu\nu} \equiv C^{(i)}_A P^{(i)}_{F_A} + C^{(i)}_B P^{(i)}_{F_B} + C^{(i)}_C. \tag{7}$$

The coefficients $C^{(i)}_\mu$ depend linearly on the initial correlation probabilities, $p_{\mu\nu}$, and on the response probabilities of the other player, $P^{(-i)}_{F_\mu}$. The exact form of these coefficients is given in App. A.

An equilibrium for such a correlated game is reached when none of the players can improve their expected payoff by changing their own response probabilities, while keeping the other players' probabilities fixed. If we consider Eqn. 7, then a Nash equilibrium is achieved by requiring that for both players the coefficients $C^{(i)}_A$ and $C^{(i)}_B$ are zero, unless the equilibrium response probability $P^{(i)*}_{F_\mu}$ is either 0 or 1, in which case the corresponding coefficient $C^{(i)}_\mu$ should be consistently nonzero.[1]

More formally, we can define a Nash equilibrium in the response strategy in a similar way as for the correlated strategy in Sec. 2.1.3. A Nash equilibrium is reached when no player i has an incentive to change $P^{(i)}_{F_\mu}$, while the response probabilities of the other players, $P^{(-i)}_{F_\mu}$, remain fixed. Thus, if for all players i and $\forall \mu' \in \{A, B\}$

$$\sum_{\mu\nu\nu'} u^{(i)}_{\mu\nu} P^{(i)*}_{\mu\leftarrow\mu'} P^{(-i)*}_{\nu\leftarrow\nu'} p_{\mu'\nu'} \geq \sum_{\mu\nu\nu'} u^{(i)}_{\mu\nu} P^{(i)}_{\mu\leftarrow\mu'} P^{(-i)*}_{\nu\leftarrow\nu'} p_{\mu'\nu'}, \tag{8}$$

then the response strategies form a Nash equilibrium for the correlated game.

Finally, as we now need to work with two independent probabilities per player, we will introduce the notation

$$\Pi^{(i)}_{P_{F_A},P_{F_B}} = \left(P^{(i)}_{F_A}, P^{(i)}_{F_B}\right), \tag{9}$$

which completely defines the response strategy of player i. A response strategy profile can now be written compactly as $\Pi = \{\Pi^{(1)}_{P_{F_A},P_{F_B}}, ..., \Pi^{(N)}_{P_{F_A},P_{F_B}}\}$.

2.2 Spin systems

In physics, we know of two types of angular momentum:[4] orbital, which we associate with the motion of the center of mass around a central axis, and spin, which we associate with motion of an object about the center of mass itself. Classically, however, spin can be seen as the net sum of orbital angular momenta of all particles an object consists of, such that the distinction between orbital and spin angular momentum is merely a matter of semantics. However, in quantum mechanics this distinction is truly fundamental, since now the spin angular momentum has nothing to do with spatial motion. Elementary particles are believed to be structureless point particles and hence have no axis to spin around. They carry an intrinsic form of angular momentum, which we call *spin*, in addition to their extrinsic orbital angular momentum.

The existence of this *intrinsic* form of angular momentum was inferred from experiments such as the Stern-Gerlach experiment,[5] an experiment which was decisive in convincing physicists of the reality of quantized angular momentum. Quantized, since the spin can only take integer and half-integer values.

Systems consisting of spin-$\frac{1}{2}$ objects belong to the simplest quantum systems, as it admits just two basis states.[4] Understanding spin-$\frac{1}{2}$ dynamics is of great importance to describe magnetic systems, as these systems are described in terms of the magnetic dipole moment of atomic spins. Simply put, a spin-$\frac{1}{2}$ can be in either one of two states, which are commonly referred to as the up and down states. Respectively, the up and down state are assigned the values $+1$ and -1. The net sum of all magnetic dipole moments in a material determines to a large extend the magnetic properties of a material.

We will now elaborate on spin-$\frac{1}{2}$ models which are of importance for this thesis.

2.2.1 Noninteracting spin-$\frac{1}{2}$ particles

The simplest spin system one can imagine is that of a single spin-$\frac{1}{2}$ in an external magnetic field with strength B in the z-direction. The energy of such a system is given by[4]

$$E(s) = -\gamma B s. \tag{10}$$

Here, γ is a proportionality constant, known as the gyromagnetic ratio, and $s = \pm 1$ is the spin value. It follows immediately that the ground state, i.e. the lowest-energy state, is the one where the spin aligns with the the field, i.e. $s = +1$. Note that the energy is often also expressed as

$$E(\mu) = -\mu B,$$

where $\mu = \gamma s$ is the magnetic dipole moment associated with a spinning charged particle.

According to basic statistical physics,[6,7] the probability we find the particle in the state s is, in thermal equilibrium, Boltzmann distributed, i.e.

$$P(s) = \frac{\mathrm{e}^{-\beta E(s)}}{Z}.$$

Here, $\beta = 1/(k_B T)$ with k_B Boltzmann's constant and T the temperature. Furthermore, Z is known as the partition function, which acts as a normalization factor and is given by

$$Z = \sum_s \mathrm{e}^{-\beta E(s)} = 2\cosh(\beta\gamma B).$$

The partition function in general plays an important role in statistical physics, as many thermodynamic variables can be expressed in terms of the partition function or its derivatives.

In our case, it would be interesting to compute the *magnetization* M. It follows from the free energy F as the derivative[7]

$$M = -\frac{\partial F}{\partial B},$$

where the free energy itself is given by

$$F = -\frac{1}{\beta}\ln(Z).$$

Hence, the magnetization of a single spin-$\frac{1}{2}$ particle in an external field is given by

$$M = \gamma \tanh(\beta\gamma B).$$

Now consider N spin-$\frac{1}{2}$ particles in an external field. To determine the partition function, we need to sum over the possible states of all spins, i.e.

$$\begin{aligned}
Z_N &= \sum_{s_1=\pm 1}\sum_{s_2=\pm 1}\cdots\sum_{s_N=\pm 1} \mathrm{e}^{-\beta\sum_{i=1}^N E(s_i)} \\
&= \sum_{s_1=\pm 1}\sum_{s_2=\pm 1}\cdots\sum_{s_N=\pm 1}\prod_{i=1}^N \mathrm{e}^{\gamma\beta s_i} \\
&= \prod_{i=1}^N \left(\sum_{s_i=\pm 1} \mathrm{e}^{\gamma\beta s_i}\right) \\
&= \left(\sum_{s_1=\pm 1} \mathrm{e}^{\gamma\beta s_1}\right)^N \\
&= \left(2\cosh(\beta\gamma B)\right)^N .
\end{aligned}$$

This gives us for the net magnetization

$$M = N\gamma\tanh(\beta\gamma B), \tag{11}$$

which is simply N times the magnetization of a single particle.

To make the connection with the Battle of the Sexes later on, we need to accommodate for not one particle species, but two, due to the difference in preference of the players, which makes them distinguishable. We can include this by assigning gyromagnetic ratio γ_A to one species, and γ_B to the other. Supposing we have N_A particles of type A and N_B of type B, the partition function becomes

$$Z_{N_A+N_B} = \sum_{s_1^A=\pm 1}\sum_{s_2^A=\pm 1}\cdots\sum_{s_{N_A}^A=\pm 1}\sum_{s_1^B=\pm 1}\cdots\sum_{s_{N_B}^B=\pm 1} \mathrm{e}^{-\beta\left[\sum_{i=1}^{N_A} E(\gamma_A|s_i^A)+\sum_{i=1}^{N_B} E(\gamma_B|s_i^B)\right]}.$$

For clarity we made the distinction between spins of different types explicit by writing s_i^A and s_i^B for type A and B spins, respectively.

By going through the same procedure as before, one can easily show that we now have

$$Z_{N_A+N_B} = \left(2\cosh(\beta\gamma_A B)\right)^{N_A}\left(2\cosh(\beta\gamma_B B)\right)^{N_B}$$

and

$$M = N_A\gamma_A\tanh(\beta\gamma_A B) + N_B\gamma_B\tanh(\beta\gamma_B B). \tag{12}$$

We thus find that the two different species can be treated separately, as one would expect for noninteracting particles, such that the net magnetization is simply the sum of all individual components' magnetization.

2.2.2 The Ising model and its exact solution in 1 dimension

Now consider N (identical) spin-$\frac{1}{2}$ particles in an external field, but this time neighboring spins interact with each other with interaction strength J. For simplicity, we consider our model in 1 dimension only, and with periodic boundary conditions, i.e. $s_{i+N} = s_i$. It is then described by the famous Ising model, a model by Wilhelm Lenz and first solved by Ernst Ising in his 1924 thesis.[8] The total system's energy is given by[7]

$$E(\{s_i\}) = -\sum_{i=1}^{N}\left[Js_is_{i+1} + \frac{\gamma B}{2}\left(s_i + s_{i+1}\right)\right]. \tag{13}$$

The partition function is now given by

$$\begin{aligned} Z &= \sum_{s_1=\pm1}\sum_{s_2=\pm1}\cdots\sum_{s_N=\pm1}\exp\left\{\beta\sum_{i=1}^{N}\left[Js_is_{i+1} + \frac{\gamma B}{2}\left(s_i + s_{i+1}\right)\right]\right\} \\ &= \sum_{s_1=\pm1}\sum_{s_2=\pm1}\cdots\sum_{s_N=\pm1}\prod_{i=1}^{N}\exp\left\{\beta\left[Js_is_{i+1} + \frac{\gamma B}{2}\left(s_i + s_{i+1}\right)\right]\right\} \\ &= \sum_{s_1=\pm1}\sum_{s_2=\pm1}\cdots\sum_{s_N=\pm1}\prod_{i=1}^{N}T_{s_i,s_{i+1}}. \end{aligned}$$

Here we used

$$\langle s_i|\mathbf{T}|s_{i+1}\rangle \equiv T_{s_i,s_{i+1}} = \exp\left\{\beta\left[Js_is_{i+1} + \frac{\gamma B}{2}\left(s_i + s_{i+1}\right)\right]\right\},$$

where we introduced the *transfer matrix*

$$\mathbf{T} = \begin{pmatrix} T_{1,1} & T_{1,-1} \\ T_{-1,1} & T_{-1,-1} \end{pmatrix}.$$

The partition function can now be written as

$$\begin{aligned} Z &= \sum_{s_1}\sum_{s_2}\ldots\sum_{s_N}\prod_{i=1}^{N}\langle s_i|\mathbf{T}|s_{i+1}\rangle \\ &= \sum_{s_1}\sum_{s_2}\ldots\sum_{s_N}\langle s_1|\mathbf{T}|s_2\rangle\langle s_2|\mathbf{T}|s_3\rangle\ldots\langle s_N|\mathbf{T}|s_1\rangle \\ &= \sum_{s_1}\left\langle s_1|\mathbf{T}^N|s_1\right\rangle. \end{aligned}$$

Hence, the partition function is given by the trace of the N-th power of the transfer matrix

$$Z = \mathrm{tr}\left[\mathbf{T}^N\right].$$

The transfer matrix is symmetric, and thus one can perform a linear transformation to write $\mathbf{T} = \mathbf{U}^{-1}\mathbf{D}\mathbf{U}$, with $\mathbf{D}$ a diagonal matrix and $\mathbf{U}$ unitary. Then

$$\begin{aligned} Z &= \mathrm{tr}\left[\left(\mathbf{U}^{-1}\mathbf{D}\mathbf{U}\right)^N\right] = \mathrm{tr}\left[\left(\mathbf{U}^{-1}\mathbf{D}\mathbf{U}\right)\left(\mathbf{U}^{-1}\mathbf{D}\mathbf{U}\right)\ldots\left(\mathbf{U}^{-1}\mathbf{D}\mathbf{U}\right)\right] \\ &= \mathrm{tr}\left[\mathbf{U}^{-1}\mathbf{D}^N\mathbf{U}\right] \\ &= \mathrm{tr}\left[\mathbf{D}^N\right], \end{aligned}$$

where we used that the trace is invariant under cyclic permutations. Now, since $\mathbf{D}$ is diagonal, we have

$$Z = \lambda_+^N + \lambda_-^N,$$

where the eigenvalues

$$\lambda_\pm = \mathrm{e}^{\beta J}\cosh(\beta\gamma B) \pm \mathrm{e}^{\beta J}\sqrt{\sinh^2(\beta\gamma B) + \mathrm{e}^{-4\beta J}}$$

are the roots of the equation $\det(\lambda\mathbf{1} - \mathbf{T}) = 0$. The eigenvalues are always positive and $\lambda_+ > \lambda_-$, except for $B = 0$. In the latter case $\lambda_+ = 2\cosh(\beta J)$ and $\lambda_- = 2\sinh(\beta J)$, such that $\lambda_+ \geq \lambda_-$, with the degeneracy ($\lambda_+ = \lambda_-$) occurring in the limit $\beta \to \infty$ (i.e. $T \to 0$).

One can rewrite the partition function as

$$Z_\beta = \lambda_+^N \left[1 + \left(\frac{\lambda_-}{\lambda_+}\right)^N\right],$$

such that in the *thermodynamic limit* ($N \to \infty$) one finds the free energy per spin

$$f(\beta, B) = -\frac{1}{\beta N}\ln Z_\beta = -\frac{1}{\beta}\ln\lambda_+.$$

Hence

$$f(\beta, B) = -\frac{1}{\beta}\ln\left[\mathrm{e}^{\beta J}\cosh(\beta\gamma B) + \mathrm{e}^{\beta J}\sqrt{\sinh^2(\beta\gamma B) + \mathrm{e}^{-4\beta J}}\right],$$

from which we can determine all thermodynamic properties. The magnetization per spin is given by

$$m = \frac{M}{N} = -\frac{\partial f}{\partial B} = \frac{\gamma\sinh(\beta\gamma B)}{\sqrt{\sinh^2(\beta\gamma B) + e^{-4\beta J}}}, \tag{14}$$

which vanishes for $B = 0$, just as for the uncoupled case. Note that the magnetization reduces to the result for noninteracting spins (Eqn. 11) for $J = 0$, as it should.

Just like we did for noninteracting spins, we will now introduce a second particle species and assign them different gyromagnetic ratios γ_A and γ_B. For clarity we will once more write the respective spin values as s^A and s^B. We will consider a bipartite lattice, i.e. a lattice on which a particle of species A has nearest-neighbors of type B and vice versa. For such a lattice, it is important to note that spins of the same type do not interact with each other, but only with spins of the other type (at least for nearest-neighbor interactions). If we assume particles of type A are on the even lattice sites and particles of type B on the odd sites, we can write

$$E(\{s_i\}) = -J\sum_{i=1}^{N/2}\left(s_{2i}^A s_{2i+1}^B + s_{2i+1}^B s_{2i+2}^A\right) - \frac{B}{2}\sum_{i=1}^{N}\left[(\gamma_A s_{2i} + \gamma_B s_{2i+1}) + (\gamma_B s_{2i+1}^B + \gamma_A s_{2i+2}^A)\right]. \tag{15}$$

For the partition function we now need to sum all states of both species, i.e.

$$\begin{aligned} Z &= \sum_{\{s_i^A\}}\sum_{\{s_j^B\}}\prod_{i=1}^{N/2} \mathrm{e}^{\beta\left[Js_{2i}^A s_{2i+1}^B + \frac{B}{2}(\gamma_A s_{2i}^A + \gamma_B s_{2i+1}^B)\right]}\mathrm{e}^{\beta\left[Js_{2i+1}^B s_{2i+2}^A + \frac{B}{2}(\gamma_B s_{2i+1}^B + \gamma_A s_{2i+2}^A)\right]} \\ &= \sum_{\{s_i^A\}}\sum_{\{s_j^B\}}\prod_{i=1}^{N/2} \tilde{T}_{s_{2i}^A, s_{2i+1}^B}\hat{T}_{s_{2i+1}^B, s_{2i+2}^A}, \end{aligned}$$

Here we have matrix elements

$$\begin{aligned} \tilde{T}_{s_{2i}^A, s_{2i+1}^B} &= \mathrm{e}^{\beta\left[Js_{2i}^A s_{2i+1}^B + \frac{B}{2}(\gamma_A s_{2i}^A + \gamma_B s_{2i+1}^B)\right]} \equiv \langle s_{2i}^A|\tilde{\mathbf{T}}|s_{2i+1}^B\rangle \\ \hat{T}_{s_{2i+1}^B, s_{2i+2}^A} &= \mathrm{e}^{\beta\left[Js_{2i+1}^B s_{2i+2}^A + \frac{B}{2}(\gamma_B s_{2i+1}^B + \gamma_A s_{2i+2}^A)\right]} \equiv \langle s_{2i+1}^B|\hat{\mathbf{T}}|s_{2i+2}^A\rangle, \end{aligned}$$

with

$$\tilde{\mathbf{T}} = \begin{pmatrix} \tilde{T}_{1,1} & \tilde{T}_{1,-1} \\ \tilde{T}_{-1,1} & \tilde{T}_{-1,-1} \end{pmatrix} \quad \text{and} \quad \hat{\mathbf{T}} = \begin{pmatrix} \hat{T}_{1,1} & \hat{T}_{1,-1} \\ \hat{T}_{-1,1} & \hat{T}_{-1,-1} \end{pmatrix}.$$

By computing the matrices explicitly, one can easily show that $\hat{\mathbf{T}} = \tilde{\mathbf{T}}^{\intercal}$. The matrix $\tilde{\mathbf{T}}\tilde{\mathbf{T}}^{\intercal}$ is still diagonalizable, such that we can proceed as we did for the single particle species. Hence

$$Z = \lambda_+^{N/2} + \lambda_-^{N/2},$$

where the power $N/2$ is the result of having split the lattice in two sublattices consisting of the even and odd labeled sites. Note that $\lambda_\pm$ are now the eigenvalues of $\tilde{\mathbf{T}}\tilde{\mathbf{T}}^{\intercal}$. Using once more that $\lambda_+ > \lambda_-$, λ_+ is the only surviving term in the thermodynamic limit. Computing the free energy and taking its derivative with respect to B, we now find for the magnetization per spin

$$\begin{aligned} m = \Big(& \cosh(\beta B\gamma_A)\sinh(\beta B\gamma_B)[\gamma_A \sinh(2\beta J) + \gamma_B \cosh(2\beta J)] \\ & + \sinh(\beta B\gamma_A)\cosh(\beta B\gamma_B)[\gamma_A \cosh(2\beta J) + \gamma_B \sinh(2\beta J)]\Big) \\ \Big/ \Big(& \sinh(4\beta J)\sinh(2\beta B\gamma_A)\sinh(2\beta B\gamma_B) + \cosh(2\beta B\gamma_B) - \cosh(4\beta J) + 2 \\ & + \cosh(2\beta B\gamma_A)[\cosh(4\beta J)\cosh(2\beta B\gamma_B) + 1]\Big)^{1/2}, \end{aligned} \tag{16}$$

which reduces to Eqn. 14 when taking $\gamma_A = \gamma_B = \gamma$ and to Eqn. 12 when setting $J = 0$.

3 Response strategy analysis

We will now turn to a complete analysis of the response strategy for the correlated Battle of the Sexes game, as described in Sec. 2.1.3, with the main focus on the equilibrium solutions.

We assume for simplicity, but this can easily be generalized, symmetric probabilities of the intrinsic correlations, or initial correlation device, i.e. $p_{AB} = p_{BA} = \frac{1}{2}(1 - p_{AA} - p_{BB})$, with $p_{AA} + p_{BB} \leq 1$. The coefficients of player i then depend on p_{AA} and p_{BB} only, and on the response probabilities of the other players. To find all possible Nash equilibria in the $p_{AA} - p_{BB}$ plane, we simply impose the conditions on the coefficients in the payoff function, $C_A^{(i)}$ and $C_B^{(i)}$, as discussed in Sec. 2.1.3. Since these coefficients are linear in the response strategy of the other player, they can be solved exactly. To obtain a self-consistent solution, each slope is calculated while assuming the other player's response probabilities are in equilibrium, since no single player should have any incentive to deviate from a Nash equilibrium.

To recap, the equilibrium conditions we impose are the coefficients $C_\mu^{(i)}(p_{AA}, p_{BB}, P_{F_A}^{(-i)}, P_{F_B}^{(-i)})$ being zero, unless the equilibrium response strategy $P_{F_\mu}^{(i)}$ is at an extremum, i.e. equal to 0 or 1, in which case the corresponding coefficient should be nonzero and negative or positive, respectively. We thus have possible Nash equilibria for $P_{F_\mu}^{(i)} = 0$, $P_{F_\mu}^{(i)} = 1$ and $0 < P_{F_\mu}^{(i)} < 1$. Note that the response probabilities of both players need not be equal to reach equilibrium.

We will now look into each of the possible Nash equilibria.

3.1 Pure-strategy equilibria

We start with a response strategy profile corresponding to the pure-strategy equilibria, in the sense that the final outcomes of the game are the pure-strategy Nash equilibria of the pure-strategy Battle of the Sexes game. To consistently get (A, A) as the final outcome, both players should always follow the instruction to play A, while they should never obey the correlation device if the instruction is to play B. Hence $P_{F_A}^{(i)*} = 1$ and $P_{F_B}^{(i)*} = 0$, such that the conditions on the coefficients become

$$\begin{cases} C_A^{(i)}\left(p_{AA}, p_{BB}, 1, 0\right) > 0 \\ C_B^{(i)}\left(p_{AA}, p_{BB}, 1, 0\right) < 0 \end{cases} \forall\, i \in \{1, 2\}. \tag{17}$$

Similarly, to reach (B, B) consistently, the response probabilities must be $P_{F_A}^{(i)*} = 0$ and $P_{F_B}^{(i)*} = 1$, such that

$$\begin{cases} C_A^{(i)}\left(p_{AA}, p_{BB}, 0, 1\right) < 0 \\ C_B^{(i)}\left(p_{AA}, p_{BB}, 0, 1\right) > 0 \end{cases} \forall\, i \in \{1, 2\} \tag{18}$$

are the conditions for the coefficients.

We find that both sets of conditions hold true in the entire $p_{AA} - p_{BB}$ plane. Hence, regardless of the correlation device, we have a Nash equilibrium for the response strategy profile

$$\Pi^* = \{\Pi_{1,0}^{(1)}, \Pi_{1,0}^{(2)}\} \quad \text{and for} \quad \Pi^* = \{\Pi_{0,1}^{(1)}, \Pi_{0,1}^{(2)}\}, \tag{19}$$

corresponding to pure-strategy equilibria (A, A) and (B, B), respectively.

3.2 Mixed-strategy equilibrium

Next, we impose slope conditions with both players having $0 < P_{F_\mu}^{(i)*} < 1$ and thus where the coefficients are all zero. Hence

$$\begin{cases} C_A^{(i)}\left(p_{AA}, p_{BB}, P_{F_A}^{(-i)*}, P_{F_B}^{(-i)*}\right) = 0 \\ C_B^{(i)}\left(p_{AA}, p_{BB}, P_{F_A}^{(-i)*}, P_{F_B}^{(-i)*}\right) = 0 \end{cases} \forall\, i \in \{1, 2\}. \tag{20}$$

We find once more that these conditions hold for any point in the $p_{AA} - p_{BB}$ plane, and that the equilibrium response strategies are given by

$$P_{F_A}^{(1)*} = \frac{1}{1+S} = P_A^{(1)*} \quad \text{and} \quad P_{F_B}^{(1)*} = \frac{S}{1+S} = P_B^{(1)*}, \tag{21a}$$

$$P_{F_A}^{(2)*} = \frac{S}{1+S} = P_A^{(2)*} \quad \text{and} \quad P_{F_B}^{(2)*} = \frac{1}{1+S} = P_B^{(2)*}. \tag{21b}$$

Hence

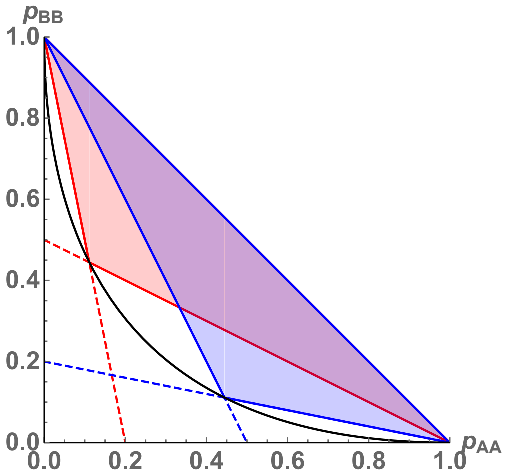
$$\Pi^* = \{\Pi_{P_A, P_B}^{(1)*}, \Pi_{P_A, P_B}^{(2)*}\} \tag{22}$$

is a Nash equilibrium in the response strategy, and corresponds to the mixed-strategy Nash equilibrium. Note that this equilibrium is uncorrelated, as the renormalized probabilities reduce to $p_{\mu\nu}^R = P_\mu^{(1)*} P_\nu^{(2)*}$, regardless of the initial probabilities of the correlation device.

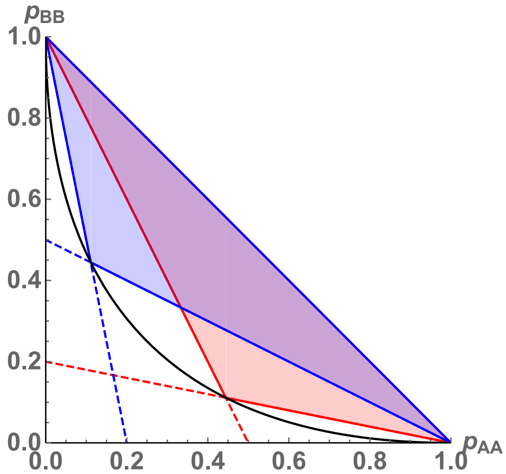
3.3 Correlated Equilibria

Now consider the response probabilities $P_{F_\mu}^{(i)*} = 1$, which correspond to the correlated equilibrium as described in Sec. 2.1.3. With these response probabilities, the slope conditions become

$$\begin{cases} C_A^{(i)}\left(p_{AA}, p_{BB}, 1, 1\right) > 0 \\ C_B^{(i)}\left(p_{AA}, p_{BB}, 1, 1\right) > 0 \end{cases} \forall\, i \in \{1, 2\} \tag{23}$$



(a) Always obey correlation device



(b) Never obey correlation device

Figure 1: Regions where the conditions given in ***(a)*** *Eqn. 23, and as given in* ***(b)*** *Eqn. 28, are obeyed. The area where the conditions hold for player 1 are indicated in red, while it is indicated in blue for player 2. The straight lines are the lower boundaries that follow from the imposed conditions. The black solid line separates the region where both players use the same response strategy (above the line) from where both players use opposite response strategies (below the line). Here, $S = 1/2$ was used.*

First of all, these conditions lead to the additional constraint

$$p_{BB} > 1 - 2\sqrt{p_{AA}} + p_{AA}, \tag{24}$$

which is indicated as the solid black line in Fig. 1.

For both players, the conditions in Eqn. 23 furthermore lead to inequalities defining an area in the $p_{AA} - p_{BB}$ plane for which the conditions hold true. In this case the area wherein player 1's conditions hold is defined by

$$\frac{1 - p_{AA} - p_{BB}}{2p_{BB}} < S < \frac{2p_{AA}}{1 - p_{AA} - p_{BB}}, \tag{25}$$

while for player 2 it is defined by

$$\frac{1 - p_{AA} - p_{BB}}{2p_{AA}} < S < \frac{2p_{BB}}{1 - p_{AA} - p_{BB}}. \tag{26}$$

By comparing these areas, we find the region where the solutions are self-consistent (Fig. 1a). Within this area we recover the correlated equilibrium, i.e. both players always follow the instructions of the correlation device. The equilibrium response strategy profile for this region is therefore

$$\Pi^* = \{\Pi^{(1)}_{1,1}, \Pi^{(2)}_{1,1}\}. \tag{27}$$

Similarly, an equilibrium is reached for $P^{(i)*}_{F_\mu} = 0$. By imposing

$$\begin{cases} C^{(i)}_A(p_{AA}, p_{BB}, 0, 0) < 0 \\ C^{(i)}_B(p_{AA}, p_{BB}, 0, 0) < 0 \end{cases} \quad \forall i \in \{1, 2\}, \tag{28}$$

we find that the area where these conditions hold is given by

$$\frac{1-p_{AA}-p_{BB}}{2p_{AA}} < S < \frac{2p_{BB}}{1-p_{AA}-p_{BB}}, \tag{29}$$

for player 1, while for player 2

$$\frac{1-p_{AA}-p_{BB}}{2p_{BB}} < S < \frac{2p_{AA}}{1-p_{AA}-p_{BB}}. \tag{30}$$

The areas are shown for both players in Fig. 1b for $S = 1/2$, from which we can see that the area wherein the solutions are self-consistent is the same as in Fig. 1a. This shows that in this area (VII in Fig. 3) we have an equilibrium given by Eqn. 27, as well as for

$$\Pi^* = \{\Pi^{(1)}_{0,0}, \Pi^{(2)}_{0,0}\}. \tag{31}$$

The reason both of these strategy profiles are equilibria for the same correlation devices can easily be understood, since it simply constitutes to redefining the actions players take when receiving certain information. Instead of choosing A when receiving the information to play A, they will play B, and similarly when receiving the instruction to play B, they will play A instead. In our example in Sec. 2.1.3 with the coin flip, it would be equivalent to instructing B for heads and A for tails, instead of the other way around.

3.4 New Nash equilibria

Up until this point we considered response strategies corresponding to already known equilibria. However, there are more strategy profiles possible, which, as we will see, will lead to new equilibria.

The first profile we consider is the one for which player 1 always follows the instructions given, while player 2 never follows given instructions. Hence we impose

$$\begin{cases} C^{(1)}_A(p_{AA}, p_{BB}, 0, 0) > 0 \\ C^{(1)}_B(p_{AA}, p_{BB}, 0, 0) > 0 \end{cases} \wedge \begin{cases} C^{(2)}_A(p_{AA}, p_{BB}, 1, 1) < 0 \\ C^{(2)}_B(p_{AA}, p_{BB}, 1, 1) < 0 \end{cases}. \tag{32}$$

These conditions hold in the area defined by

$$p_{BB} < \frac{S}{S+2}(1-p_{AA}) \quad \wedge \quad p_{BB} < 1-(1+2S)p_{AA}, \tag{33}$$

which is visualized in Fig. 2a for $S = 1/2$. In this area (I and III in Fig. 3) we thus have the Nash equilibrium given by

$$\Pi^* = \{\Pi^{(1)}_{1,1}, \Pi^{(2)}_{0,0}\}. \tag{34}$$

We also obtain the additional constraint

$$p_{BB} < 1 - 2\sqrt{p_{AA}} + p_{AA}, \tag{35}$$

similar to Eqn. 24. This shows that the curve $p_{BB} = 1 - 2\sqrt{p_{AA}} + p_{AA}$ divides the $p_{AA} - p_{BB}$ plane in an area where the coefficients of both players have equal signs (above the curve), from an area where they have opposite signs (below the curve). In other words, above the curve the players use the same response strategy (at equilibrium), while below the curve they will play with opposite strategies.

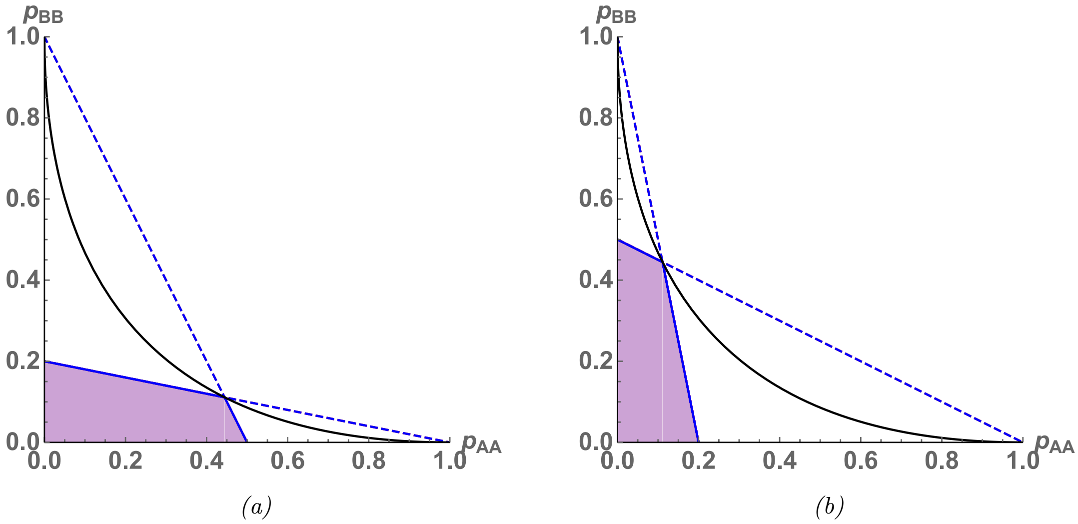


*Figure 2: **(a)** shows the area where the conditions as given in Eqn. 32 are obeyed. In this area an equilibrium is achieved if player 1 always follows the instructions of the correlation device, while player 2 never follows the instructions. In **(b)** the area is shown where the conditions of Eqn. 36 are met. Here, an equilibrium is achieved if player 1 never follows the instructions of the correlation device, while player 2 always follows. For these figures $S = 1/2$ was used.*

Similarly, we can impose the conditions corresponding to player 1 never obeying the correlation device and player 2 always obeying, i.e.

$$\left\{ \begin{array}{l} C_A^{(1)}\left(p_{AA}, p_{BB}, 1, 1\right) < 0 \\ C_B^{(1)}\left(p_{AA}, p_{BB}, 1, 1\right) < 0 \end{array} \right. \wedge \left\{ \begin{array}{l} C_A^{(2)}\left(p_{AA}, p_{BB}, 0, 0\right) > 0 \\ C_B^{(2)}\left(p_{AA}, p_{BB}, 0, 0\right) > 0 \end{array} \right., \tag{36}$$

which hold in the area defined by

$$p_{BB} < \frac{1}{1+2S}\left(1 - p_{AA}\right) \quad \wedge \quad p_{BB} < 1 - \frac{2+S}{S} p_{AA}, \tag{37}$$

shown in Fig. 2b for $S = 1/2$. In this area (I and II in Fig. 3) the strategy profile

$$\Pi^* = \{\Pi_{0,0}^{(1)}, \Pi_{1,1}^{(2)}\} \tag{38}$$

is a Nash equilibrium. Note that in the overlap of the areas shown in Fig. 2 both of the corresponding strategy profiles are Nash equilibria.

As opposed to having both coefficients nonzero, i.e. all response strategies either 0 or 1, we can also take one coefficient equal to zero, with the other coefficient nonzero. Consider

$$\left\{ \begin{array}{l} C_A^{(1)}\left(p_{AA}, p_{BB}, 0, P_{F_B}^{(2)*}\right) > 0 \\ C_B^{(1)}\left(p_{AA}, p_{BB}, 0, P_{F_B}^{(2)*}\right) = 0 \end{array} \right. \wedge \left\{ \begin{array}{l} C_A^{(2)}\left(p_{AA}, p_{BB}, 1, P_{F_B}^{(1)*}\right) < 0 \\ C_B^{(2)}\left(p_{AA}, p_{BB}, 1, P_{F_B}^{(1)*}\right) = 0 \end{array} \right.. \tag{39}$$

These conditions hold in the areas II, IV and VI (Fig. 3), which are defined by

$$\frac{S}{S+2}\left(1 - p_{AA}\right) < p_{BB} < 1 - 2\sqrt{p_{AA}} + p_{AA}, \tag{40}$$

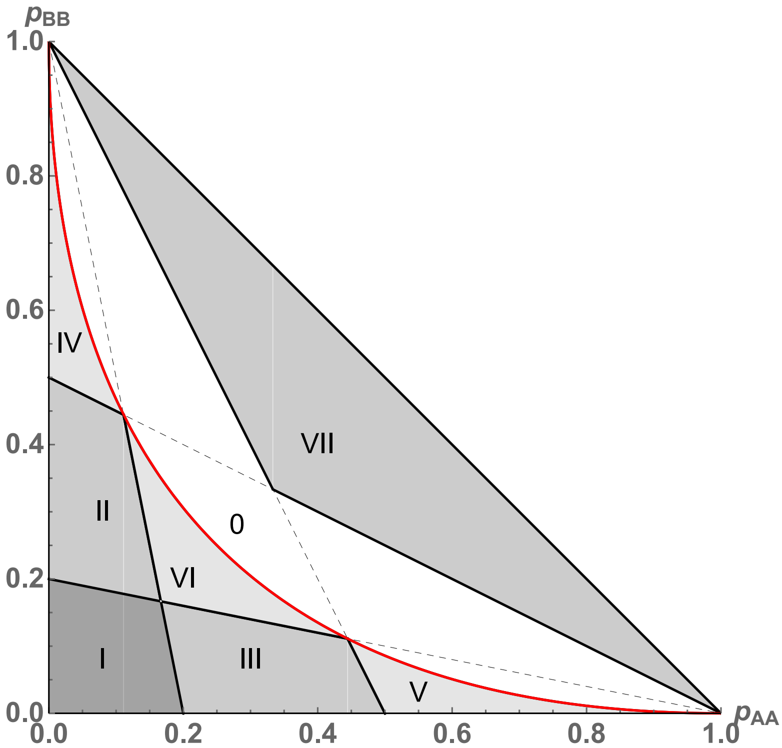


Figure 3: All Nash equilibria in the response strategy of the correlated Battle of the Sexes game, represented in the $p_{AA} - p_{BB}$ plane, with $S = 1/2$. In App. C one can see this plane for other S values.

and the equilibrium response strategies we find are given by

$$P_{F_B}^{(1)*} = \frac{S}{S+1}\frac{1-p_{AA}+p_{BB}}{2p_{BB}} \quad \wedge \quad P_{F_B}^{(2)*} = \frac{1}{S+1}\frac{S(p_{AA}-1)+(S+2)p_{BB}}{2p_{BB}}, \tag{41}$$

such that the strategy profile for this Nash equilibrium becomes

$$\Pi^* = \{\Pi^{(1)}_{1,P_{F_B}}, \Pi^{(2)}_{0,P_{F_B}}\}. \tag{42}$$

Note that $P_B^{(1)*} < P_{F_B}^{(1*)} < 1$ and $0 < P_{F_B}^{(2)*} < P_B^{(2)*}$, such that the response probabilities lie between the mixed-strategy equilibrium value and the value of the other response probability at one of the extrema (either 0 or 1).

Similarly, the set of conditions

$$\left\{ \begin{array}{l} C_A^{(1)}\left(p_{AA}, p_{BB}, P_{F_A}^{(2)*}, 1\right) = 0 \\ C_B^{(1)}\left(p_{AA}, p_{BB}, P_{F_A}^{(2)*}, 1\right) < 0 \end{array} \right. \quad \wedge \quad \left\{ \begin{array}{l} C_A^{(2)}\left(p_{AA}, p_{BB}, , P_{F_A}^{(1)*}, 0\right) = 0 \\ C_B^{(2)}\left(p_{AA}, p_{BB}, , P_{F_A}^{(1)*}, 0\right) > 0 \end{array} \right. \tag{43}$$

is true for a correlation device with the property

$$1 - \frac{2+S}{S}p_{AA} < p_{BB} < 1 - 2\sqrt{p_{AA}} + p_{AA}, \tag{44}$$

which correspond to areas III, V and VI in Fig. 3. The response probabilities we find are then

$$P_{F_A}^{(1)*} = \frac{1}{S+1}\frac{(S+2)p_{AA} + S(p_{BB}-1)}{2p_{AA}} \quad \wedge \quad P_{F_A}^{(2)*} = \frac{S}{S+1}\frac{1+p_{AA}-p_{BB}}{2p_{AA}}, \tag{45}$$

such that the response-strategy profile of this Nash equilibrium is given by

$$\Pi^* = \{\Pi^{(1)}_{P_{F_A},0}, \Pi^{(2)}_{P_{F_A},1}\}. \tag{46}$$

Here, we find once more that $0 < P_{F_A}^{(1)*} < P_A^{(1)*}$ and $P_A^{(2)*} < P_{F_A}^{(2)*} < 1$.

Next, the conditions given by

$$\begin{cases} C_A^{(1)}\left(p_{AA}, p_{BB}, 1, P_{F_B}^{(2)*}\right) < 0 \\ C_B^{(1)}\left(p_{AA}, p_{BB}, 1, P_{F_B}^{(2)*}\right) = 0 \end{cases} \wedge \quad \begin{cases} C_A^{(2)}\left(p_{AA}, p_{BB}, 0, P_{F_B}^{(1)*}\right) > 0 \\ C_B^{(2)}\left(p_{AA}, p_{BB}, 0, P_{F_B}^{(2)*}\right) = 0 \end{cases} \tag{47}$$

lead to the area in the $p_{AA} - p_{BB}$ plane defined by

$$\frac{1}{1+2S}(1-p_{AA}) < p_{BB} < 1 - 2\sqrt{p_{AA}} + p_{AA}, \tag{48}$$

corresponding to area IV in Fig. 3, with equilibrium probabilities

$$P_{F_B}^{(1)*} = \frac{1}{S+1}\frac{p_{AA} - 1 + (1+2S)p_{BB}}{2p_{BB}} \quad \wedge \quad P_{F_B}^{(2)*} = \frac{1}{S+1}\frac{1-p_{AA}+p_{BB}}{2p_{BB}}, \tag{49}$$

for which $0 < P_{F_B}^{(1)*} < P_B^{(1)*}$ and $P_B^{(2)*} < P_{F_B}^{(2)*} < 1$. The strategy profile corresponding to this Nash equilibrium is

$$\Pi^* = \{\Pi^{(1)}_{0,P_{F_B}}, \Pi^{(2)}_{1,P_{F_B}}\}. \tag{50}$$

Finally, the conditions

$$\begin{cases} C_A^{(1)}\left(p_{AA}, p_{BB}, P_{F_A}^{(2)*}, 0\right) = 0 \\ C_B^{(1)}\left(p_{AA}, p_{BB}, P_{F_A}^{(2)*}, 0\right) > 0 \end{cases} \wedge \quad \begin{cases} C_A^{(2)}\left(p_{AA}, p_{BB}, P_{F_A}^{(1)*}, 1\right) = 0 \\ C_B^{(2)}\left(p_{AA}, p_{BB}, P_{F_A}^{(1)*}, 1\right) < 0 \end{cases} \tag{51}$$

hold for a correlation device obeying

$$1 - (1+2S)p_{AA} < p_{BB} < 1 - 2\sqrt{p_{AA}} + p_{AA}, \tag{52}$$

i.e. area V in Fig. 3, and the equilibrium response strategies for such a device are

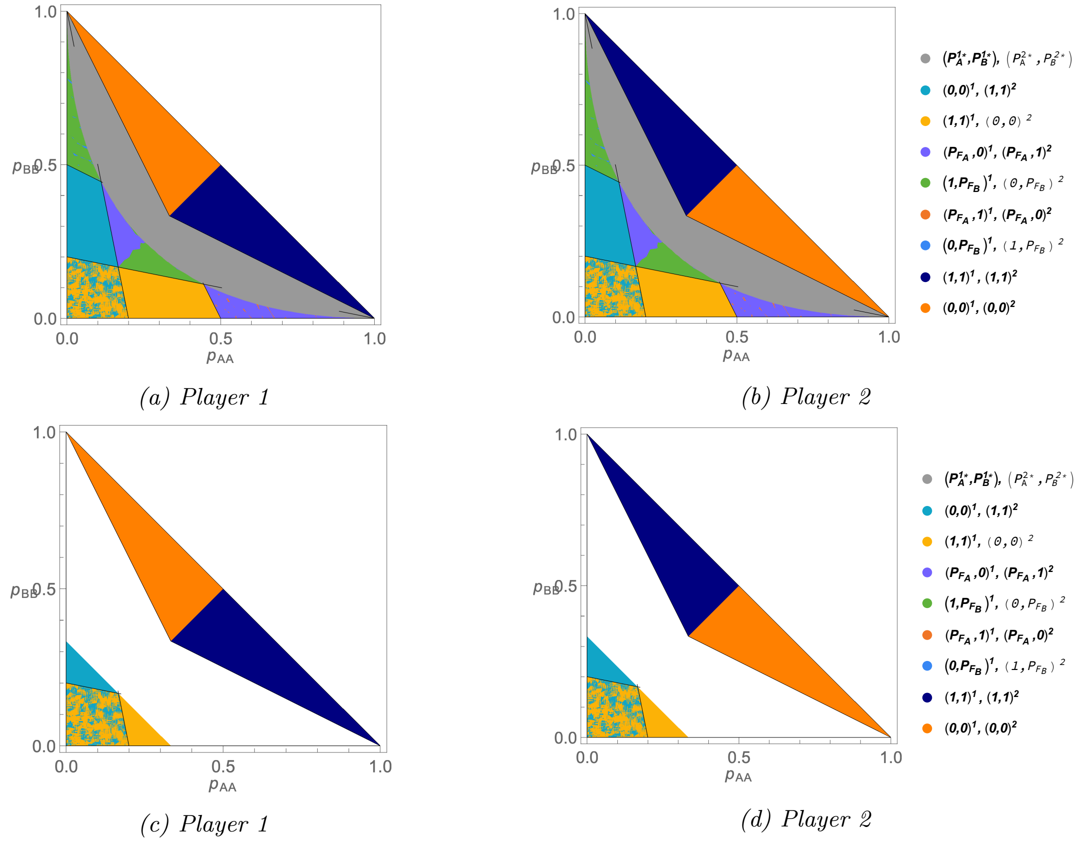
$$P_{F_A}^{(1)*} = \frac{1}{1+S}\frac{1+p_{AA}-p_{BB}}{2p_{AA}} \quad \wedge \quad P_{F_A}^{(2)*} = \frac{1}{1+S}\frac{(1+2S)p_{AA}+p_{BB}-1}{2p_{AA}}, \tag{53}$$

such that the equilibrium profile is given by

$$\Pi^* = \{\Pi^{(1)}_{P_{F_A},1}, \Pi^{(2)}_{P_{F_A},0}\}. \tag{54}$$

Here we find again that the equilibrium response probabilities only have values for which $P_A^{(1)*} < P_{F_A}^{(1)*} < 1$ and $0 < P_{F_A}^{(2)*} < P_A^{(2)*}$. Note that all other possible sets of conditions for which we set one coefficient equal to zero, which all hold true above the curve $p_{BB} = 1 - 2\sqrt{p_{AA}} + p_{AA}$, do not lead to self-consistent solutions.

To conclude the discussion of all possible Nash equilibria, we once more refer to Fig. 3 showing all areas in the $p_{AA} - p_{BB}$ plane, for which the previously discussed Nash equilibria exist. We hereby remark that the pure and mixed-strategy Nash equilibria are solutions for any correlation device, whereas the other equilibria we found only exist for devices with specific probability distributions. A visualization of the Nash equilibria in the space of response probabilities is given in App. B.



(a) Player 1 (b) Player 2

(c) Player 1 (d) Player 2

*Figure 4: **(a)** shows which response strategy profiles are optimal for player 1 and **(b)** for player 2. In **(c)** and **(d)** we see the same, but now we only show where the payoffs are better than S, i.e. better than the payoff one would get from consistently going to one's least-preferred option. Here, we used $S = 1/2$.*

3.5 Optimal payoff

Before moving on, we will compute which response-strategy is optimal for a player, given a certain correlation device. This is necessary, as we assume a rational player will choose the highest-rewarding strategy which still leads to an equilibrium.

We already know all Nash equilibria in the full $p_{AA} - p_{BB}$ plane, thus we can easily compute a player's optimal strategy by plotting Eqn. 7. In Fig. 4 we have shown, for all values of the correlation device, which strategy profile leads to the highest payoff. In App. D we show the full 3-dimensional plots of the payoff.

We see that there is only a difference in optimal strategy between the players in area VII (of Fig. 3), in which above the line $p_{BB} = p_{AA}$ the strategy profile $\{\Pi^{(1)}_{0,0}, \Pi^{(2)}_{0,0}\}$ is optimal for player 1 and below this line $\{\Pi^{(1)}_{1,1}, \Pi^{(2)}_{1,1}\}$, whereas for player 2 it is the other way around.

We hereby note that for some initial probabilities of the correlation device it is possible to have multiple response-strategy profiles yielding the same payoff. We have furthermore left out the response strategies corresponding to the pure-strategy equilibria, since this is always the best a single player can do, but it would take us back to original discussion of 'fairness' and which of the two equilibria would ultimately be reached. From a statistical physics point of view it is also the least interesting case to study, as the outcome of the game would always be the same. We do, however, show in 4c and 4d which strategy profiles are more rewarding than picking one's least-preferred option consistently. This shows that, even for correlated games, the pure-strategy equilibria are, for a large part of the $p_{AA} - p_{BB}$ plane, still the better choice.

4 Effective spin models

We will now apply our knowledge of statistical physics to the results of the previous section, to model the Battle of the Sexes when played on some simple 1-dimensional networks. To model an N-player game, we assume players are placed on a bipartite lattice and only interact with their nearest-neighbors. Furthermore, we assume players in each pair pick the response strategy in equilibrium that leads to the highest payoff in a two-player game.

As discussed in Sec. 2.2, the probability to reach a certain state s in thermal equilibrium is given by

$$P(s) = \frac{1}{Z}\mathrm{e}^{-\beta E(s)}, \tag{55}$$

with $E(s)$ the energy of the corresponding state and Z the partition function.

In our game of Battle of the Sexes, we would consider the final state $\mu\nu$, e.g. (A, B), and its corresponding renormalized probability $p^R_{\mu\nu}$. Since we know which response probabilities optimize payoff for each correlation device, and thus which renormalized probability matrices are optimal, we can write them as effective Ising models by writing

$$p^R_{\mu\nu} = \frac{1}{\mathcal{Z}}\mathrm{e}^{-\beta H_{\mu\nu}} \tag{56}$$

Here, H the effective energy, which will depend on the Nash equilibrium we consider, and $\mathcal{Z}$ is an effective partition function, i.e. normalization constant.

To complete the analogy between the Battle of the Sexes and the Ising model, we let each player correspond to a particle with spin 1/2, and let A correspond to $\uparrow$ and B to $\downarrow$. The reason for this particular choice is to make sure that $E^{(1)}_A < E^{(1)}_B$ and $E^{(2)}_B < E^{(2)}_A$, such that the player's preference corresponds to the single-spin ground state.

4.1 Uncorrelated games

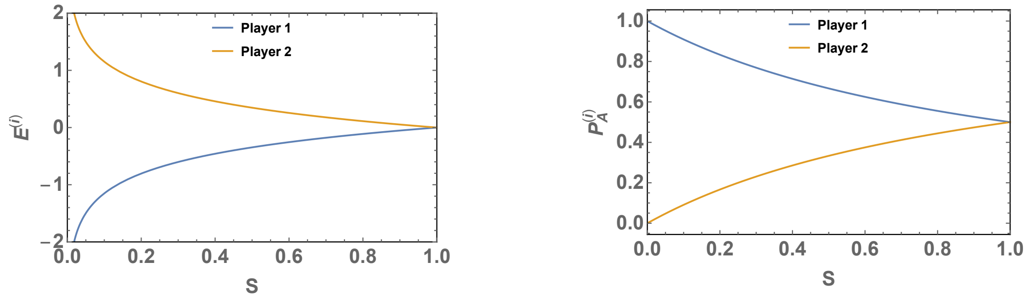
When the players use response strategies $P^{(i)*}_{F_\mu} = P^{(i)*}_\mu$, i.e. the situation as described in Sec. 3.2, one can show by using Eqn. 6 that, regardless of the initial correlation device, the renormalized probability matrix reduces to

$$p^R_{\mu\nu} = P^{(1)*}_\mu P^{(2)*}_\nu, \tag{57}$$

which is clearly an uncorrelated probability matrix. This therefore corresponds to two noninteracting spin-1/2 particles in a magnetic field, as described in Sec. 2.2.1. Since the particles are uncorrelated, we can treat them separately and make use of the fact that the final probability to reach $\mu\nu$ is simply the product of the individual particles' probabilities.

For each of the players we now have

$$P^{(i)}_\mu = \frac{1}{Z}\mathrm{e}^{-\beta H^{(i)}_\mu}, \tag{58}$$



(a) *Energy when playing A*

(b) *Probability to play A*

Figure 5: As a function of the parameter S, we show the energy experienced by the players when placed in an external field given by Eqn. 60, in ***(a)****, and the probabilities the players choose A, in* ***(b)****, when players use a response strategy corresponding to the mixed-strategy Nash equilibrium.*

with H_μ corresponding to the energy given in Eqn. 10, where we take $s = +1$ if $\mu = A$ and $s = -1$ if $\mu = B$. Hence, the probabilities the players pick A are given by

$$P_A^{(1)} = \frac{1}{1+S} = \frac{e^{\beta\gamma_A B}}{e^{-\beta\gamma_A B} + e^{\beta\gamma_A B}} \quad \text{and} \quad P_A^{(2)} = \frac{S}{1+S} = \frac{e^{\beta\gamma_B B}}{e^{-\beta\gamma_B B} + e^{\beta\gamma_B B}}, \tag{59}$$

which gives us

$$B = -\frac{1}{2\beta}\ln(S), \quad \gamma_A = 1, \quad \gamma_B = -1. \tag{60}$$

Note that in the case of game theory we do not have an external magnetic field like we would have in a physics problem. However, one can instead have an external influence, e.g. some form of propaganda, that influences the players' choices in the same way an external magnetic field influences the spin values.

Fig. 5 shows the energy of the players in the external field given by Eqn. 60 and the probability to play A. This shows that the energy when player 1 chooses A is negative, which it should be since A is his preference, and that this energy becomes less negative for increasing S. This also makes, sense, since if S increases, his preference towards A decreases with respect to B. This also shows in his probability to play A, which decreases as S increases.

For player 2 the energy he experiences when playing A is positive, which is also expected from a physics point of view, as it is an unfavorable state. Similar as for player 1, the energy now decreases for increasing S and the probability to play A increases, both which are due to his preference for B becoming less prominent.

By now using Eqn. 12 and plugging in the above mentioned values for B, γ_A and γ_B, we find a "magnetization" per player given by

$$m = \frac{S-1}{S+1}. \tag{61}$$

Note that one would obtain the same result if one were to use Eqn. 16 with $J = 0$.

4.2 Correlated games

With the exception of the response-strategy profile corresponding to the mixed-strategy equilibrium, all strategy profiles lead to correlated final outcomes. We hereby note that it was not possible to find analytical results for the external field B and the coupling constant J for all correlated games. We will here present the results for correlation devices with initial probabilities in regions $\mathrm{I}, \mathrm{II}, \mathrm{III}$ and VII (see Fig. 3), as these have exact solutions.

For the correlated games, we need to use the energy corresponding to Eqn. 15, which can be written in matrix form for two players as

$$H_{\mu\nu} = \begin{pmatrix} -J - B(\gamma_A + \gamma_B) & J - B(\gamma_A - \gamma_B) \\ J + B(\gamma_A - \gamma_B) & -J + B(\gamma_A + \gamma_B) \end{pmatrix}, \tag{62}$$

where we once more take μ, ν equal to 1 if they equal A and equal to -1 if they equal B. Using again Eqn. 56, we now have a set of four equations, which can be solved to determine the interaction strength J, the field B and the couplings γ.

In the region denoted by I we have equilibria for two strategy profiles that are equally good payoff-wise.

We find an external magnetic field given by

$$B = \frac{1}{4\beta} \ln \left(\frac{p_{AA}}{p_{BB}} \right), \tag{63}$$

which is the same for any correlation device. It is how the players couple to this field what changes between the various Nash equilibria. For $\Pi^* = \{\Pi^{(1)}_{1,1}, \Pi^{(2)}_{1,1}\}$, i.e. both players always following the instructions, we find $\gamma_A = \gamma_B = 1$, whereas for $\Pi^* = \{\Pi^{(1)}_{0,0}, \Pi^{(2)}_{0,0}\}$, i.e. both never following the correlation device, we find $\gamma_A = \gamma_B = -1$. Similarly, for $\Pi^* = \{\Pi^{(1)}_{1,1}, \Pi^{(2)}_{0,0}\}$, we find $\gamma_A = -\gamma_B = 1$, while for $\Pi^* = \{\Pi^{(1)}_{0,0}, \Pi^{(2)}_{1,1}\}$ we obtain $\gamma_A = -\gamma_B = -1$. This result shows us that if a player, in equilibrium, follows the instructions of the correlation device with response probabilities $P^{(i)}_{F_A} = P^{(i)}_{F_B} = 1$, he will couple to the field given in Eqn. 63 with $\gamma = 1$, whereas if his equilibrium response probabilities are $P^{(i)}_{F_A} = P^{(i)}_{F_B} = 0$, he will couple with $\gamma = -1$.

For the interaction strength between the players, we find that below the curve given by $p_{BB} = 1 - 2\sqrt{p_{AA}} + p_{AA}$ it is given by

$$J = \frac{1}{2\beta} \ln \left(\frac{1 - p_{AA} - p_{BB}}{\sqrt{2 p_{AA} p_{BB}}} \right), \tag{64}$$

whereas above this curve

$$J = -\frac{1}{2\beta} \ln \left(\frac{1 - p_{AA} - p_{BB}}{\sqrt{2 p_{AA} p_{BB}}} \right). \tag{65}$$

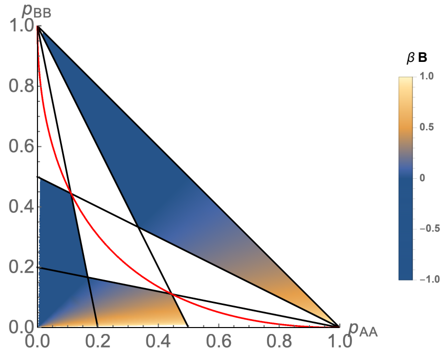
Although at first these seem two different interaction types, one ferromagnetic and one antiferromagnetic, one can check that above the aforementioned curve the argument of the logarithm is smaller than 1, whereas in area VII it is greater than 1, such that in both cases the resulting interaction strength is positive. This implies that we always see an interaction of the ferromagnetic type, i.e. the players want to play either both A or both B, which makes sense, since this yields them the highest payoff.

Considering now players placed on a bipartite lattice; N players in total, half with preference A and half having preference B. We use Eqn. 16 with the above discussed external field B, interaction strength J and coupling constants γ, and find that the magnetization per player is given by

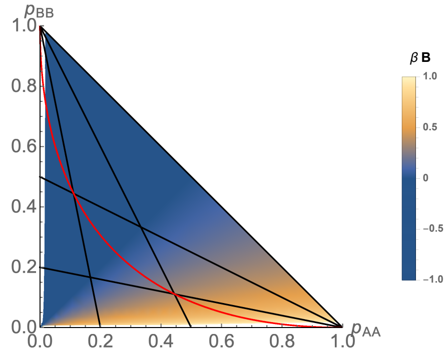
$$m = \frac{p_{AA} - p_{BB}}{(1 - p_{AA} - p_{BB}) \sqrt{3 - \frac{(2p_{AA} - 1)(2p_{BB} - 1)}{(1 - p_{AA} - p_{BB})^2} + \sqrt{\frac{p_{BB}}{p_{AA}}} + \sqrt{\frac{p_{AA}}{p_{BB}}}}}, \tag{66}$$

for all the Nash equilibria discussed here, i.e. the ones corresponding to areas I, II, III and VII in Fig. 3.

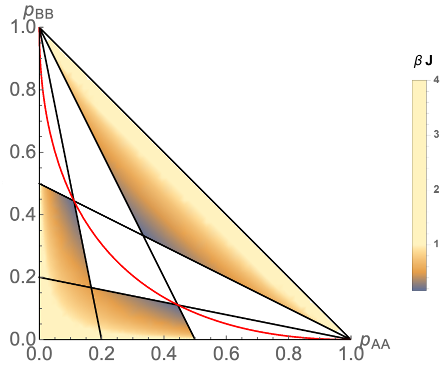
In Fig. 6 we show all our analytical results for the external field, interaction strength and the magnetization per player. Since the external field and magnetization were the same for all



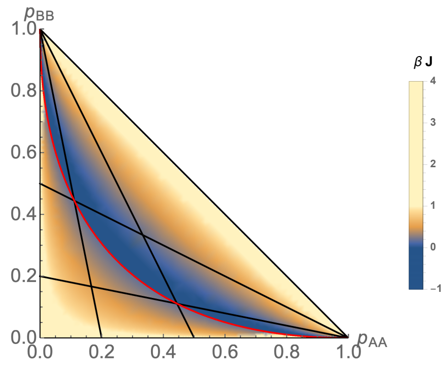
(a) External field (exact)



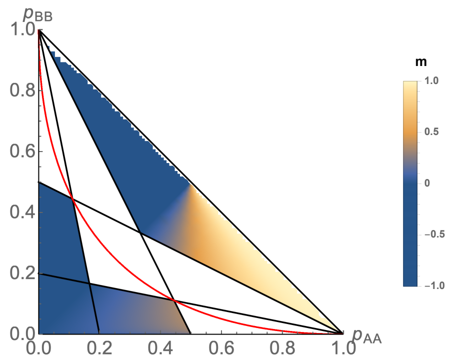
(b) External field (interpolation)



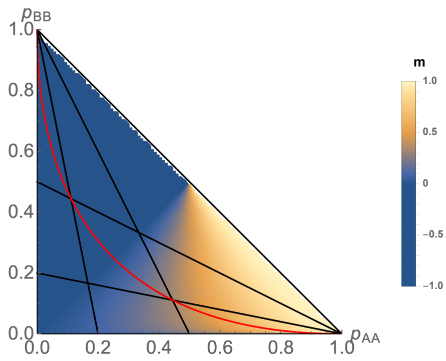
(c) Interaction strength (exact)



(d) Interaction strength (interpolation)



(e) magnetization (exact)



(f) magnetization (interpolation)

*Figure 6: Figures **(a)**, **(c)** and **(e)** show our exact results for the external field B, interaction strength J and the resulting magnetization per player m for the correlated Battle of the Sexes game. Figures **(b)**, **(d)** and **(f)** show the same parameters, but interpolated over the regions we could not solve analytically. For reference, the solid lines of Fig. 3 are also shown. In all figures $S = 1/2$.*

equilibria, and the interaction strength appears to be continuous, we interpolated the analytical results to obtain their values also for parameters for which we could not solve exactly. The interpolation simply consisted of using the same functions we found analytically, and applying them in white areas in Fig. 6. This allows us to describe the Nash equilibria corresponding to areas IV, V and VI using the same bipartite Ising model. However, one needs to keep in mind they are interpolations, hence not exact results. Also, for the interaction strength we find an area where $J < 0$, i.e. antiferromagnetic interactions. However, this area is the same as where the mixed-strategy equilibrium is the only self-consistent solution. The results we found by interpolation can therefore not be used for this specific area (area 0 in Fig. 3). Lastly, the interpolations do not tell us with which coupling constants γ the players couple to the external field.

The fact that we found the same magnetization for all the Nash equilibria, leads us to believe the final dynamics do not differ (on average) between the various equilibria in the response strategy for correlated games, at least for a 1 dimensional bipartite lattice. Since the 1 dimensional Ising model does not show a phase transition, whereas higher dimensional Ising models do, it is likely one would find more interesting dynamics for the Battle of the Sexes on 2 dimensional lattices.

5 Conclusion

To summarize, we analyzed a model by Correia *et al.*[1] when applied to the Battle of the Sexes. We started out with an introduction to game theory, elaborating on possible Nash equilibria in the Battle of the Sexes with pure, mixed and correlated strategies. We then explained the notion of a correlated game; making a correlation device part of the game itself, thus creating games with intrinsic correlations. The model by Correia *et al.*, which we discussed in detail, lets players respond to these correlations by following the instructions of the correlation device only with a certain response probability. Furthermore, we gave an introduction to some methods of statistical physics to deal with spin models, such that we could use these techniques to extend the two-player Battle of the Sexes game to simple network structures.

We then explored the entire probability space of the correlation device to determine all possible Nash equilibria. This led us to the recovery of the pure-strategy, mixed-strategy and correlated equilibria, as well as several previously unknown equilibria. The pure and mixed-strategy equilibria appear to exist for any correlation device, whereas the correlated equilibrium only exists for correlation devices with strong correlations on the diagonal of the probability matrix. For correlation devices with strong off-diagonal correlations, we found equilibria for strategy profiles where one of the players will always follow the instructions of the correlation device, while the other player won't follow given instructions. This is due to the fact that off-diagonal states are unfavorable; if only one player obeys the instructions guiding towards such a state, the players will ultimately reach one of the more favorable states on the diagonal.

To conclude the discussion of the response-strategy for a correlated Battle of the Sexes game, we computed the optimal strategy profiles. The probabilities corresponding to these optimal response-strategy profiles were then used as a starting point to make the connection with statistical physics. We computed the external field and the inter-player interaction strength for some of the Nash equilibria. We found that the external field would have the same (logarithmic) dependence on $p_{\mu\nu}$ for all Nash equilibria, but that the way the players couple to this field was different for the various equilibria. In general, a player would couple to the field with a coupling constant equal to $+1$ if he always obeys the instructions in equilibrium, and couples with a coupling constant equal to -1 if he never follows given instructions. Furthermore, the interaction between players was always of a ferromagnetic type, which once more shows that players would preferably end up both having made the same choice. We used the field and couplings we found to determine the magnetization of one-dimensional networks, specifically players on a bipartite lattice.

To improve the model, there are three things we urge to consider. First of all, we have only taken into account symmetric initial probabilities of the correlation device. However, to generalize, one should allow for any correlation device for which the probabilities sum to unity. This might then lead to even more new Nash equilibria in the probability space of the correlation device, which would now become a three-dimensional volume, instead of a two-dimensional plane. Secondly, one might also consider to use the generalized payoff table we briefly mentioned (Table 2). This introduces yet another asymmetry to the Battle of the Sexes game, leading presumably to different equilibria and response strategies. Lastly, since real-life networks are rarely one dimensional, one should consider higher-dimensional networks. These can be treated using for example the analytical solution for the two-dimensional Ising model,[7] or by using a mean-field approach for random networks such as Erdős-Rényi networks.

As a final remark, we note that it was not possible to find analytical solutions for B and J in the case one of the equilibrium response probabilities is $0 < P^{(i)}_{F_\mu} < 1$. However, from our results, it is to be expected that B and J are the same as we found for the other Nash equilibria, but the constants with which the players couple to B will probably be different, most likely some function of the correlation probabilities $p_{\mu\nu}$. To find these coupling constants, one should resort to numerical methods.

References

[1] A.D. Correia and H.T.C. Stoof. Nash equilibria in the response strategy of correlated games. 2018, arXiv:1809.03860.

[2] J.J. Broere, V.W. Buskens, J. Weesie, and H.T.C. Stoof. Network effects on coordination in asymmetric games. *Scientific Reports*, 7(1):17016, 2017.

[3] D. Fudenberg and J. Tirole. *Game Theory*. Mit Press. MIT Press, 1991.

[4] D.J. Griffiths. *Introduction to Quantum Mechanics (2nd Edition)*. Pearson Prentice Hall, 2nd edition, 2005.

[5] Walther Gerlach and Otto Stern. Der experimentelle nachweis der richtungsquantelung im magnetfeld. *Zeitschrift für Physik*, 9(1):349–352, Dec 1922.

[6] S.J. Blundell and K.M. Blundell. *Concepts in Thermal Physics*. OUP Oxford, 2010.

[7] K. Huang. *Statistical Mechanics, 2nd ed.* Wiley India Pvt. Limited, 2008.

[8] Ernst Ising. Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik*, 31(1):253–258, 1925.

A Coefficients of the expected payoff

By introducing probability vectors $\mathbf{p} = (p_{AA}, p_{AB}, p_{BA}, p_{BB})$ and $\mathbf{p}^R = (p^R_{AA}, p^R_{AB}, p^R_{BA}, p^R_{BB})$, one can find the renormalized probabilities by the transformation

$$\mathbf{p}^R = \mathbf{A}\mathbf{p}, \tag{67}$$

where the transformation matrix is given by the direct product $\mathbf{A} = \mathbf{A}^{(1)} \otimes \mathbf{A}^{(2)}$ with

$$\mathbf{A}^{(i)} = \begin{pmatrix} p_{F_A}^{(i)} & p_{NF_B}^{(i)} \\ p_{NF_A}^{(i)} & p_{F_B}^{(i)} \end{pmatrix}. \tag{68}$$

For example, by using this transformation one finds that the renormalized probability to get (A, A) is given by $p_{AA}^R = p_{AA} p_{F_A}^{(1)} p_{F_A}^{(2)} + p_{AB} p_{F_A}^{(1)} p_{NF_B}^{(2)} + p_{BA} p_{NF_B}^{(1)} p_{F_A}^{(2)} + p_{BB} p_{NF_B}^{(1)} p_{NF_B}^{(2)}$. Plugging all renormalized probabilities into Eqn. 7 and collecting terms, one finds the coefficients

$$C_A^{(1)} = p_{AA} \left((S+1) P_{F_A}^{(2)} - S \right) - p_{AB} \left((S+1) P_{F_B}^{(2)} - 1 \right), \tag{69a}$$

$$C_B^{(1)} = p_{BB} \left((S+1) P_{F_B}^{(2)} - 1 \right) - p_{BA} \left((S+1) P_{F_A}^{(2)} - S \right), \tag{69b}$$

$$C_C^{(1)} = p_{\mathrm{AA}} S \left(1 - P_{F_A}^{(2)} \right) + p_{AB} S P_{F_B}^{(2)} + p_{BA} P_{F_A}^{(2)} - p_{BB} \left(P_{F_B}^{(2)} - 1 \right) \tag{69c}$$

and

$$C_A^{(2)} = p_{AA} \left((S+1) P_{F_A}^{(1)} - 1 \right) - p_{\mathrm{BA}} \left((S+1) P_{F_B}^{(1)} - S \right), \tag{70a}$$

$$C_B^{(2)} = p_{BB} \left((S+1) P_{F_B}^{(1)} - S \right) - p_{\mathrm{AB}} \left((S+1) P_{F_A}^{(1)} - 1 \right), \tag{70b}$$

$$C_C^{(2)} = -p_{AA} \left(P_{F_A}^{(1)} - 1 \right) + p_{AB} S P_{F_A}^{(1)} + p_{BA} P_{F_B}^{(1)} + p_{BB} S \left(1 - P_{F_B}^{(1)} \right). \tag{70c}$$

B Equilibria in the response-strategy space

In Fig. 7 through 14 we show the Nash equilibria in the 4-dimensional space of response strategies, for each area in the $p_{AA} - p_{BB}$ plane as indicated in Fig. 3. Each point in a player's strategy space corresponds to an equilibrium. The color of the point shows with which point in the other player's strategy space it forms an equilibrium. For all figures we used $S = 1/2$, just as in the main text.

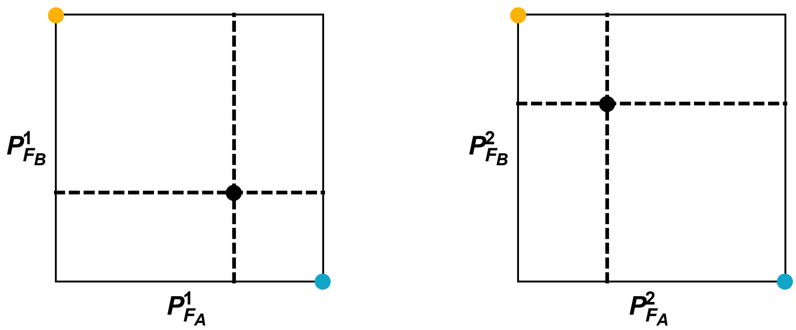


Figure 7: Equilibrium values of response probabilities in region 0.

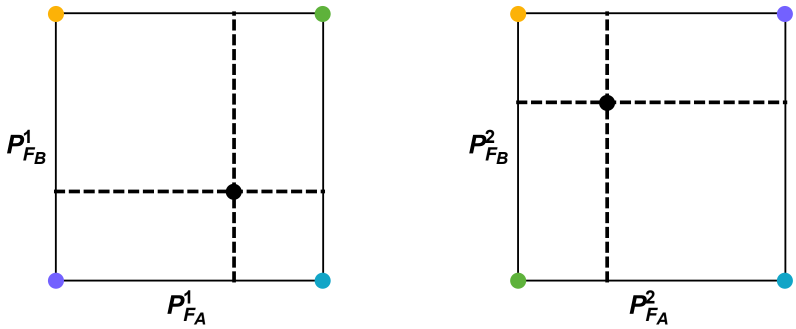


Figure 8: Equilibrium values of response probabilities in region I.

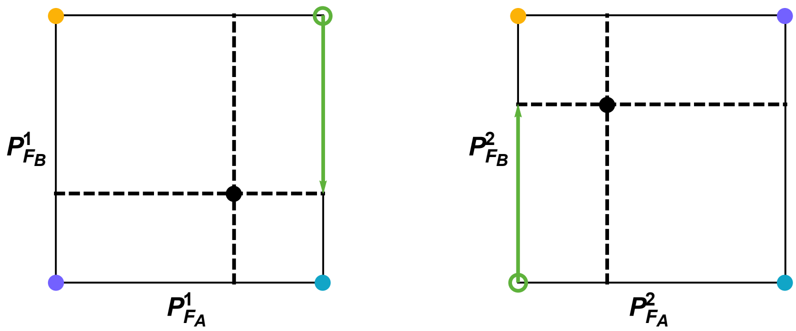


Figure 9: Equilibrium values of response probabilities in region II.

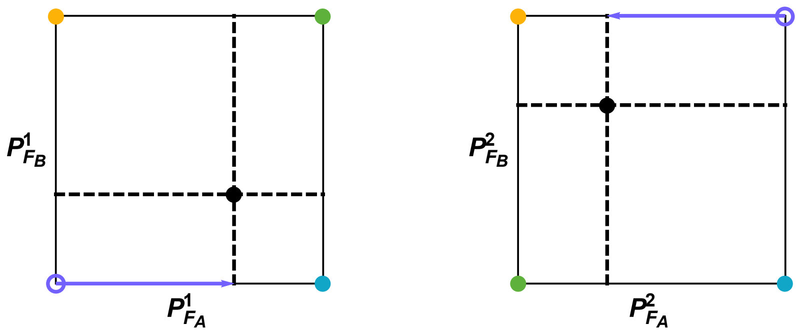


Figure 10: Equilibrium values of response probabilities in region III.

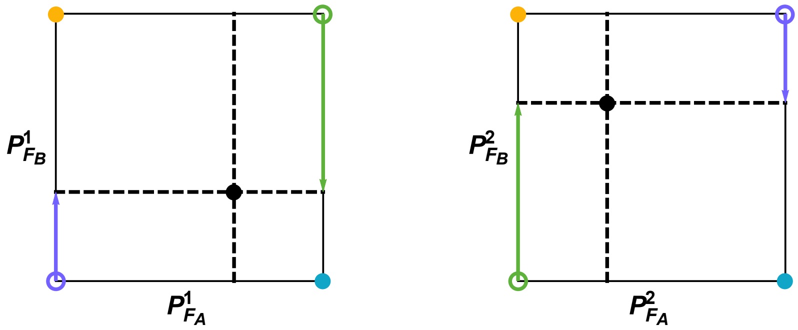


Figure 11: Equilibrium values of response probabilities in region IV.

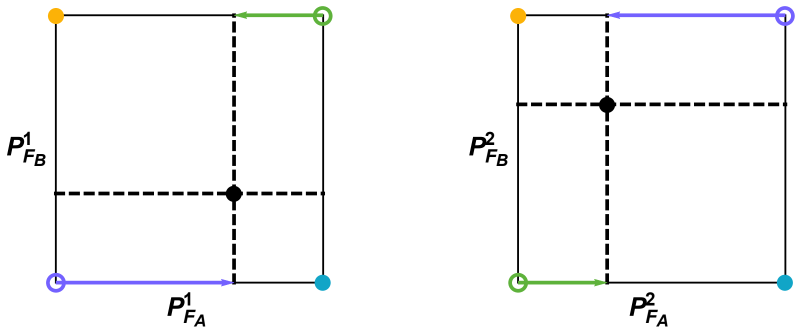


Figure 12: Equilibrium values of response probabilities in region V.

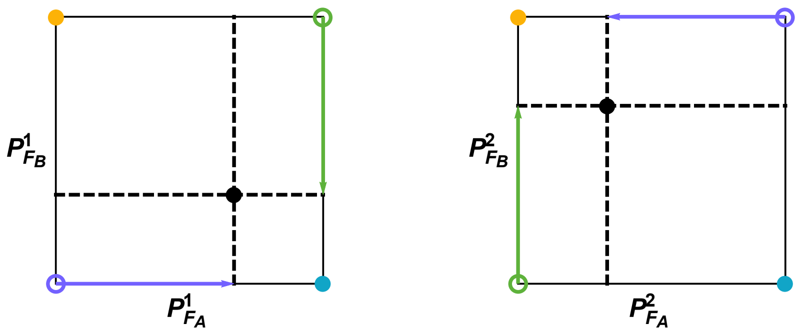


Figure 13: Equilibrium values of response probabilities in region VI.

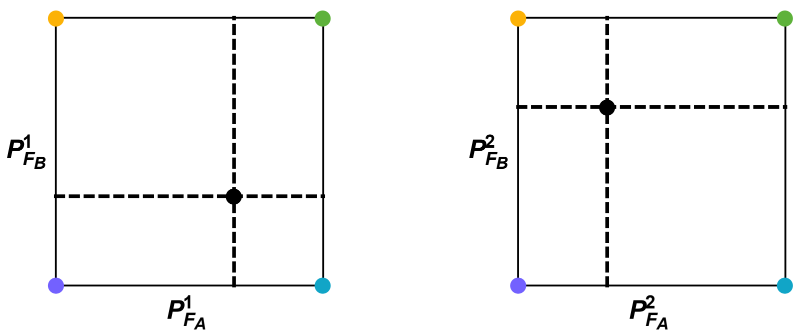
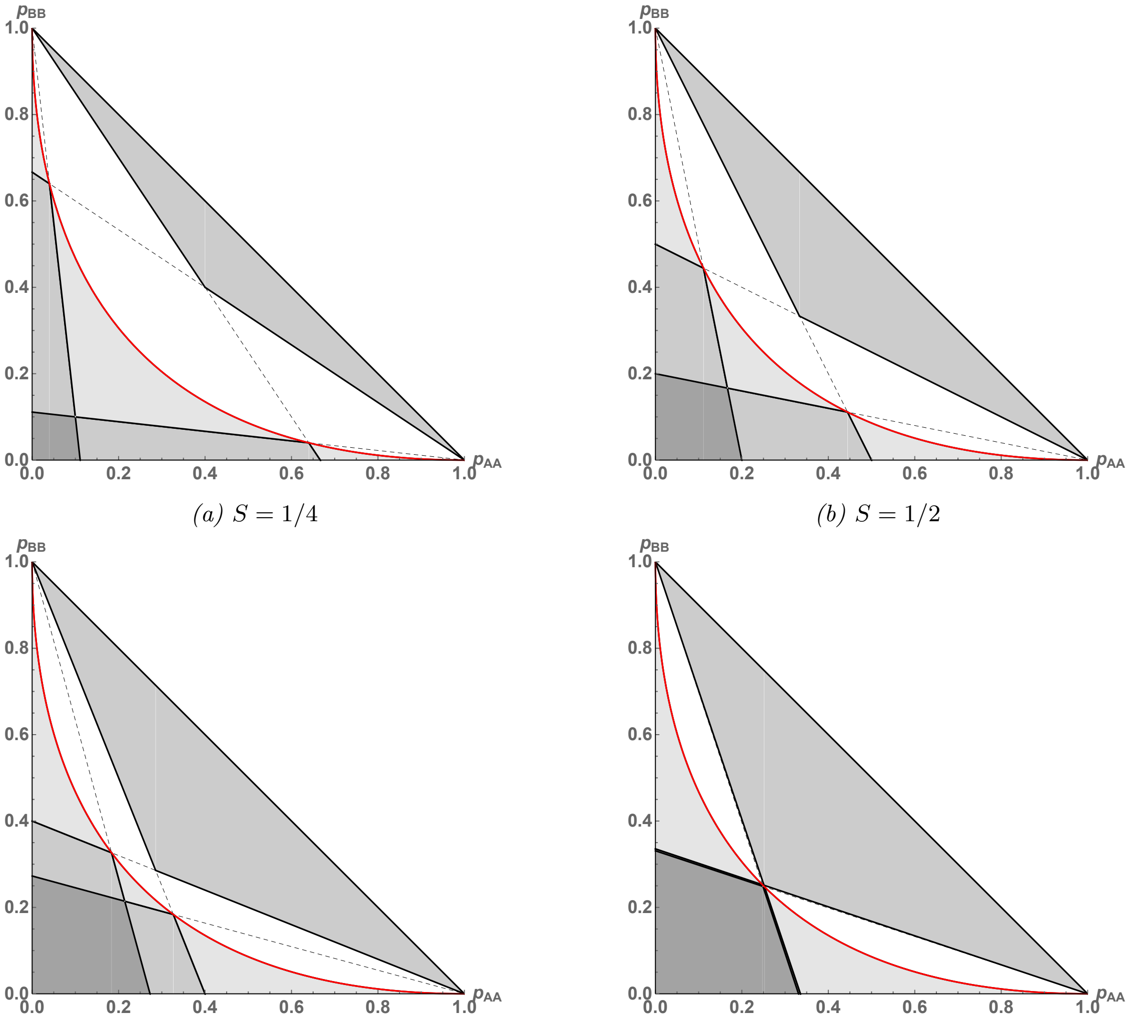


Figure 14: Equilibrium values of response probabilities in region VII.

C Nash equilibria for various S

Fig. 15 shows how the areas wherein the various Nash equilibria exist, change as we increase the parameter S.

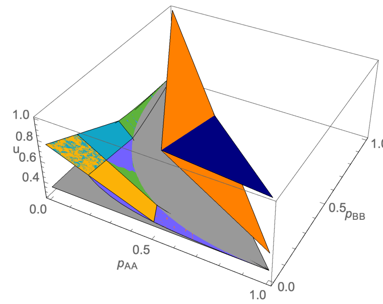


(a) $S = 1/4$ (b) $S = 1/2$

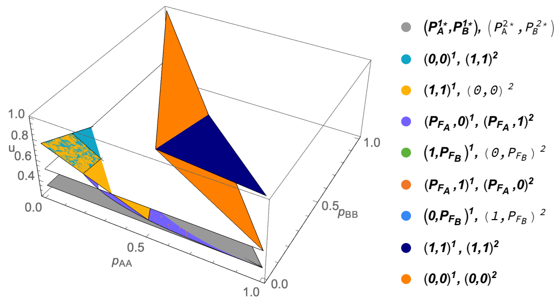
(c) $S = 3/4$ (d) $S = 1$

*Figure 15: **(a) - (d)** Plots of the Nash equilibria in the probability space of the correlation device for increasing S.*

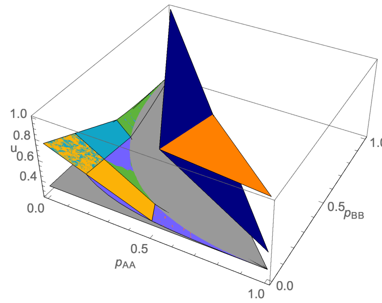
D 3 Dimensional plots of payoff



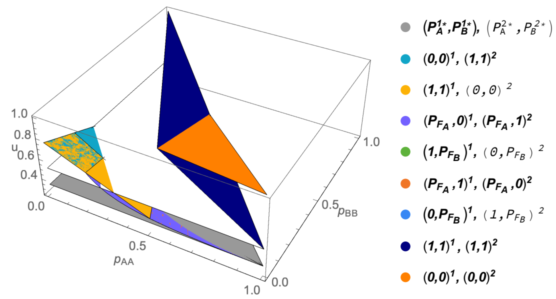
(a) Player 1



(b) Player 1



(c) Player 2



(d) Player 2

Figure 16: ***(a)*** *and* ***(c)*** *show a 3D plot of the payoffs for each Nash equilibrium in the probability space of the correlation device. In* ***(b)*** *and* ***(d)*** *we see the same, but now we have included the constant payoff* S*, i.e. the payoff one would get from consistently going to one's least-preferred option. In all figures* $S = 1/2$.