Vrije Universiteit Brussel

Faculty of Science and Bio-Engineering Sciences
Department of Computer Science

Emergence of Honest Signaling through Learning and Evolution

Thesis submitted to obtain the degree of Doctor of Philosophy in Sciences

David Catteeuw

Supervisor: prof. dr. Bernard Manderick

January, 2015

2015 Uitgeverij VUBPRESS Brussels University Press
VUBPRESS is an imprint of ASP nv (Academic and Scientific Publishers nv)
Keizerslaan 28
B-1000 Brussels
Tel. +32 (0)2 289 26 50
Fax +32 (0)2 289 26 59
E-mail: info@vubpress.be
www.vubpress.be

ISBN 978 90 5718 294 5
NUR 958
Legal deposit D/2014/11.161/008

To my dearest Adeline & Éline

Ph.D. Committee

Supervisor	prof. dr. Bernard Manderick
Internal	prof. dr. Ann Nowé prof. dr. Beat Signer prof. dr. Jean Paul Van Bendegem
External	prof. dr. Philippe De Wilde University of Kent, Canterbury, United Kingdom
	prof. dr. Tom Lenaerts Université Libre de Bruxelles, Brussels, Belgium
	prof. dr. Francisco C. Santos Universidade de Lisboa, Lisbon, Portugal

Acknowledgments

First of all, I'd like to thank Bernard, my supervisor, for providing feedback, proofreading, many inspiring discussions, and pointing me to the relevant literature again and again. I also realize that without you, I would not had this amazing opportunity to do research in all freedom. It was a very unique and instructive period indeed. Thanks for that.

I am also grateful to my co-authors. Together with Joachim, I wrote my first paper on signaling and together with The Anh, I wrote two other papers on signaling near the end of this period. Both of you were inspiring, often looking at things from a different point of view than I did. Without your expertise and advice I'd not stand where I am today. Together with Yailen, Bert, Ann, and later with Madalina, I did some research that did not fitted into this text, but that I enjoyed nevertheless. Thank you all.

I'd like to thank the members of the Ph.D. committee for investing their precious time and giving insightful comments. Especially, Francisco for all those flattering compliments on my text and Tom for the many discussions during the game theory research meetings we've had now for more than a year.

Writing a thesis is not just about work, but also about a pleasant working environment, inspiring (and sometimes ... less inspiring) research meetings, going to conferences in good company, 'after-hours' drinks, BBQs, and other (often improvised) festivities. That was only possible due to all of you: Ann, Anna, Abdel, Bart, Bernard, Bert, Cosmin, 'other' David, Dip, Frederik, Ioannis, Ivan, Ivomar, Jean-Sébastien, Jelena, Joachim, Jonatan, Joris, Katrien, Kevin, Kevin, Kristof, Lara, Luis, Maarten, Marjon, Matteo, Mike, Pasquale, Peter, Pieter, Pieter, Ruben, Saba, Steven, Stijn, Sven, The Anh, Tim, Tom, Yailen, Yann-Aël, Yann-Michaël. Some of you made my life as a teaching assistant a lot easier by taking a lot of workload of my back. I truly appreciate that!

Last but not least, I'd like to thank the home front. First of all, my parents for their huge and unconditional support. Without them, I would not even had this great opportunity. I'd like to thank my friends, especially Niko and Alexander who made decompressing during the weekends a lot of fun. I regret I don't have as much time as a few years ago. And of course, my wife and daughter. Adeline, you are definitely an important part of this work. We met when I had merely started. Time flies and I know you're glad this is over. You often had to cope with my absence, because I was abroad, excluded in my office late at night and during weekends, or simply because my mind was elsewhere. You supported me through the more stressful moments and gave me a lot of confidence. Thanks! Éline, you don't realize it but we've signaled a lot this past year. I am sorry if I did not always respond honestly ;-).

Brussels, December 2014.

Abstract

The emergence of honest (or reliable) signaling is a *multi-disciplinary* problem. Linguists and philosophers have long wondered how conventions, such as human language, can emerge without a pre-existing language. Biologists noticed that the many signals in nature can only exist because they are honest. Otherwise they would be ignored and so, not worth the trouble sending. Economists created a real breakthrough by recognizing that many interactions are characterized by private information—where one party knows more than the other—and signals may, or may not, reveal that information. It explains, for example, why the free market does not work for health insurance: those willing to buy costly insurance are most likely those who expect to need it the most. I contributed to this research in three domains: common interest; costly signals; and costly, social punishment.

One reason why signals are honest is *common interest*: both the sender and the receiver of the signal benefit from conveying the correct information. Under common interest, the only question that remains is how a signal acquires its meaning. One explanation that may also explain the origins of language is that this happens by chance. My findings support this idea. In Chapter 3,

- I introduce a new behavioral rule, called 'win-stay/lose-inaction' or 'WSLI:' initially play random, repeat forever what was once successful. When two repeatedly interacting players apply WSLI they always end up signaling honestly in all Lewis signaling games (the standard game-theoretic model to study the emergence of signaling under common interest). I prove that the expected number of iterations is only polynomial in the number of signals. No such algorithm was known before.

- I show that three well-known reinforcement learning algorithms (Q-

learning, Roth-Erev learning, and Learning Automata) behave exactly like WSLI in Lewis signaling games for certain parameter configurations.

- While WSLI is not robust to errors, these reinforcement learning algorithms are robust for certain parameter configurations and still reach honest signaling in a polynomial number of iterations.

Economists and biologists independently discovered that when interests conflict signals may be honest if they are costly. This is known as the *'handicap principle'* and is almost exclusively studied assuming infinite populations and by means of static equilibrium analyses—verifying if honest signaling is an equilibrium while ignoring the dynamics that may or may not lead to it. In Chapter 4, I apply learning and evolutionary dynamics in finite populations to the Philip Sidney game:

- In many cases where honest signaling is an equilibrium, it does not emerge: equilibrium analyses wrongfully predict honest signaling.
- Dynamics reveal (partially) honest signaling in some cases where it is not an equilibrium: equilibrium analyses fail to predict (partially) honest signaling.

Costly, social *punishment* is known to promote the evolution of cooperation but its effect on the evolution of honest signaling is merely studied. In Chapter 5, I distinguish four ways of deviating from honest signaling: the sender can lie or be timid and the receiver can be greedy or worried. I extend the Philip Sidney game to explicitly allow for punishment of such behavior and study its effect on the evolution of honest signaling:

- When punishment targets lying individuals, honest signaling emerges also for cost-free signals. So, punishment provides an alternative to the handicap principle.
- When punishment targets greedy individuals, honest signaling emerges also in cases with strong conflicts, similar to the punishment of defectors to promote cooperation.

- The evolution of honest signaling does not benefit from punishment of timid or worried individuals.

Samenvatting

Het ontstaan van eerlijke (of betrouwbare) signalen is een probleem in *meerdere onderzoeksdomeinen*. Taalkundigen en filosofen hebben lang gezocht hoe conventies, zoals menselijke taal, kunnen ontstaan zonder vooraf bestaande, gemeenschappelijke taal. Biologen hebben ontdekt dat veel signalen in de natuur enkel bestaan omdat ze eerlijk zijn. Zoniet zouden ze genegeerd worden en dus nutteloos zijn. Economisten hebben een echte doorbraak teweeg gebracht toen ze ondervonden dat veel interacties gekenmerkt worden door private informatie—een partij weet meer dan de anderen—en signalen kunnen eventueel die informatie overbrengen. Het verklaart onder andere waarom de vrije markt niet werkt voor ziekteverzekeringen: zij die een dure verzekering willen betalen zijn juist diegene die vermoeden het meest nodig te hebben. Ik droeg bij aan dit onderzoek in drie domeinen: gemeenschappelijke belangen, kostelijke signalen en kostelijk, sociaal straffen.

Eén reden waarom signalen eerlijk zijn is *gemeenschappelijke belangen*: zowel de zender als de ontvanger van het signaal heeft baat bij het correct overbrengen van de informatie. Bij gemeenschappelijke belangen is de enige vraag hoe signalen hun betekenis krijgen. Een verklaring die ook de oorsprong van taal kan verklaren is dat dit gebeurt door toeval. Mijn onderzoek ondersteunt deze verklaring. In Hoofdstuk 3 doe ik het volgende:

- Ik definieer een nieuwe gedragsregel, 'win-stay/lose-inaction' of kortweg 'WSLI': kies willekeurige acties en herhaal voor altijd de eerste succesvolle actie. Wanneer twee individuen herhaaldelijk interageren en WSLI toepassen, dan zullen ze uiteindelijk optimaal en eerlijk communiceren in elk 'Lewis signaling game' (het standaard spel theoretisch model om het ontstaan van signalen bij gemeenschappelijke belangen te bestuderen). Ik bewijs dat het verwachte aantal interacties slechts een veelterm is van het aantal signalen. Zo'n algoritme was nog niet bekend.

- Ik toon aan dat drie bekende 'reinforcement learning' (leren door middel van versterking) algoritmes (Q-leren, Roth-Erev leren en leerautomaten) zich identiek gedragen als WSLI in Lewis signaling games voor bepaalde parameters.

- Terwijl WSLI niet robuust is voor fouten, zijn deze reinforcement learning algoritmes dat wel voor bepaalde parameters en kunnen ze nog steeds optimaal communiceren na een veelterm van interacties.

Economisten en biologen ontdekten onafhankelijk van elkaar dat bij tegenstrijdige belangen signalen toch eerlijk kunnen zijn als ze kostelijk zijn. Dit staat bekend als het *handicap principe*. Het is bijna uitsluitend bestudeerd in de veronderstelling van oneindige populaties en door middel van statische analyse van evenwichten—nagaan of het gebruik van eerlijke signalen in evenwicht is maar negeren of dynamische processen er wel of niet toe leiden. In Hoofdstuk 4 pas ik dynamische processen, gebaseerd op leren en evolutie, toe op eindige populaties en het Philip Sidney spel:

- In veel gevallen waar het gebruik van eerlijke signalen een evenwicht vormt, ontstaat het niet: statische analyse van evenwichten voorspelt verkeerdelijk het gebruik van eerlijke signalen.

- Dynamische processen leiden tot het (gedeeltelijk) gebruik van eerlijke signalen in bepaalde gevallen waar het niet in evenwicht is: statische analyse van evenwichten faalt om het (gedeeltelijk) gebruik van eerlijke signalen te voorspellen.

Kostelijk, sociaal *straffen* staat bekend om samenwerking te bespoedigen maar het effect op het ontstaan van eerlijke signalen werd nauwelijks bestudeerd. In Hoofdstuk 5 onderscheid ik vier manieren om af te wijken van het gebruik van eerlijke signalen: de zender kan liegen of te bescheiden zijn en de ontvanger kan hebzuchtig of overbezorgd zijn. Ik breid het Philip Sidney spel uit zodat zo'n gedrag expliciet gestraft kan worden en bestudeer het effect daarvan op het ontstaan van eerlijke signalen:

- Wanneer leugenaars gestraft worden, kunnen ook goedkope signalen eerlijk zijn. Dus, straffen is een alternatief voor het handicap principe.

- Wanneer hebzuchtige individuen gestraft worden, ontstaan er eerlijke signalen in situaties met zeer tegenstrijdige belangen, gelijkaardig aan hoe straffen van zij die niet meewerken samenwerking bespoedigt.

- Het ontstaan van eerlijke signalen wordt niet geholpen door het straffen van te bescheiden of overbezorgde individuen.

Contents

Chapter 1

Introduction

This chapter introduces the emergence of honest signaling, summarizes my contributions to the subject, and lists my publications.

Contents

1.1 Signaling

Signaling, or conveying information through arbitrary symbols and actions, is everywhere in nature. Humans signal extensively, not only in the form of human language but also by means of facial expressions (Ekman, 1992) and unconscious body movements (Pentland, 2010). Two classic examples in other animals are the vervet monkey's alarm calls (Seyfarth et al., 1980) and the honey bee's waggle dance (Riley et al., 2005; Von Frisch, 1967).

Example 1.1. When a vervet monkey spots a predator, he uses an alarm call to warn his group members. Which alarm call he uses depends on the type of the predator: raptor, snake, or leopard. Vervet monkeys also respond differently to different alarm calls: for raptors they hide in the bushes, for snakes they stand up and look around, and for leopards they quickly climb in the nearest tree.

Example 1.2. Honey bees communicate the direction and distance of flowers to their colony members by performing a so-called 'waggle dance' (Figure 1.1). The bees repeatedly waggle in the direction of the flowers and go back to their starting position alternately turning left and right. When the dance is performed on a vertical surface, the upward direction refers to the direction of the sun and the angle between the upward direction and the direction of the waggle correlates with the angle between the direction of sun and the direction of the flowers. The distance from the hive to the flowers correlates with the duration of the waggle.

Here is a more general description of signaling. Signaling is a interaction between two agents. One agent, like the monkey that spotted a leopard or the bee that just discovered flowers full of nectar, has some *private information*: he knows something that the other does not know. The informed agent can signal to the uninformed one and share his information so that the uninformed agent can respond appropriately to the current situation. An agent's private information is also called his '*type*.'

For clarity, let me contrast this with two examples of what signaling is not. Conveying information *unwillingly* is not signaling. For example, a

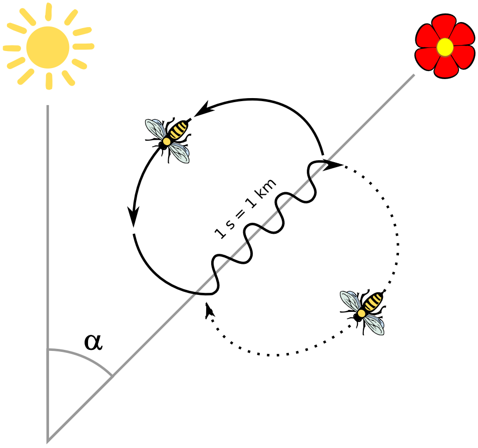


Figure 1.1: The honey bees' waggle dance communicates the location and the direction of flowers to their colony members. See Example 1.2 for more information. (This figure was reproduced from (Jüppsche, 2011) with slight modifications.)

mouse that moves through the grass and makes noise, gives away its location to a nearby predator, like a cat. I consider such noise not a signal from the mouse to the cat, but rather an (unfortunate) side effect of the mouse's behavior (moving through the grass). Maynard Smith and Harper (2003) call this a 'cue.'

Forcing another agent to behave in some way is not signaling either. For example, to get someone to leave the room, you can point with your finger towards the door or push him towards and through the door. The former is signaling, but the latter is not. Signaling informs the other agent, such that he decides for himself what the best action is.

Signaling must benefit both the sender and the receiver of the signal. The receiver must benefit from the information gained otherwise he would ignore the signal and the signal must generate a response in its receiver that benefits the sender otherwise he would have no interest in sending signals. This requires that both the sender and receiver attach the same meaning to the same signal and that the signals are reliable. This reliability can be disrupted if the communication channel is too noisy, which is studied in information theory (Shannon, 1948), or if the agents are not 'honest.'

Since this thesis concentrates on honesty, most of it assumes there is no *noise* so that the signal that the sender intends to send is exactly the same as what the receiver observes. Section 3.5 deviates from this setting and introduces noise, not just in sending and receiving signals but in general. There is a small probability that an agent misinterprets a situation or makes a mistake.

Honesty refers to the agents' intentions. First, the sender's signals should correlate with his type, or private information, so that as much information is revealed as possible. This happens if the sender always uses the same signal for the same type and every signal is used for only one type. I will assume that there are exactly as many signals as types so that this is always possible. If there are too many, some of them could go unused or synonyms may emerge (Skyrms, 2010, ch. 9). If there are not enough signals, agents could invent new ones (Skyrms, 2010, ch. 10). Second, an honest signal should be maximally informative about the receiver's response and that response should correspond to the sender's type. Which response corresponds to which private information will depend on the exact model, but in this thesis there will always be a unique one. Again, this assumption seems reasonable (Skyrms, 2010, ch. 9). If there are too many responses, some of them will go unused. If there are not enough responses, then there is no use in distinguishing between some of the types and they can be mapped to the same signal and response.

This rises two questions:

1. How do arbitrary symbols and actions acquire meaning?
2. Why are signals reliable or honest?

I call these two questions combined: 'the emergence of honest signaling.'

The emergence of honest signaling is a multi-disciplinary problem. Linguists and philosophers (Lewis, 1969; Skyrms, 2010; Steels, 1999) have long wondered how conventions, such as human language, can emerge without a preexisting language.

Biologists noticed that animal signals can only exist because they are honest. Otherwise they would be ignored and not worth the trouble sending. Zahavi (1975, 1977) proposed that signals can be honest if they are costly to send. The typical example is that of the peacock's tail (Petrie et al., 1991).

Example 1.3. The peacock has a large tail that signals his quality as a parent to potential mating partners. Only the strongest peacocks can afford the largest tails, because a large tail makes the peacock less agile and increases the risk of being caught by predators. A large tail is also costly because it requires a lot of resources that could otherwise be spent to defeat diseases and parasites. These costs make the signal honest and peahens trust it: they prefer males with longer tails.

Economists Akerlof (1970) and Spence (1973) created a real breakthrough by recognizing that many interactions are characterized by private information. With an example of the market of second hand cars Akerlof (1970) illustrates how the free market may collapse under private information.

Example 1.4. Higher quality cars deserve a higher price than lower quality cars; but since the buyer does not know the quality of the second hand car, a dishonest seller can sell a bad second hand car for the price of a good one. This motivates buyers to offer a lower price and owners of good second hand cars will step out of the market because their vehicle is worth more than the market price. When good quality cars leave the market, the average quality of a second hand car drops, which lowers the average price offer again, and so on. In the end, no one is selling his second hand car even though many buyers and sellers could benefit from trading.

The problem that the quality of the supply decreases with decreasing price, is called *adverse selection* and is caused by private information. In some cases, it can be avoided by signaling. Spence (1973) showed how this may work in the job market.

Example 1.5. The job market is sensible to adverse selection, because an employer cannot directly observe a potential employee's productivity. If he decreases the wage offer, he will no longer attract the best candidates. But, candidates can signal their productivity with a university degree. Education is costly because of the time spent at university. Less skilled individuals will not invest in higher education, since they will require too much time to graduate, while highly skilled individuals need less time. Employers trust the signal and offer higher wages to individuals with higher degrees.

This thesis contributes to a better understanding of the emergence of signaling in two ways:

1. *How* do signals emerge?
2. *Why* are signals honest?

I first discuss these two problems separately. In the next section, I summarize my contributions.

1.1.1 How do signals emerge?

For a signal to emerge, an otherwise arbitrary symbol or action must acquire a meaning. It must refer to the same object or concept for both the sender and receiver of the signal. There are roughly three ways *how* arbitrary symbols and actions may acquire meaning: prearrangement, focal points, and chance.

Some signals have *prearranged meanings*. The meaning of traffic signs, mathematical symbols, and hand signals in financial trading floors was decided and agreed upon. Prearrangement requires a preexisting common language, so it cannot explain the origins of language, the vervet monkeys' alarm calls (Example 1.1), or the honey bees' waggle dance (Example 1.2).

Another explanation is that some signals have an obvious or *natural meaning*, a so-called 'focal point' (Lewis, 1969; Schelling, 1960). When a dog shows his teeth, the meaning is obvious: he is ready to attack and will bite if he must. In many cases, the natural meaning requires some common knowledge and the capacity to interpret the signal's context. While the dog showing his teeth, clearly expresses "I will bite you," a human showing his, is conveying a more friendly message: "I am happy." A good example of a context-dependent signal is the ‘thumbs up’ sign. Scuba divers use the signal to indicate they will go back to the surface (Recreational Scuba Training Council, 2005), a broker on a financial trading floor uses it to indicate an order is filled (Chicago Mercantile Exchange, 2006), sometimes it means “I am OK,” and in other contexts it has still other meanings.

The origins of language and signals such as smiling, can only be explained as follows: signals acquire their *meaning by chance* (Skyrms, 2010). A signal’s meaning is a convention. It does not matter which signal has which meaning, what matters is that everyone uses the same signal for the same meaning. This thesis supports this idea and demonstrates on several occasions how meaning emerges from random processes within learning individuals (Chapter 3 and Section 4.3) or evolving populations (Section 4.4 and Chapter 5).

1.1.2 Why are signals honest?

All signals observed in nature are honest (on average), because dishonest signals would be ignored, thus become useless, and finally unused. This thesis studies three reasons why signals are honest: common interest, costly signals, and punishment. Számadó (2011) lists some extra possibilities.

The most obvious reason is *common interest* between the sender and the receiver: it is in the sender’s best interest to correctly convey his information and in the receiver’s best interest to correctly respond to the signal. Common interest is the topic of Chapter 3.

Unfortunately, the interests of interacting agents often conflict. Pursuit-deterrent signals seem to be obvious examples of signals used in conflict situations (Hasson, 1991). When a gazelle spots an approaching cheetah, it stots—jumping up and down in place—instead of running away. Some birds, like skylarks, sing while hunted by a predator. Singing and stotting clearly does not help prey to escape predators. On the contrary, it consumes precious energy and oxygen needed to escape, so they are probably signals meaning: “Do not bother chasing me. I am so fast I can afford to waste time and energy by stotting/singing.”

Biologist Zahavi (1975, 1977) and economist Spence (1973) independently discovered that signals can still be honest under conflict of interest if they are costly. This is known as the '*handicap principle.*' More specifically, signals must be costly such that, a signal has a lower cost or a higher benefit when it is used honestly than when it is used dishonestly. A typical example in biology is the peacock's tail (Example 1.3) and in economics, the job market (Example 1.5). Costly signals is the topic of Chapter 4.

Chapter 5 studies the effects of punishment on the emergence of honest signaling. *Punishment*, even when costly to the punisher, is known to promote the evolution of cooperation (Boyd and Richerson, 1992) and may also promote the evolution of honest signaling provided there is a possibility to verify whether a signal was truthful or not. An example of punishment in a signaling context is found rhesus macaques (Hauser and Marler, 1993).

Example 1.6. Rhesus macaques use food calls to alert group members when food is found. Individuals that find food and refrain from sending food calls (using the signal 'quiet' with the meaning 'no food') are punished by their group members whenever they are discovered.

1.2 Overview and Contributions

The thesis is mostly based on four publications. In the overview below, I mention which publications relate to each of the chapters. Section 1.3 provides a complete list of my publications.

To study the emergence of signaling, I rely on *game theory* which is a mathematical framework that models interactions between agents, such as signaling, by means of games. Game theory predicts the outcome of an interaction by identifying equilibria—behavior from which no agent wants to deviate. Some solution concepts verify equilibria without considering any dynamics that may, or may not, lead to them, while other solution concepts rely explicitly on dynamical processes that model learning or evolution. Chapter 2 provides the minimal background on game theory needed to understand the rest of the text. In later chapters (for example Chapter 4) I will contrast the results from learning and evolution with those from static equilibrium analyses.

Chapter 3 studies the emergence of signaling under *common interest* and is based on

> David Catteeuw and Bernard Manderick (2014). "The Limits and Robustness of Reinforcement Learning in Lewis Signaling Games." In: *Connection Science* 26.2, pp. 161–177.

When both sender and receiver benefit from conveying the correct information and responding appropriately, the only question that remains is how a signal acquires its meaning. My findings support the idea that this happens by chance, as advocated by Skyrms (2010):

- I introduce a new behavioral rule, called 'win-stay/lose-inaction' or 'WSLI:' initially play random and repeat forever what was once successful. When two repeatedly interacting players apply WSLI they always end up signaling honestly in all Lewis signaling games (the standard game-theoretic model to study the emergence of signaling under common interest). I prove that the expected number of iterations is only polynomial in the number of signals. No such algorithm was known before.

- I show that three well-known reinforcement learning algorithms (Q-learning, Roth-Erev learning, and learning automata) behave exactly like WSLI in Lewis signaling games for certain parameter configurations.

- While WSLI is not robust to errors, these reinforcement learning algorithms are robust for certain parameter configurations and still reach honest signaling in a polynomial number of iterations.

Chapter 4 is based on

> David Catteeuw and Bernard Manderick (in press). "Honesty and deception in populations of selfish, adaptive individuals." In: *The Knowledge Engineering Review* 31.2

and

> David Catteeuw, Bernard Manderick, and The Anh Han (2013). "Evolutionary Stability of Honest Signaling in Finite Populations." In: *Proceedings of the IEEE Congress on Evolutionary Computation.* Ed. by Luis Gerardo de la Fraga and Carlos A. Coello Coello. Cancun, Mexico: IEEE Computer Society, pp. 2864–2870.

It studies the *handicap principle*: when interests conflict, signals can be honest only if they are costly. Whereas most of the literature assumes infinite populations and considers only static equilibrium analyses—verifying if honest signaling is an equilibrium while ignoring the dynamics that may or may not lead to it—I consider both evolutionary and learning *dynamics* in *finite* populations and find some surprising results:

- In many cases where honest signaling is an equilibrium, it does not emerge: equilibrium analyses wrongfully predict honest signaling.
- Dynamics reveal (partially) honest signaling in some cases where it is not an equilibrium: equilibrium analyses fail to predict (partially) honest signaling.

Costly, social punishment is known to promote the evolution of cooperation but its effect on the evolution of honest signaling is merely studied. In Chapter 5, based on

> David Catteeuw, The Anh Han, and Bernard Manderick (2014a). "Evolution of Honest Signaling by Social Punishment." In: *Proceedings of the 2014 Genetic and Evolutionary Computation Conference.* Ed. by Christian Igel and Dirk V. Arnold. Vancouver, BC, Canada: ACM Press, pp. 153–160,

I distinguish four ways of deviating from honest signaling: the sender can lie or be timid and the receiver can be greedy or worried. I extend the Philip Sidney game to explicitly allow for punishment of such behavior and study its effect on the evolution of honest signaling:

- When punishment targets lying individuals, honest signaling emerges also for cost-free signals. So, punishment provides an alternative to the handicap principle.

- When punishment targets greedy individuals, honest signaling emerges also in cases with strong conflicts, similar to the punishment of defectors to promote cooperation.

- The evolution of honest signaling does not benefit from punishment of timid or worried individuals.

Chapter 6 provides some discussion and a summary.

1.3 Publication List

The following is my complete publication list at the time of writing, including work that does not concern signaling.

Articles in journals

1. David Catteeuw and Bernard Manderick (in press). "Honesty and deception in populations of selfish, adaptive individuals." In: *The Knowledge Engineering Review* 31.2.

2. David Catteeuw and Bernard Manderick (2014). "The Limits and Robustness of Reinforcement Learning in Lewis Signaling Games." In: *Connection Science* 26.2, pp. 161–177.

3. David Catteeuw and Bernard Manderick (2011b). "Heterogeneous Populations of Learning Agents in the Minority Game." In: *Lecture Notes in Computer Science, Adaptive and Learning Agents* 7113, pp. 100–113.

4. David Catteeuw and Bernard Manderick (2011c). "Learning in Minority Games with Multiple Resources." In: *Lecture Notes in Computer Science, Advances in Artificial Life* 5778. Ed. by George Kampis, István Karsai, and Eörs Szathmáry, pp. 326–333.

Articles at international, peer-reviewed conferences

1. David Catteeuw, The Anh Han, and Bernard Manderick (2014a). "Evolution of Honest Signaling by Social Punishment." In: *Proceedings of the 2014 Genetic and Evolutionary Computation Conference*. Ed. by Christian Igel and Dirk V. Arnold. Vancouver, BC, Canada: ACM Press, pp. 153–160.

2. David Catteeuw, Bernard Manderick, and The Anh Han (2013). "Evolutionary Stability of Honest Signaling in Finite Populations." In: *Proceedings of the IEEE Congress on Evolutionary Computation.* Ed. by Luis Gerardo de la Fraga and Carlos A. Coello Coello. Cancun, Mexico: IEEE Computer Society, pp. 2864–2870.

3. David Catteeuw and Bernard Manderick (2012b). "Honest Signaling: Learning Dynamics versus Evolutionary Stability." In: *Proceedings of the 21st Belgian-Dutch Conference on Machine Learning.* Ed. by Bernard De Baets, Bernard Manderick, Michael Rademaker, and Willem Waegeman. Ghent, Belgium, pp. 1–6.

4. David Catteeuw, Joachim De Beule, and Bernard Manderick (2011). "Roth-Erev Learning in Signaling and Language Games." In: *Proceedings of the 23rd Benelux Conference on Artificial Intelligence.* Ed. by Patrick De Causmaecker et al. Ghent, Belgium, pp. 65–74.

5. Yailen Martinez, Bert Van Vreckem, David Catteeuw, and Ann Nowé (2010). "Application of Learning Automata for Stochastic Online Scheduling." In: *Recent Advances in Optimization and its Applications in Engineering, Post-proceedings of the 14th Belgian-French-German Conference on Optimization.* Ed. by Moritz Diehl, Francois Glineur, Elias Jarlebring, and Wim Michiels. Springer-Verlag, pp. 491–498.

6. David Catteeuw and Bernard Manderick (2009). "Learning in the Time-Dependent Minority Game." In: *Proceedings of the 11th annual conference on Genetic and Evolutionary Computation.* Montréal, Canada: ACM Press, pp. 2011–2016.

Articles at international, peer-reviewed workshops

1. David Catteeuw and Bernard Manderick (2013). "The Limits of Reinforcement Learning in Lewis Signaling Games." In: *Proceedings of the 13th Adaptive and Learning Agents workshop.* Ed. by Sam Devlin, Daniel Hennes, and Enda Howly. Saint Paul, MN, USA, pp. 22–30.

2. David Catteeuw and Bernard Manderick (2012a). "Emergence of Honest Signaling through Reinforcement Learning." In: *Proceedings of the 12th Adaptive and Learning Agents workshop.* Ed. by Enda Howley, Peter Vrancx, and Matt Knudson. Valencia, Spain, pp. 81–86.

3. David Catteeuw and Bernard Manderick (2011a). "Heterogeneous Populations of Learning Agents in Minority Games." In: *Proceedings of the 11th Adaptive and Learning Agents workshop.* Ed. by Peter Vrancx, Matt Knudson, and Marek Grzes. Taipei, Taiwan, pp. 15–20.

Abstracts

1. David Catteeuw, Madalina M. Drugan, and Bernard Manderick (2014c). "'Guided' Restarts Hill-Climbing." In: *International Conference on Metaheuristics and Nature Inspired Computing.* Marrakech, Morocco, pp. 1–2.
2. David Catteeuw (2014). "The Emergence of Honest Signaling." In: *European Conference on Complex Systems.* Lucca, Italy, p. 1.
3. David Catteeuw, Madalina M. Drugan, and Bernard Manderick (2014b). "'Guided' Restarts Hill-Climbing." In: *In Search of Synergies between Reinforcement learning and Evolutionary Computation, Workshop at the 13th International Conference on Parallel Problem Solving from Nature.* Ed. by Madalina M. Drugan and Bernard Manderick. Ljubljana, Slovenia, pp. 1–4.
4. Yailen Martinez, Bert Van Vreckem, and David Catteeuw (2009). "Multi-Stage Scheduling Problem with Parallel Machines." In: *Book of Abstracts of the 14th Belgian-French-German Conference on Optimization.* Leuven, Belgium: Katholieke Universiteit Leuven, p. 162.

Chapter 2

Game Theory

This chapter reviews the necessary background on game theory (Section 2.2) and learning in games (Section 2.3). Section 2.2 is largely based on the book of Binmore (2007).

Contents

2.1 Introduction

Signaling is an interaction, in the simplest case, between two agents: a sender and a receiver. The former has some private information and sends a signal. The latter observes that signal and responds. The signal may, or may not, decrease the receiver's uncertainty about the sender's private information and this may help him to choose an appropriate response.

To study signaling, this thesis relies on game theory (Neumann and Morgenstern, 1944) which is a mathematical framework to model interactions between agents (which may be individuals, companies, countries, animals, . . .) as games and predict their outcome.

This chapter introduces only those concepts of game theory necessary to understand the rest of the text. Section 2.2 discusses basic concepts of games (Section 2.2.1 and 2.2.2), two solution concepts (Section 2.2.3), and situates signaling games in the broader classes of incomplete information games and games in extensive form (Section 2.2.4). The former class models interactions where one or more agents have private information. The latter class explicitly models sequential interactions, where the agents take actions one at a time. This thesis studies two well-known signaling games: the Lewis signaling game (Chapter 3) and the Philip Sidney game (Chapter 4). It also introduces two extensions to these games: the Lewis signaling game with noise (Section 3.5) and the Philip Sidney game with punishment (Chapter 5). These are not signaling games but belong to the same class of incomplete information games. All games in this thesis are two-player games: one player is the sender, the other is the receiver.

Game theory provides different solution concepts. These are models that predict a game's outcome. The classic approach (Section 2.2.3) is to assume that the agents are fully rational: they have unlimited computing power and always choose the option that has the best expected outcome. Alternate approaches do not assume agents are fully rational but only boundedly rational (Fudenberg and Levine, 1998a; Hart, 2005). They model evolution (Section 2.3.1) or individual learning (Section 2.3.2). In evolution, populations of agents evolve, possibly towards rational behavior, while the agents themselves are blindly executing their genetically determined strategy. In individual learning, agents improve their behavior based on previous experiences applying ‘trial-and-error.’ In this thesis, I apply both evolution and individual learning, and compare the results with the classic approach.

2.2 Game Theory

Game theory models interactions by *games*. The interacting agents are called the *players*. The *outcome* of an interaction depends on the *actions* of all players. Each player has clear preferences for each possible outcome, but these are not necessarily aligned. For example, in the Lewis signaling game (Chapter 3), the players' preferences are perfectly aligned: they have common interests. In the Philip Sidney game (Chapter 4) the players' preferences are not always aligned: they may have conflicting interests. Since your best action generally depends on what the other players do, deciding what the best action is, is rarely trivial. A player's preference for different outcomes is modeled by a *payoff function* which assigns a *payoff*, a real number, to each possible outcome. Higher payoffs are assigned to more preferred outcomes.

Game theory assumes that all players

- know the rules of the game: the number of players, the possible actions they have, which actions lead to which outcomes, and all players' payoffs for all outcomes; and
- are *rational*.

The latter means players *always* take the action that will yield the highest payoff, or, in the case of uncertainty, the action they *expect* to yield the highest payoff. Game theory further assumes that the game's rules and the fact that all players are rational are *common knowledge*: everybody knows it, everybody knows everybody knows it, everybody knows everybody knows everybody knows it, and so on.

These assumptions are sufficient, but not necessary (Gintis, 2000), to show that all players will, or should play a strategy that leads to an equilibrium outcome. See for example (Aumann and Brandenburger, 1995; Nash, 1950b). Such outcomes are called 'equilibria' because players cannot improve their payoff by deviating from it. Section 2.2.3 gives an example of an equilibrium: the Nash equilibrium.

There are some problems with equilibria (Fudenberg and Levine, 1998b; Gintis, 2000).

- The equilibrium selection problem (Harsanyi and Selten, 1988): which equilibrium should you select if there is more than one?
- Humans do not always play equilibrium behavior and are only boundedly rational (Binmore, 2007, sect. 2.9.2).
- Players do not always know each other's payoffs or preferences so what is rational according to the game may not be rational from the player's point of view.

An alternate explanation is that an equilibrium is the end result of dynamical processes—evolution or learning—that maximize payoff by changing behavior until a stable fixed point (an equilibrium) is reached. I discuss evolution in Section 2.3.1 and learning in Section 2.3.2.

2.2.1 Games in extensive form

Most interactions have a temporal aspect: actions are not taken simultaneously but one at a time. Such interactions are naturally modeled by games in extensive form.

A *game in extensive form* or extensive form game is represented by a tree, such as the one in Figure 2.1. At each non-terminal node of the tree, one of the players must take an action. Non-terminal nodes are therefore called '*decision nodes*.' Taking an action at a given node corresponds to choosing a branch and going to the corresponding child node. At each terminal node of the tree, the game ends and payoffs are defined for each player. An *outcome* of a game corresponds to a path from the initial node all the way down to a terminal node. It consists of an action at each of the decision nodes on that path. *Chance nodes* represent stochasticity in an interaction, such as rolling dice. They are implemented as decision nodes

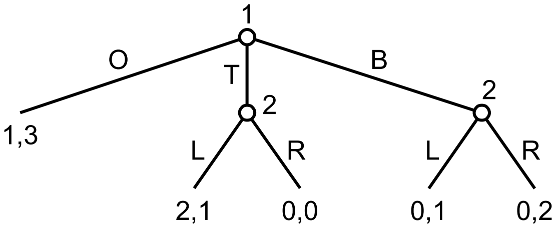


Figure 2.1: A game in extensive form is represented by a tree. Each non-terminal node (circles) is labeled with the name of the player (in this case '1' or '2') who decides which action (or branch) to take. Each terminal node represents a unique outcome of the game and defines a payoff for each player. The payoffs are listed in the order in which players appear in the tree (from top to bottom and from left to right).

of *Nature*—a special player who does not play strategically but according to a fixed strategy which is part of the game's definition.

In Figure 2.1, at the root node, player 1 must choose between actions `O`, `T`, and `B`. At the two other decision nodes, player 2 can take either action `L` or `R`. At each terminal node, payoffs are given. The first number is the payoff for player 1, the second for player 2. So, if player 1 takes action `T` and player 2 action `L`, the outcome of the game is $(\mathtt{T}, \mathtt{L})$ and the payoff is 2 for player 1 and 1 for player 2. I use u to represent the payoff function and subscripts to refer to the payoff of a specific player, for example, $u_1(\mathtt{T}, \mathtt{L}) = 2$ and $u_2(\mathtt{O}) = 3$ in Figure 2.1. For clarity, action names are printed in a `typewriter font`.

Perfect vs. imperfect information

If there is a player that cannot uniquely identify all of his decision nodes, the game is an *imperfect information game*. Two different nodes which cannot be uniquely distinguished from each other are in the same *information set*. In the game tree, such nodes are connected by a dashed line. Figure 2.2 shows an example of an imperfect information game. In that game, player 2 cannot distinguish between player 1's actions `T` and `B`.

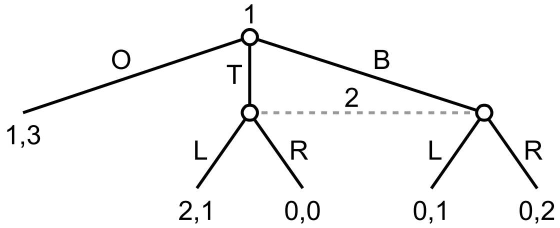


Figure 2.2: An imperfect information game. The same game as in Figure 2.1, but here player 2 does not know whether player 1 took action `T` or `B`. Both of player 2's decision nodes are in the same information set.

When all players know exactly what actions were previously taken in the game, the game is a *perfect information game* and the players can uniquely identify each of their decision nodes. In other words, all information sets are singletons.

Strategies

Three types of strategies are defined for games in extensive form: pure, behavioral, and mixed strategies.

- A *pure strategy* consists of an action at each of a player's decision nodes. To avoid confusion, the elements of a pure strategy follow the order of the decision nodes: from top to bottom and from left to right. In Figure 2.1 player 2's pure strategies (`LL`, `LR`, `RL`, and `RR`) have two actions: one for his decision following action `T` and one for his decision following action `B`.

- A *mixed strategy* is a probability distribution over all pure strategies of a player and is represented by a linear combination. For example, player 2's mixed strategy $^2/_3$`LL` + $^1/_3$`RL` means he uses pure strategy `LL` with probability $^2/_3$ and pure strategy `RL` with probability $^1/_3$.

- A *behavioral strategy* is a probability distribution over the actions in each decision node of a player. For example, player 2's behavioral strategy $(\sfrac{2}{3}\texttt{L} + \sfrac{1}{3}\texttt{R}, \texttt{L})$ means that after action `T` he plays action `L` with probability $\sfrac{2}{3}$ and action `R` with probability $\sfrac{1}{3}$ and after action `B` he plays always action `L`. Again, the elements of a behavioral strategy follow the order of the decision nodes: from top to bottom and from left to right.

Given a strategy s_i for each player $i = 1, 2, \ldots, M$, $s = (s_1, \ldots, s_M)$ is a *strategy profile* and player i's payoff $u_i(s)$ given this strategy profile is the payoff assigned to the outcome that results when each player i sticks to his strategy s_i. For mixed and behavioral strategies or games with chance nodes, more than one outcome is possible. In that case, player i's payoff $u_i(s)$ is the average of his payoff at every outcome weighted by the probability of each outcome.

This text only uses games with a finite number of players and a finite number of pure strategies for each player, also called '*finite games*.'

2.2.2 Games in strategic form

Interactions where players act simultaneously and only once are usually modeled by *games in strategic form* (also known as normal form). Such games are represented by an M-dimensional table, where M is the number of players. An example of a game in strategic form is shown in Figure 2.3. Along each dimension $i = 1, 2, \ldots, M$ are the possible actions of player i. The combined actions of all players lead to a unique outcome with payoffs for each player given in the corresponding entry of the table.

From extensive to strategic form

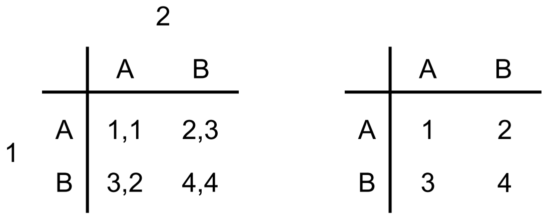
In Section 2.2.3 and also further on, I will describe some solution concepts. These predict how players behave in a game or, from another point of view, solution concepts advice players how to play. Some solution concepts (such as the Nash equilibrium, Section 2.2.3) are only, or more easily, defined for games in strategic form. Luckily, games in extensive form can also be represented in strategic form.

		2	
		C	D
1	A	1,1	2,2
	B	3,3	4,4

Figure 2.3: Two-player game in strategic form. Player 1 has actions `A` and `B`. Player 2 has actions `C` and `D`. If player 1 plays action `A` and player 2 plays action `C`, the outcome is $(\mathtt{A}, \mathtt{C})$ and both players get a payoff of 1.

		2			
		LL	LR	RL	RR
	O	1,3	1,3	1,3	1,3
1	T	2,1	2,1	0,0	0,0
	B	0,1	0,2	0,1	0,2

Figure 2.4: Strategic form game of the extensive form game with perfect information in Figure 2.1.



		2	
		A	B
1	A	1,1	2,3
	B	3,2	4,4

	A	B
A	1	2
B	3	4

Figure 2.5: Symmetric game in strategic form. *Left*: Standard representation. *Right*: Simplified representation, where the names of the player roles are left out and only the payoffs of the first player are given.

Figure 2.4 shows the strategic form of the game in Figure 2.1. Given a game in extensive form, its strategic form is defined as follows.

Definition 2.1. The *strategic form* of a game in extensive form is a game where:

- The players are the same as in the extensive form.
- The players' actions correspond to their pure strategies in the extensive form.
- The outcomes' payoffs are the expected payoffs when applying the corresponding pure strategies to the extensive form.

The resulting game in strategic form is not necessarily equivalent to the original game in extensive form (for example (Cooper and Van Huyck, 2003; Harsanyi and Selten, 1988; Schelling, 1960; Seidenfeld, 1994)). Converting a game from extensive form to strategic form may throw away information, namely, the sequence in which players take actions.

Symmetric games

A game is *symmetric* if all player roles are equivalent. Figure 2.5 shows a symmetric, two-player game. Its player roles are equivalent: both players have the same set of actions $\{\mathtt{A}, \mathtt{B}\}$ and the payoff for player 1 for outcome $(\texttt{A}, \texttt{B})$ is the same as the payoff for player 2 for outcome $(\texttt{B}, \texttt{A})$: $u_1(\texttt{A}, \texttt{B}) = u_2(\texttt{B}, \texttt{A}) = 2$. The payoff function of a symmetric, two-player game in strategic form can be given by a matrix with the payoffs of the first player. I use standard matrix notation in that case: $u_{\texttt{A},\texttt{B}} = 2$.

From asymmetric to symmetric

Some solution concepts (such as the evolutionarily stable strategy, Section 2.3.1) are defined for symmetric games. An asymmetric game must therefore be made symmetric before such a solution concept can be applied.

Assuming that each player can be in any role with equal probability, it is easy to define the symmetric version of any asymmetric game (Sigmund, 2010, sect. 2.5).

Definition 2.2. The *symmetrized game* or the symmetric version of an asymmetric game is a game where:

- The number of players is the same as in the asymmetric game.
- All players have the same action set: one action per element of the Cartesian product of the original action sets.
- The payoffs are computed from the original game, assuming that each player is equally likely to be in each of the player roles of the asymmetric game.

Figure 2.6 shows an example of an asymmetric game and its symmetric version with action set $\{\texttt{AC}, \texttt{AD}, \texttt{BC}, \texttt{BD}\}$ each of which correspond to an element of the Cartesian product of the original action sets $\{\texttt{A}, \texttt{B}\} \times \{\texttt{C}, \texttt{D}\} = \{(\texttt{A}, \texttt{C}), (\texttt{A}, \texttt{D}), (\texttt{B}, \texttt{C}), (\texttt{B}, \texttt{D})\}$. The payoff for strategy AC against strategy BD is

$$v_{\texttt{AC},\texttt{BD}} = \frac{1}{2}u_1(\texttt{A}, \texttt{D}) + \frac{1}{2}u_2(\texttt{B}, \texttt{C}) = \frac{1}{2}2 + \frac{1}{2}0 = 1,$$

where u is the payoff function of the asymmetric game and v the payoff function of the symmetric game.

		2	
		C	D
1	A	1,1	2,0
	B	3,0	0,4

	AC	AD	BC	BD
AC	1	1.5	0.5	1
AD	0.5	1	2.5	3
BC	2	0.5	1.5	0
BD	1.5	0	3.5	2

Figure 2.6: *Left*: Asymmetric game in strategic form. *Right*: The symmetric version, simplified as in Figure 2.5, where each player is assumed to be half of the time in role 1 of the asymmetric game and half of the time in role 2 of the asymmetric game.

2.2.3 Solution concepts

Given a game, one likes to predict what the players will or should do. Several solution concepts attempt this. Here, I discuss the Nash equilibrium. Later, in Section 2.3.1, I also discuss evolutionarily stable strategies.

In the following, s is a strategy profile—a tuple of strategies, one per player—where each player i plays strategy s_i and s_{-i} is a set of strategies for all players except player i. So, (s_{-i}, s'_i) is used to indicate that player i deviates from the strategy profile s: all players $j \neq i$ play strategy s_j but player i plays strategy s'_i.[1]

Definition 2.3. A *Nash equilibrium* of a game in strategic form is a strategy profile s in pure or mixed strategies such that every player's strategy is a best response to the other players' strategies: $u_i(s) \geq u_i(s_{-i}, s'_i)$ for all players i and strategies s'_i (Nash, 1950a). A *strict* Nash equilibrium is a Nash equilibrium where every player's strategy is a unique best response: $u_i(s) > u_i(s_{-i}, s'_i)$ for all players i and strategies s'_i. A Nash equilibrium that is not strict is sometimes called 'weak.'

A player may still be able to improve his payoff if more than one player

[1] This notation is standard in game theory. It is used for example in (Gintis, 2000).

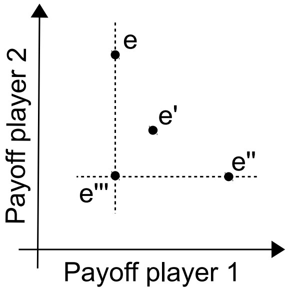


Figure 2.7: Three Pareto optimal equilibria e, e', and e'' which Pareto dominate equilibrium e'''.

deviates at the same time. Any finite game, which has a finite number of players with a finite number of pure strategies, has at least one Nash equilibrium (Nash, 1950a). Since the Nash equilibrium is defined in pure or mixed strategies, the Nash equilibria of a game in extensive form are the same as those of its corresponding strategic form.

Many games have more than one equilibrium. It is not clear, in general, which one is preferred, but there is an obvious partial order: *Pareto dominance*

Definition 2.4. Equilibrium e *Pareto dominates* equilibrium e' if none of the players prefer e' to e but some prefer e to e': $u_i(e) \geq u_i(e')$ for all players i and $u_i(e) > u_i(e')$ for at least one player i. An equilibrium which is not Pareto dominated is *Pareto optimal*.[2]

For a two-player game, Figure 2.7 shows three Pareto optimal equilibria e, e', and e'' which Pareto dominate equilibrium e'''.

2.2.4 Private information

In some games, some player may know something right from the start of the game that the others do not. Those are called games with private informa-

[2] Pareto optimal is sometimes called ‘Pareto efficient’ and Pareto dominant is sometimes called ‘payoff dominant.’

tion, also known as *incomplete information games*. The private information of a player is called his '*type*' and the probability distribution over the different types is common knowledge. Just as in imperfect information games, incomplete information games have at least one information set that contains more than one decision node. Whereas in an imperfect information game, a player is unable to distinguish different decision nodes, because *he cannot observe all actions*; in an incomplete information game, a player is unable to distinguish different decision nodes, because *he lacks some information right from the start.* Figure 2.8 shows a classification of some games and game classes according to these information criteria.

Representing incomplete information games

Interactions where one player has private information can be modeled by letting *Nature* pick the type of that player before any player takes any action (Harsanyi, 1967). Figure 2.9 shows an example where *Sender* has private information. For each possible type, the root node of the game tree has a branch and each of these branches is followed by a copy of the same subtree $\mathcal{T}$ representing the actual decision nodes and actions of the players. Players that cannot distinguish between *Nature*'s moves will have information sets connecting all decision nodes occurring on the same position in the different copies of the subtree $\mathcal{T}$. The payoffs for each subtree $\mathcal{T}$ may be different. This technique can be generalized to games where more than one player has private information (Harsanyi, 1967).

Adverse selection

The introduction of incomplete information games was a breakthrough for economics, because such games model a huge set of real-world situations that could not be captured before: insurance, advertising, and bargaining are examples where one player has private information. A person seeking health insurance is better informed about his condition than the insurance company. A young adult whose family members all have the same genetic disorder that can only be cured with costly surgery better gets a full-option

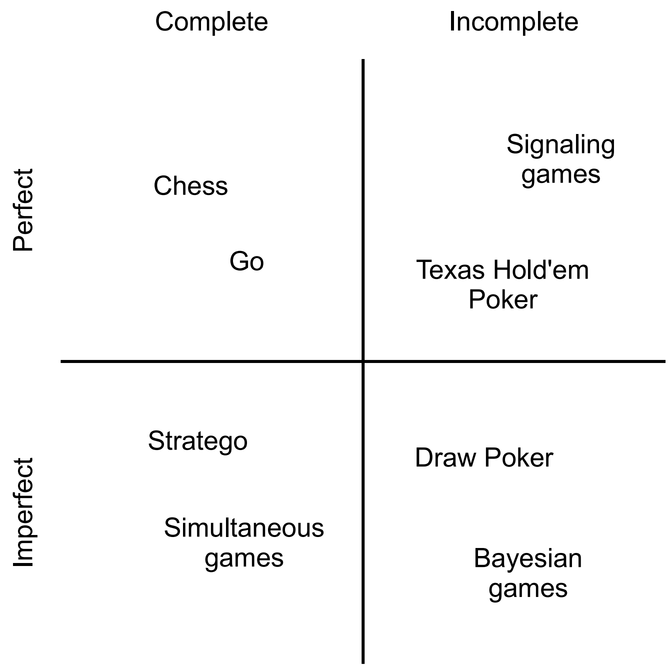


Figure 2.8: Classification of games according to information. In imperfect information games, some players cannot observe all actions. For example, in Stratego, the players do not see how their opponents initially arranges his pieces. In incomplete information games, some players know something that others do not. For example, in Texas Hold’em Poker, players receive two cards which their opponents cannot see.

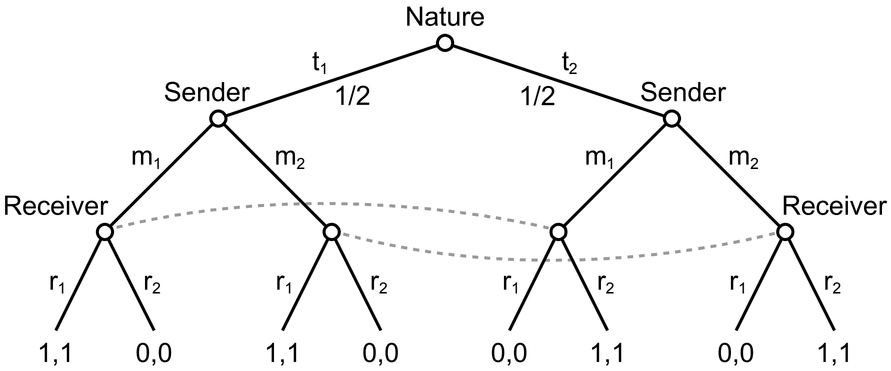


Figure 2.9: A Lewis signaling game—an incomplete information game where *Nature* decides *Sender*'s type (`t`$_1$ or `t`$_2$ with equal probability) before any player takes an action. *Receiver* cannot observe this type, so the corresponding nodes of the two subtrees (following `t`$_1$ and `t`$_2$) are part of the same information set (dashed lines).

health insurance, while someone whose grandparents all died of old age should go for a cheap insurance.

The key problem of private information is *adverse selection*: the quality of the supply decreases with decreasing price until the entire market collapses. Akerlof (1970) introduced this concept with an example of the market of second hand cars (Example 1.4). Higher quality cars deserve a higher price than lower quality cars; but since the buyer does not know the quality of the second hand car, a dishonest seller can sell a bad second hand car for the price of a good one. This motivates buyers to offer a lower price and owners of good second hand cars will step out of the market because their vehicle is worth more than the market price. When good quality cars leave the market, the average quality of a second hand car drops, which lowers the average price offer again, and so on.

Adverse selection applies to consumer goods of unknown or hard to determine quality just as to second hand cars. It also applies to insurance and health insurance in particular (Harford, 2006). If the insurance premium increases, the most healthy people will no longer take an insurance since they expect not to benefit from it and the average cost per client will rise which in turn may force the insurance company to further increase the premium. To avoid market failures, health insurance is compulsory in many countries (Harford, 2006).

In some situations, adverse selection can be avoided by means of signals. For example, an employer cannot directly observe a potential employee's productivity, but the latter can signal his abilities with his education (Spence, 1973). This was explained in Example 1.5. Another typical example is car insurance. Car insurance companies provide a low cost, partial insurance and a high cost, full insurance such that the customer can signal his risk aversion with his choice (Wilson, 1977).

In the job market example the informed player (the student and potential employee) moves first and signals by acquiring education while the uninformed player (the employer) responds to the signal by hiring the potential employee or not. Game theory models such an interaction as a *signaling game*. I discuss it in the next section.

In the car insurance market the uninformed player (the insurance company) moves first by setting insurance options and prices while the informed player (the car owner) responds and signals by choosing an insurance option. Game theory models such an interaction as a *screening game*. This thesis studies only signaling games. Riley (2001) reviews both signaling and screening games.

Signaling games

Signaling games (see for example Figure 2.9) are two-player, incomplete information games, where the first player, called *Sender*, has some private information. *Nature* selects *Sender*'s type t (his private information) from set $\mathcal{T}$ according to probability distribution π. I denote the probability of *Sender* having type t by π_t. *Sender* observes his type t, selects a signal m from set $\mathcal{M}$, and sends it to the second player, called *Receiver*, who observes the signal m and selects a response r from set $\mathcal{R}$. The set of types $\mathcal{T}$, signals $\mathcal{M}$, and responses $\mathcal{R}$ are finite sets or real intervals. The set of signals $\mathcal{M}$ may depend on type t and the set of possible responses $\mathcal{R}$ may depend on signal m. The payoff function $u : \mathcal{T} \times \mathcal{M} \times \mathcal{R} \to \mathbb{R}^2$ determines a payoff for both players for each possible outcome of the game. In this thesis, I only consider signaling games where the available signals are independent of the current type; the available responses are independent of the current signal; and where the available types, signals, and responses are finite sets. So, I define a signaling game as follows.

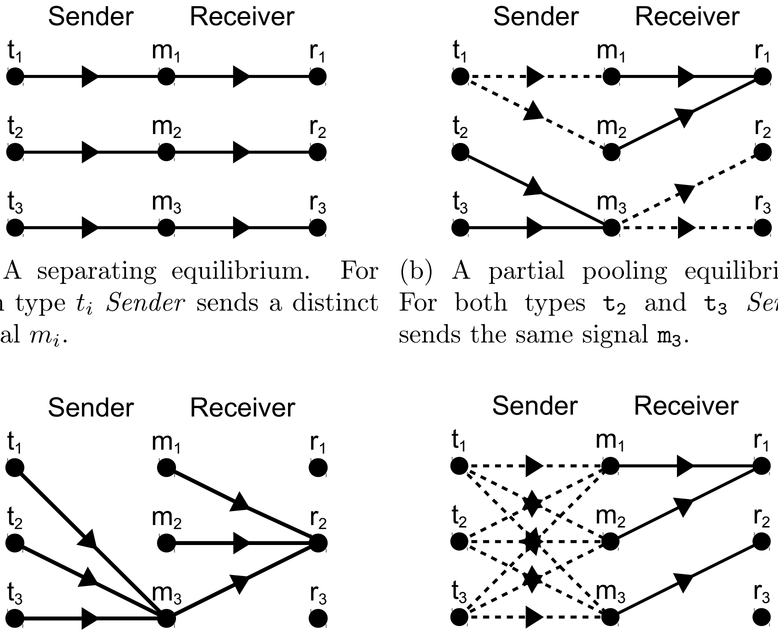
Definition 2.5. A *signaling game* is a two-player, incomplete information game. The first player, called *Sender*, has a finite set of types $\mathcal{T}$ with distribution π. Each type t occurs with non-zero probability ($\pi_t > 0$). For each type, *Sender*'s actions are the finite set of signals $\mathcal{M}$. The second player, called *Receiver*, has no private information. For each signal, his actions are the finite set of responses $\mathcal{R}$. A signaling game has a payoff function $u : \mathcal{T} \times \mathcal{M} \times \mathcal{R} \to \mathbb{R}^2$ and is fully described by the 5-tuple $(\mathcal{T}, \mathcal{M}, \mathcal{R}, \pi, u)$.

Figure 2.9 shows an example of a signaling game (more particularly a Lewis signaling game, see Chapter 3). In this game, *Sender* has two types which occur with equal probability: $\mathcal{T} = \{\mathtt{t}_1, \mathtt{t}_2\}$ and $\pi = (1/2, 1/2)$. In both cases he can choose between two signals: $\mathcal{M} = \{\mathtt{m}_1, \mathtt{m}_2\}$. *Receiver* observes the signal, but not the type, hence the dashed lines connecting the two decision nodes following signal $\mathtt{m}_1$ and those following signal $\mathtt{m}_2$. He has two options to respond: $\mathcal{R} = \{\mathtt{r}_1, \mathtt{r}_2\}$.

This game is successful if *Receiver*'s response corresponds to *Sender*'s type as follows:

$$u_i(t_j, m_k, r_l) = \begin{cases} 1 & \text{if } j = l, \\ 0 & \text{otherwise,} \end{cases} \qquad \text{for all players } i = \textit{Sender}, \textit{Receiver}.$$

Since both players always receive the same payoff, the game is fully cooperative and both players benefit from signaling. We expect *Sender* to send signals that allow *Receiver* to infer his type and *Receiver* to respond correspondingly. Game theory tries to predict the outcome by determining the equilibria.



(a) A separating equilibrium. For each type t_i *Sender* sends a distinct signal m_i.

(b) A partial pooling equilibrium. For both types $\mathtt{t}_2$ and $\mathtt{t}_3$ *Sender* sends the same signal $\mathtt{m}_3$.

(c) A pooling equilibrium where *Sender* uses $\mathtt{m}_3$ for all types.

(d) A babbling equilibrium. *Sender* uses all signals with equal probability for all types.

Figure 2.10: Four equilibria in behavioral strategies for a signaling game with 3 types, 3 signals, and 3 responses. *Sender*'s strategies map types t_i to signals m_j and *Receiver*'s strategies map signals m_j to responses r_k. A solid line represents a probability of 1. A dashed line represents a probability between 0 and 1.

Equilibria in signaling games

Signaling games have many Nash equilibria with different degrees of signaling: separating, partial pooling, and pooling equilibria (Sobel, 2009). Some also have babbling equilibria (Sobel, 2009). In a *separating equilibrium* (Figure 2.10a) *Sender* uses a different signal for all types allowing *Receiver* to infer with certainty *Sender*'s type. In a *pooling equilibrium* (Figure 2.10c) *Sender* uses the same signal for all types and *Receiver* can infer nothing. In between the two extremes are *partial pooling equilibria* (Figure 2.10b) where *Sender* uses the same signal for some, but not all, types. Some signaling games, such as the one in Figure 2.9, have cost-free signals. These are signals that do not influence the payoff function—though they may still influence *Receiver*'s behavior. Such signaling games also have a *babbling equilibrium* (Figure 2.10d) where *Sender* uses all signals with equal probability for all types (Sobel, 2009). Just as in pooling equilibria, in babbling equilibria signals convey no information.

Even the small signaling game of Figure 2.9 has many equilibria.[3] It has two separating equilibria which are strict Nash equilibria: $((\mathtt{s}_1, \mathtt{s}_2), (\mathtt{r}_1, \mathtt{r}_2))$ and $((\mathtt{s}_2, \mathtt{s}_1), (\mathtt{r}_2, \mathtt{r}_1))$. The equilibria here are written as tuples of behavioral strategies. So, the strategy $(\mathtt{s}_1, \mathtt{s}_2)$ means that *Sender* always uses signal $\mathtt{s}_1$ at his first decision node, which follows type $\mathtt{t}_1$, and always uses signal $\mathtt{s}_2$ at his second decision node, which follows type $\mathtt{t}_2$. The game in Figure 2.9 also has a connected set of weak Nash equilibria which includes pooling and babbling equilibria. The pooling equilibria are those where *Sender* always sends $\mathtt{s}_1$ or always sends $\mathtt{s}_2$. *Receiver*'s best responses to these *Sender* strategies $((\mathtt{s}_1, \mathtt{s}_1)$, $(\mathtt{s}_2, \mathtt{s}_2)$, and $({}^1\!/\!{}_2\mathtt{s}_1 + {}^1\!/\!{}_2\mathtt{s}_2, {}^1\!/\!{}_2\mathtt{s}_1 + {}^1\!/\!{}_2\mathtt{s}_2))$ are the strategies $(x\mathtt{r}_1 + (1-x)\mathtt{r}_2, x\mathtt{r}_1 + (1-x)\mathtt{r}_2)$ where $x \in [0, 1]$.

When many Nash equilibria exist it is hard for players to coordinate on the same equilibrium (Harsanyi and Selten, 1988). Economists tried to refine the set of Nash equilibria by means of extra rationality. Unfortunately, these refinements do not restrict the possibilities to a single equi-

[3] I use Gambit's implementation (McKelvey et al., 2014) of Mangasarian's algorithm (Mangasarian, 1964) to compute the equilibria of specific signaling games. The algorithm is guaranteed to find all Nash equilibria of any finite two player game (Shapley, 1974).

librium or are not always applicable to signaling games (Riley, 2001). This made signaling an important testing ground for equilibrium refinements (Appendix B).

The basic assumption of rationality and common knowledge is debatable (Fudenberg and Levine, 1998b; Gintis, 2000), even more so the extra rationality needed for these equilibrium refinements. Therefore, this thesis takes a different approach and looks at which equilibria are the likely outcome of different dynamical processes modeling evolution or learning. I contrast the results with those of static equilibrium analyses.

Applications of signaling

That signaling games model a wide variety of applications was clear from the beginning. Philosopher Lewis (1969, ch. 4) introduced signaling games to study the emergence of conventions, such as human language, in cooperative settings. Economist Spence (1973) and biologist Zahavi (1975) both independently suggested what is now known as the handicap principle—signals must be costly in order to be honest if there is a conflict of interest—and applied it to different domains. Spence (1973) suggested that university degrees act as honest signals in the job market because they are costly to acquire (Example 1.5). Zahavi (1975) suggested that the peacock's tail is a costly and hence honest signal of its quality as a mating partner (Example 1.3). Riley (2001) and Sobel (2009) discuss several applications of signaling games in great detail.

In the following chapters, I will use signaling games to study the emergence of honest signaling in cooperative settings by means of the Lewis signaling game (Chapter 3) and in competitive settings by means of the Philip Sidney game and an extension thereof (Chapters 4 and 5). I introduce these games in the relevant chapters.

2.3 Learning in Games

The classic interpretation of equilibria is that they are the outcome selected by rational agents, assuming they know the rules of the game, know that the other players are also rational, and that all this is common knowledge (Section 2.2). An alternate interpretation is that equilibria are the outcome of dynamical processes such as evolution, imitation and experimentation, or individual learning (Fudenberg and Levine, 1998b).

On the one hand, such processes have not been able to explain all equilibrium concepts in literature. For example, it is currently known that there cannot exist an 'uncoupled' dynamical process that leads to a Nash equilibrium in all games, where 'uncoupled' means that agents only see their own payoff, not those of their opponents (Hart and Mansour, 2010; Hart and Mas-Colell, 2003, 2006). On the other hand, they form a solution concept on their own, that can even help to refine equilibrium concepts (Fudenberg and Levine, 1998b).

This thesis considers models based on evolution, social learning, and individual learning.

2.3.1 Evolution and social learning

Evolutionary game theory (Maynard Smith, 1982) studies how genetically encoded strategies spread through a population of agents that repeatedly and strategically interact so that a strategy's success depends on the frequencies of all strategies in the population—*frequency dependent selection*. Offspring inherit the strategy of their parents and natural selection ensures that the most successful agents (are more likely to) reproduce the fastest. As a result, successful strategies take over the population while unsuccessful strategies go extinct.

The same models that hold for strategies spreading through evolution hold for strategies spreading through social learning, where agents prefer to imitate the strategy of better performing agents (Fudenberg and Levine, 1998b; Hofbauer and Sigmund, 1998; Imhof et al., 2005; Sigmund, 2010; Traulsen and Hauert, 2009).

Contrary to classic game theory, evolutionary game theory provides a solution concept without relying on rationality. In evolution, the agents execute the strategy they inherited from their parents, they do not even need to make conscious decisions; in social learning, the agents simply imitate the strategies of better performing agents.

This section only discusses the concepts and techniques used in later chapters: the evolutionarily stable strategy, an evolutionary dynamics in finite populations, and the evolutionarily stable strategy in finite populations. All three assume *well-mixed* populations: every agent is equally likely to interact (play a game) with all other agents in the population. In this thesis, the interactions are modeled by one of the games mentioned before (the Lewis signaling game, the Philip Sidney game, and their extensions) so they are pairwise. Each agent plays a pure strategy of the game and is equally likely to take the role of *Sender* as the role of *Receiver*.

Evolutionarily stable strategies

The key solution concept of evolutionary game theory is the *evolutionarily stable strategy* or ESS (Maynard Smith and Price, 1973). Assume a well-mixed and infinitely large population where all agents adopt the (genetically encoded) strategy W—the *wild type*.[4] Sooner or later, a *mutant* appears in the population and adopts strategy M. Assuming pairwise and symmetric interactions, we have a symmetric, two-player game with payoff matrix u and can define the notion of evolutionary stability:

Definition 2.6. Strategy W is *evolutionarily stable against* strategy M if the expected payoff $f_{\mathtt{W}}$ of the wild type W is greater than the expected payoff $f_{\mathtt{M}}$ of the mutant strategy M:

$$(1-\epsilon)\, u_{\mathtt{W},\mathtt{W}} + \epsilon\, u_{\mathtt{W},\mathtt{M}} > (1-\epsilon)\, u_{\mathtt{M},\mathtt{W}} + \epsilon\, u_{\mathtt{M},\mathtt{M}}, \tag{2.1}$$

where ϵ is the fraction of mutants in the population and $u_{\mathtt{X},\mathtt{Y}}$ is the payoff of an X-strategist interacting with a Y-strategist. An *evolutionarily stable strategy* (ESS) is a strategy that is evolutionarily stable against all other strategies of the game.

[4] The *initially* most frequent strategy in the population is considered the wild type, the one without mutation. 'Initially,' since a mutant can appear that takes over the entire population.

Since the fraction of mutants is infinitely small, strategy W is evolutionarily stable against M if either the wild type strategy W is a best response to itself ($u_{\mathtt{W},\mathtt{W}} > u_{\mathtt{M},\mathtt{W}}$) or both W and M are best responses to W ($u_{\mathtt{W},\mathtt{W}} = u_{\mathtt{M},\mathtt{W}}$) but the wild type strategy W is a better response to the mutant strategy M, than the mutant strategy itself ($u_{\mathtt{W},\mathtt{M}} > u_{\mathtt{M},\mathtt{M}}$).

This leads to the following relation between the evolutionarily stable strategy and the Nash equilibrium. All strict Nash equilibria are evolutionarily stable strategies and all evolutionarily stable strategies are Nash equilibria: strict Nash $\subseteq$ ESS $\subseteq$ Nash (Hofbauer and Sigmund, 1998). For asymmetric games, the set of evolutionarily stable strategies and the set of strict Nash equilibria coincide because the ESSs of an asymmetric game are defined as those of its symmetric version and the set of ESSs of that symmetric game coincides with the set of strict Nash equilibria of the original asymmetric game (Selten, 1980).

Evolutionary stability can be quickly verified, but, just as the Nash equilibrium, it is defined for games in strategic form and it is a static solution concept. The latter means it is only concerned with the stability of an equilibrium, not with the paths that may, or may not, lead to it. In other words, whether or not an equilibrium is likely to emerge remains unclear.

Evolutionary dynamics in finite populations

I now assume, contrary to the previous section, a *finite* population, and consider the evolutionary dynamics as introduced by Taylor et al. (2004), Nowak et al. (2004), Imhof et al. (2005), and Traulsen et al. (2006). It has several benefits:

- Just like real systems it considers finite populations and is stochastic.

- It is more expressive than models of infinite populations. It allows to vary the population size and see its effects. For large populations, the results converge to those of infinite population models (Traulsen and Hauert, 2009). For small populations, results may be surprisingly different. For example, one strategy may be preferred over another while it is the other way around in large populations (Taylor et al., 2004).

Figure 2.11 shows an overview of the model's algorithm. In the next section, I will redefine evolutionary stability for finite populations according to Nowak et al. (2004).

This time, I explain the model from a social learning perspective. It is generally accepted that social learning and evolution lead to the same mathematical models (as mentioned at the beginning of this section), so I substitute one for the other without worrying whether the application concerns evolution or social learning.

The standard way to model social learning is with a *pairwise comparison process* (Traulsen et al., 2006). It goes as follows. There's a population of N agents and at each time step two agents A and B are chosen at random. The first observes the second agent's expected payoff f_B and his strategy. Given, how much better B performs than A, A must decide whether or not to adopt B's strategy. The probability that agent A adopts the strategy of another agent B is given by the Fermi function (Figure 2.12):

$$\Pr(A \to B) = \frac{1}{1 + e^{-\beta(f_B - f_A)}}, \tag{2.2}$$

where f_A and f_B are the expected payoffs of agents A and B respectively and β is the *selection pressure* or *imitation strength*.

The Fermi function has two benefits. First, it allows to tune selection between fully stochastic (selection pressure $\beta = 0$) and fully deterministic ($\beta \to \infty$). For $\beta = 0$, A imitates B with probability $1/2$ independent of the expected payoffs f_A and f_B. This is the limit of neutral or *random drift*. For large β, A imitates B if and only if B performs better ($f_B > f_A$). The choice is deterministic. Second, the function allows analytical results for any β (Traulsen et al., 2006). This is not the case for Nowak's model (Nowak and Sigmund, 2004) though both models are exactly the same for low selection pressure ($\beta \ll 1$) (Traulsen et al., 2006).

The pairwise comparison process always ends up in a *monomorphic state*—where all agents use the same strategy—unless there is a small exploration (or mutation) probability μ. If it is very small, as is usually assumed,

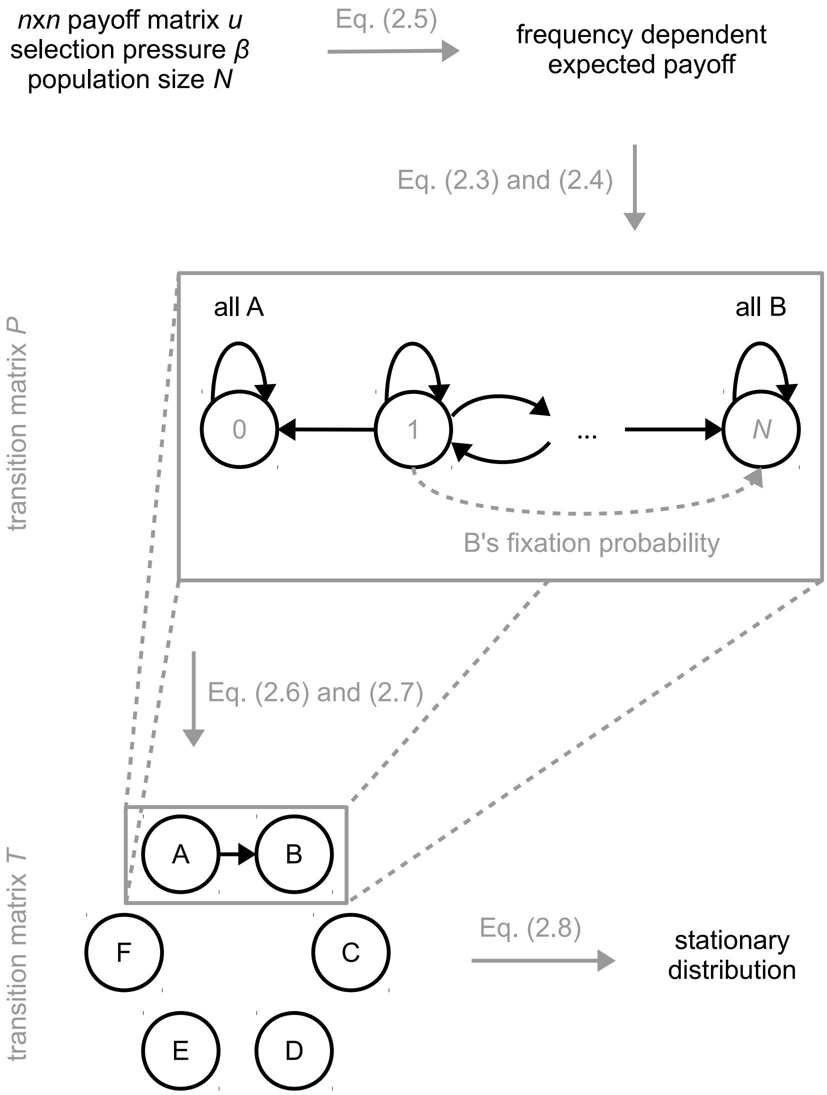


Figure 2.11: Graphical overview of the algorithm that calculates the fixation probabilities and stationary distribution for a symmetric, two-player game with n strategies (`A`, `B`, ..., `F`), given population size N, and selection pressure β. See text and equations for more information.

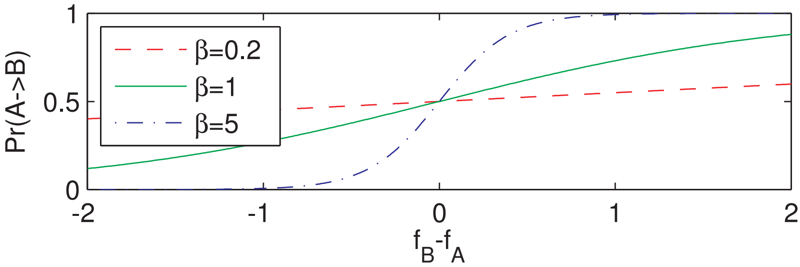


Figure 2.12: The Fermi function (Equation (2.2)) for different values of selection pressure β.

then the population consists of at most two strategies at any time: the wild type and a mutant (Wu et al., 2012). The latter either goes extinct or takes over the entire population before another mutant emerges. This process is visualized by the Markov chain[5] in the center of Figure 2.11. The probability that one mutant takes over the entire population is called the '*fixation probability.*'

The resulting population dynamics can be represented by another Markov chain: one where each state represents a monomorphic population (one for each possible strategy) and each transition probability, say from state `A` to `B`, is the fixation probability of a single mutant `B` in a population of all `A`s (Fudenberg and Imhof, 2006). The stationary distribution of this Markov chain gives the average time the population spends in each of the monomorphic states. This is visualized by the bottom part of Figure 2.11.

To calculate a mutant's fixation probability, consider again the Markov chain at the center of Figure 2.11. Each state is uniquely determined by the number of mutants `B` in the population of size N: $i = 0, \dots, N$. The number of wild types `A` is always $N - i$. The two extreme states $i = 0$ and $i = N$ are monomorphic and absorbing ($P_{0,0} = P_{N,N} = 1$). In all other states $i = 1, \dots, N-1$, the process

- transitions to state $i-1$ with probability $P_{i,i-1}$ (a mutant switches to

[5] Appendix A provides the necessary background on Markov chains.

the wild type),

- transitions to state $i+1$ with probability $P_{i,i+1}$ (a wild type switches to the mutant strategy), or
- remains in the same state with probability $P_{i,i} = 1 - P_{i,i-1} - P_{i,i+1}$.

The probability that the number of mutants decreases is the product of the probabilities that a mutant is selected (to observe), a wild type is selected (to be observed), and the mutant decides to imitate the wild type:

$$P_{i,i-1} = \frac{i}{N}\frac{N-i}{N}\frac{1}{1+e^{-\beta(f_{\mathtt{A}}(i)-f_{\mathtt{B}}(i))}}. \tag{2.3}$$

Similarly, the probability that the number of mutants increases is the product of the probabilities that a wild type is selected (to observe), a mutant is selected (to be observed), and the wild type decides to imitate the mutant:

$$P_{i,i+1} = \frac{N-i}{N}\frac{i}{N}\frac{1}{1+e^{-\beta(f_{\mathtt{B}}(i)-f_{\mathtt{A}}(i))}}. \tag{2.4}$$

A strategy's expected payoff may depend on the number of agents using each strategy: it is *frequency dependent.* For N agents of which i are mutants B the expected payoff of each strategy is (assuming random matching but excluding self-play):

$$\begin{aligned} f_{\mathtt{B}}(i) &= \frac{i-1}{N-1}u_{\mathtt{B},\mathtt{B}} + \frac{N-i}{N-1}u_{\mathtt{B},\mathtt{A}}, \\ f_{\mathtt{A}}(i) &= \frac{i}{N-1}u_{\mathtt{A},\mathtt{B}} + \frac{N-i-1}{N-1}u_{\mathtt{A},\mathtt{A}}, \end{aligned} \tag{2.5}$$

where $u_{\mathtt{X},\mathtt{Y}}$ stands for the payoff an X-strategist obtains in interaction with a Y-strategist.

Given the row stochastic and tri-diagonal[6] matrix P B's fixation probability, which is the probability that a single mutant B takes over a population

[6] A row stochastic matrix is a matrix where all elements are non-negative and each row's elements sum to one. A tri-diagonal matrix is a matrix where all elements that are not on the diagonal or directly below or above it are zero.

of As, is given by (Appendix A.4):

$$\rho_{\mathtt{A}\to\mathtt{B}} = \frac{1}{1 + \sum_{k=1}^{N-1} \prod_{i=1}^{k} \frac{P_{i,i-1}}{P_{i,i+1}}}. \tag{2.6}$$

In the limit of random drift (selection pressure $\beta = 0$), the fixation probability is the inverse of the population size: $1/N$ (Appendix A.4). A higher fixation probability means imitation (or natural selection) prefers the mutant strategy B. A lower fixation probability means the wild type A is preferred.

These fixation probabilities determine an $n \times n$ transition matrix T, where n is the number of strategies and $T_{i,j}$ is the probability that the population ends up in state i at time $t+1$ if it is in state j at time t. The off-diagonal elements of the matrix are the normalized fixation probabilities and the elements on the diagonal make sure the sum of each column adds up to 1:

$$T_{i,j} = \begin{cases} \rho_{j\to i}/(n-1) & \text{if } i \neq j, \\ 1 - \sum_k T_{k,j} & \text{if } i = j \text{ and for all } k \neq i. \end{cases} \tag{2.7}$$

The column stochastic matrix[7] T is the transition matrix of the Markov chain whose states represent the n monomorphic populations.

The stationary distribution π is the normalized (right-hand) eigenvector for eigenvalue 1 of T and describes the relative time the population spends in each of the monomorphic states (Appendix A.2):

$$\pi = T\pi. \tag{2.8}$$

Evolutionary stability in finite populations

In finite populations stability takes into account the population size N and is based on the dynamics described above.

Definition 2.7. A population adopting an *evolutionarily stable strategy in finite populations* ($\mathrm{ESS_N}$) resists invasion and replacement by any mutant strategy (Nowak et al., 2004).

[7] A column stochastic matrix is a matrix where all elements are non-negative and each column's elements sum to one.

Strategy W resists *invasion* if, in a population where all but one agent uses strategy W, the single mutant's expected payoff is lower than that of the other agents:

$$f_{\texttt{M}}(1) < f_{\texttt{W}}(1) \quad \text{for all } \texttt{M} \neq \texttt{W}.$$

Combined with Equation (2.5) this yields

$$(N-1)u_{\texttt{M},\texttt{W}} \;<\; u_{\texttt{W},\texttt{M}} + (N-2)u_{\texttt{W},\texttt{W}} \quad \text{for all } \texttt{M} \neq \texttt{W}. \tag{2.9}$$

Strategy W resists *replacement* if the probability that any mutant M fixates in a population of Ws is smaller than the neutral fixation probability:

$$\rho_{\texttt{W}\to\texttt{M}} < 1/N \quad \text{for all } \texttt{M} \neq \texttt{W}.$$

If the selection pressure is low ($\beta \to 0$), then W resists replacement if

$$u_{\texttt{M},\texttt{M}}(N-2) + u_{\texttt{M},\texttt{W}}(2N-1) < u_{\texttt{W},\texttt{M}}(N+1) + u_{\texttt{W},\texttt{W}}(2N-4) \quad \text{for all } \texttt{M} \neq \texttt{W}. \tag{2.10}$$

Low selection pressure is biologically relevant since most evolutionary changes are almost neutral (Ohta, 2002).

The relation between the traditional ESS concept (Definition 2.6) and the one in finite populations (ESS_N) is more clear when considering small and large populations separately. Strategy W is evolutionarily stable against M if $u_{\texttt{W},\texttt{W}} > u_{\texttt{M},\texttt{W}}$, or if $u_{\texttt{W},\texttt{W}} = u_{\texttt{M},\texttt{W}}$ and $u_{\texttt{W},\texttt{M}} > u_{\texttt{M},\texttt{M}}$. For small populations ($N = 2$), strategy W is ESS_N against M if $u_{\texttt{W},\texttt{M}} > u_{\texttt{M},\texttt{W}}$. The traditional ESS condition is thus neither necessary nor sufficient for ESS_N. For large populations, strategy W is ESS_N against M if $u_{\texttt{W},\texttt{W}} > u_{\texttt{M},\texttt{W}}$ and $2u_{\texttt{W},\texttt{W}} + u_{\texttt{W},\texttt{M}} > 2u_{\texttt{M},\texttt{W}} + u_{\texttt{M},\texttt{M}}$, so ESS is necessary but not sufficient.

2.3.2 Individual learning

The process discussed in the previous section describes the dynamics at the population level. In some settings, this is justified, for example, the effects of microscopic (on the agent level) behavioral rules may average out in large populations. If they do not, we must simulate the microscopic behavioral rules. One class of such rules are the individual learning rules or adaptive heuristics I describe here. When an agent repeatedly ends up and takes actions in the same situation, he can estimate the success of each action and update these estimates. Over time, an agent who learns will increasingly choose the most successful actions.

In this text, I consider two scenarios where agents can learn by repeatedly playing the same game: the fixed player model and the random matching model (Fudenberg and Levine, 1998b). The game that is repeated is called the '*stage game*.' In this thesis, the stage is the Lewis signaling game with or without noise (Chapter 3), or the Philip Sidney game (Section 4.3).

In the *fixed player model*, the same agents repeatedly play the stage game in the same roles. This is probably the simplest model in which agents can learn. Unfortunately the model also complicates matters. It allows agents to 'teach' their opponents, for example, by sticking to one pure strategy and being patient, letting opponents learn to best respond to this strategy.

To avoid such incentives, it suffices to let agents play against many different opponents. This is the case in the *random matching model*. At each iteration, M distinct players are randomly chosen from a well-mixed population of N agents to play the M-player stage game ($M = 2$ in this thesis). There are many more players than required to play the stage game ($N \gg M$). When players can identify each other, the random matching model leads to the same results as the fixed player model since each agent adapts to each opponent separately. The random matching model assumes the players cannot identify each other, so that each agent adapts to the average behavior in the population. Another difference: in the fixed player model, each agent always plays the same role, while in the random matching model, an agent can be selected to play any of the roles of the stage game.

I discuss three adaptive heuristics:

- *Roth-Erev learning* (Roth and Erev, 1995),
- *Q-learning* (Watkins, 1989), and
- *learning automata* (Narendra and Thathachar, 1974).

Roth-Erev learning and Q-learning are so-called action value methods. Such methods consist of an update rule, an action selection rule, and an *action* *value* $q_{s,a}$ for each state-action pair (s, a) which indicates the quality of taking action a in state s relative to the other actions in that state. The *update rule* determines how action values are updated based on new experience. The *action selection rule* determines which action to select, given the current state and the action values, by calculating the probability $p_{s,a}$ of taking action a in the current state s for all actions a. The usual constraints on probabilities hold: $p_{s,a} \geq 0$ for all states s and actions a, and $\sum_a p_{s,a} = 1$ for all states s. The basic idea is that action values of successful actions increase and actions with higher values get selected more often than actions with lower action values. Learning automata are similar but directly update the probability distribution over the actions.

I apply these algorithms (see Chapters 3 and 4) directly to games in extensive form and not to their corresponding strategic form to avoid information loss. To do this, each information set corresponds to a state. For signaling games, *Sender* needs an action value $q_{t,m}$ and a probability $p_{t,m}$ for all types t and signals m. Likewise, *Receiver* needs an action value $q_{m,r}$ and a probability $p_{m,r}$ for all signals m and responses r. So, signals take the role of actions for *Sender* and the role of states for *Receiver*. Of course, learning automata only have probabilities $p_{s,a}$ for all state-action pairs (s, a) and no action values $q_{s,a}$. When the game is finished, each agent updates the action values (or action probabilities) for the state he observed. This method works in extensive form games as long as a player only takes one action per game. If a player takes more than one action per game, he must apply some strategy to credit the obtained reward to the different actions he took. Q-learning as introduced by Watkins (1989) is such a method—it credits a fraction of the reward to actions which were taken some time ago. Players do not need such capabilities in signaling games because they reach just one state and take only one action per game. The reward is easily credited to that action. Therefore I only discuss a simplified version of Q-learning: the so-called '*single-state Q-learning.*'

Though these algorithms can handle stochastic rewards, I do not consider such scenarios. As mentioned earlier, a payoff represents an agent's preference for an outcome and I consider these preferences deterministic as is usual in game theory.

All three algorithms belong to the class of reinforcement learning. One of their advantages is simplicity: they require little information and little computational capacity. For example, they are unaware of other agents and learn as if they optimize against *Nature*. They are so-called boundedly rational and this makes them a more realistic model of the cognitive capabilities of many living organisms, including humans, than the perfect rationality assumed by classic game theory (Arthur, 1993). Their simplicity also allows to execute them on constrained electronic devices, such as sensor nodes (Mihaylov, 2012).

Another advantage of these algorithms is that they directly act on the extensive form game. For example, Roth-Erev learning was developed as a model of how people learn in extensive form games (Roth and Erev, 1995). Nash equilibria are only defined in strategic form and transforming a game from extensive form to strategic form may incur information loss (Section 2.2.2). There exist equilibria concepts which can be applied to extensive form games based on the notion of sequential rationality—an equilibrium strategy is optimal in all information sets, not just in those played in equilibrium (Appendix B). In signaling games, the learning algorithms are doing just that, they optimize their play in each information set.

Finally, I find ϵ-greedy Q-learning (see further on) one of the most useful adaptive heuristics there is since

- it poses no constraints on the payoffs,
- its action values are meaningful: each action value converges to the action's expected payoff if it is stationary and will otherwise fluctuate,
- although invented for single agent scenarios (Watkins, 1989), it is particularly suited to competitive multiagent scenarios where continuous adaptation and exploration is needed to avoid being exploited by other agents (Catteeuw and Manderick, 2011b), and
- in cooperative settings, where agents have common interests, it often settles down on an equilibrium (Claus and Boutilier, 1998).

Roth-Erev learning and learning automata do not have these benefits but I introduce them here for the specific setting of Chapter 3 where all three algorithms exhibit the same dynamics for certain parameters.

Roth-Erev learning

Roth-Erev learning (Roth and Erev, 1995) has two parameters:

- a discount factor $\lambda \in [0, 1]$ and
- an initial action value $Q_0 \geq 0$.

All action values are initialized to Q_0. Given the current state s, action a is selected with probability $p_{s,a}$ proportional to its action value $q_{s,a}$:

$$p_{s,a} = \frac{q_{s,a}}{\sum_{a'} q_{s,a'}}$$

This assumes that all action values (and consequently, payoffs) are non-negative: $q_{s,a} \geq 0$ for all states s and actions a. When all action values are 0, each action is selected with equal probability.

After taking action a and receiving payoff u, all action values for the current state s are discounted by factor λ and the action value of the current action is incremented with payoff u:

$$q_{s,a} \leftarrow \begin{cases} \lambda q_{s,a} + u & \text{if action } a \text{ was taken,} \\ \lambda q_{s,a} & \text{otherwise.} \end{cases}$$

Action values for states other than the current one are not updated.

In Roth and Erev's basic model the discount factor $\lambda = 1$ and the initial action value $Q_0 = 1$. A discount factor $\lambda < 1$ helps forgetting old experience. It bounds the action values to $u/(1 - \lambda)$, which makes it possible for the algorithm to settle down. If the discount factor $\lambda = 1$, the action values are unbounded and the system never settles down. Using small initial action values ($Q_0 < u$) speeds up learning in the beginning because it increases the importance of the payoffs u.

Single-state Q-learning

In Q-learning, all action values are initialized to $Q_0 \in \mathbb{R}$. High initial action values increase the amount of exploration in the beginning of the learning process (Sutton and Barto, 1998, sect. 2.7).

Q-learning can be combined with different action selection rules, like ϵ-greedy and softmax action selection. I describe and use both. With probability ϵ, *ϵ-greedy action selection* selects an action at random and with probability $1-\epsilon$, it selects the action with the highest action. In the latter case, if multiple actions have the maximum action value, they are selected with equal probability. If $\epsilon = 0$ there is no exploration. I call this 'greedy action selection.'

Softmax action selection selects an action a for current state s with probability $p_{s,a}$ according to the Boltzmann distribution:

$$p_{s,a} = \frac{e^{(q_{s,a}/\tau)}}{\sum_{a'} e^{(q_{s,a'}/\tau)}},$$

where temperature τ controls the rate of exploration: much exploration at high temperature, little exploration at low temperature.

To let the behavior stabilize, the exploration rate ϵ or temperature τ can be decreased over time. I either decrease them fast: $\epsilon(i) = \min\{\epsilon, \epsilon/i\}$, or slow: $\epsilon(i) = \min\{\epsilon, \epsilon \log(i)/i\}$, where $\epsilon(i)$ is the exploration rate used at the ith iteration and ϵ is determined by the user. For softmax action selection, simply replace ϵ by τ.

After taking action a in state s and receiving payoff u, the action value $q_{s,a}$ is updated while the other action values remain unchanged:

$$q_{s,a} \leftarrow \begin{cases} q_{s,a} + \alpha(u - q_{s,a}) & \text{if action } a \text{ was taken,} \\ q_{s,a} & \text{otherwise,} \end{cases}$$

where $\alpha \in [0,1]$ is the learning rate. A higher learning rate puts more weight on more recent payoffs. If α is 0, nothing is ever learned; if α is 1, the action value of an action simply equals the last payoff earned for that action.

The algorithm has three parameters:

- a learning rate $\alpha \in [0, 1]$,
- an initial action value $Q_0 \in \mathbb{R}$, and
- an exploration rate $\epsilon \in [0, 1]$ in the case of ϵ-greedy action selection or a temperature $\tau > 0$ in the case of softmax action selection.

Learning automata

Learning automata (Narendra and Thathachar, 1989) directly update the probability distribution over the actions and have two parameters:

- a reward factor $\alpha \in [0, 1]$ and
- a penalty factor $\beta \in [0, 1]$.

The probability distribution over the actions starts off uniform and changes as follows after taking action a in state s and receiving payoff u:

$$p_{s,a} \leftarrow \begin{cases} p_{s,a} + \alpha\, u\,(1 - p_{s,a}) - \beta\,(1 - u)\, p_{s,a} & \text{if action } a \text{ was taken,} \\ p_{s,a} - \alpha\, u\, p_{s,a} + \beta\,(1 - u)\left(\frac{1}{n-1} - p_{s,a}\right) & \text{otherwise,} \end{cases}$$

where n is the number of actions in state s. The update rule requires that the payoff $u \in [0, 1]$.

There are several well-known schemes, one is Linear-Reward-Inaction (L_{R-I}), another is Linear-Reward-ϵ-Penalty ($L_{R-\epsilon P}$). In Linear-Reward-Inaction, penalty factor $\beta = 0$, and thus, it only updates the action probabilities on reward (payoff $u > 0$). In Linear-Reward-ϵ-Penalty, β is a fraction of α. When in doubt, it is often a good idea to set the reward factor α close to 1 and the penalty factor β close to 0 (Catteeuw and Manderick, 2011b). In strategic form games, Linear-Reward-Inaction will always converge to a pure strategy Nash equilibrium if one exists, except in zero-sum games (Wheeler and Narendra, 1986). These are games where the sum of the players' payoffs is zero and player's have purely conflicting interests. One player's loss is the other's gain.

Conclusion and comparison

All three algorithms are variations on the same principle: actions yielding higher payoffs are used more frequently. Still, they are not applicable in the same settings and have different dynamics in general. These algorithms have been extensively studied before but mostly in strategic form games. In some settings, they converge to an equilibrium (Beggs, 2005; Claus and Boutilier, 1998; Wheeler and Narendra, 1986). In Chapter 3, I will prove that for some parameters all three algorithms will always reach a Pareto optimal equilibrium in Lewis signaling games.

Here are the main similarities and differences between these algorithms:[8]

Payoff constraints Q-learning poses no constraints on the payoffs and initial action values. Roth-Erev learning requires non-negative payoffs and initial action values: $u, Q_0 \geq 0$. Learning automata require payoffs between 0 and 1: $u \in [0, 1]$.

Exploration Roth-Erev learning and learning automata have no direct means of controlling the exploration rate. It is implicitly maintained by the action selection rule. Q-learning has direct means of controlling the exploration rate via the parameter ϵ, and controlling the *initial* exploration rate via the initial action value Q_0. Higher initial action values increase initial exploration.

Learning rate The user can control the *initial* learning rate of Roth-Erev learning via the initial action value Q_0 and the learning rate via discount factor λ. For both parameters, smaller values speed up learning. Q-learning has a constant learning rate α. Learning automata have a learning rate to reward and penalize: α and β.

Action values Q-learning's action values are an exponentially weighted moving average of the actions' observed payoffs. Roth-Erev's action

[8] I am aware that many of these differences can be circumvented or are non-existent in other versions of the same algorithms, but I prefer to limit this discussion to the variants introduced here.

values are the discounted sum of the actions' observed payoffs. Learning automata have no action values.

Chapter 3

Common Interest

This chapter concerns the first of three mechanisms that facilitate the emergence of honest signaling. It is based on the publication (Catteeuw and Manderick, 2014). The next two chapters concern the emergence of honest signals when interests conflict.

Contents

3.1 Introduction

This chapter discusses the emergence of signaling under common interest. When signaling is in the best interest of both the informed and uninformed agent, signals will definitely be honest. The question remains how signals can acquire their meaning. This chapter provides evidence that meaning can emerge due to very simple random processes.

The prototypical game for signaling under common interest is the *Lewis signaling game* (Section 3.2). It has many equilibria. The separating ones are Pareto optimal, the (partial) pooling ones are Pareto dominated. The Lewis signaling game is thus a coordination game where the players need to coordinate on the meaning of the signals.

I study the game with the fixed player model, where the same agents repeatedly interact in the same role. Some adaptive heuristics easily get stuck in a Pareto dominated pooling equilibrium while others always lead to a separating equilibrium in theory, but require too much time in practice for all but the smallest games (with no more than five or six types). Section 3.6 discusses this related work in detail.

In Section 3.3, I define the new adaptive heuristic *win-stay/lose-inaction* (WSLI) which initially behaves randomly, then repeats forever what was once successful. An analysis of WSLI in Lewis signaling games proves that it always reaches a separating equilibrium and predicts the number of interactions needed to do so. For Lewis signaling games with uniform type distributions, the expected number of interactions needed to find a separating equilibrium is approximately n^3, where n is the size, or the number of types, of the game.

For some parameters, Roth-Erev learning, Q-learning, and learning automata (Section 2.3.2) behave exactly like WSLI (Section 3.4).

Section 3.5 introduces errors in the Lewis signaling game. The results for WSLI in the original Lewis signaling game cannot be generalized to this case since WSLI may learn suboptimal behavior and never forget it. Roth-Erev learning, Q-learning, and learning automata still learn to signal optimally provided the parameters are slightly adjusted: they always reach a separating equilibrium in all Lewis signaling games in polynomial time even when errors may occur.

Contributions

- In Section 3.3, I define a new adaptive heuristic, called 'win-stay/lose-inaction' (WSLI).
- I prove that it always reaches a separating equilibrium in all Lewis signaling games in polynomial time: $\mathcal{O}(n^3)$, where n is the size, or the number of types, of the game. No such algorithm was known before.
- In Section 3.4, I show for which parameters Roth-Erev learning, Q-learning, and learning automata behave exactly like WSLI.
- In Section 3.5, I show for which parameters Roth-Erev learning, Q-learning, and learning automata still learn to signal optimally when errors occur.

3.2 The Lewis Signaling Game

A Lewis signaling game (Lewis, 1969) is completely determined by its type distribution π. Figure 3.1 shows the one with type distribution $\pi = (1/2, 1/2)$. It is a signaling game (Section 2.2.4 and Definition 2.5) with three constraints:

- There are an equal number of types, signals, and responses: $|\mathcal{T}| = |\mathcal{M}| = |\mathcal{R}| = n$.
- The set of signals $\mathcal{M}$ is independent of the type and the set of responses $\mathcal{R}$ is independent of the signal.
- Both players $i = \mathit{Sender}, \mathit{Receiver}$ have the same payoff function u_i:

$$u_i(t_j, m_k, r_l) = \begin{cases} 1 & \text{if } j = l, \\ 0 & \text{otherwise} \end{cases}$$

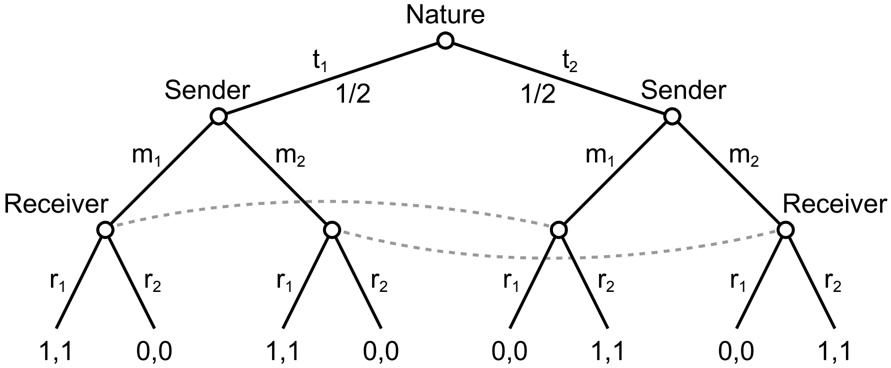


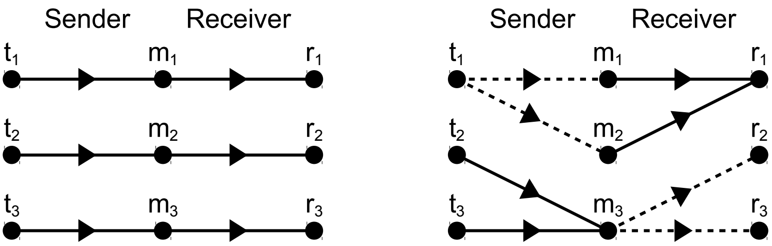
Figure 3.1: The Lewis signaling game with type distribution $\pi = (^1\!/_2, ^1\!/_2)$ and so two types, signals, and responses ($n = 2$). It is a signaling game (Section 2.2.4 and Definition 2.5) with common interest since both players always get the same payoff. The game is successful when *Receiver* chooses the response corresponding to *Sender*'s type. If the players manage to coordinate on the meaning of the signals, *Receiver* will be able to infer *Sender*'s type from his signal and they will be rewarded.

This payoff function has three characteristics:

- both players get the same payoff for all outcomes, so the game is fully cooperative;
- the signal does not (directly) influence the payoff; and
- for each type t_j there is exactly one correct response r_j, similarly, each response r_l is correct for exactly one type t_l.

This game is fully cooperative, so it is in both players' interest that *Receiver* can reliably deduce *Sender*'s type. In essence: the agents face a communication problem, which they can solve by establishing a shared language or convention. *Sender* and *Receiver* should adopt mappings (from types to signals and from signals to responses, respectively) which will be compatible if they, when applied one after the other (first *Sender*'s, then *Receiver*'s mapping), lead to the correct response for all types. Compatible mappings correspond to separating equilibria that are Pareto optimal Nash equilibria. Lewis calls them 'signaling systems.' For a Lewis signaling game with n types there are $n!$ such equilibria, corresponding to the $n!$ unique mappings from types to signals and those from signals to responses. One separating equilibrium is shown in Figure 3.2a for a game with three types ($n = 3$). In this figure, the *signaling success rate*—the probability that the two agents will have a successful interaction—is 1. Since the payoff for success is 1 and the payoff for failure is 0, the expected payoff equals the signaling success rate.

As most signaling games, a Lewis signaling game also has many pooling or partial pooling equilibria especially when the number of types is larger than two ($n > 2$). Such equilibria are Pareto dominated by the separating equilibria, thus suboptimal. Figure 3.2b shows a partial pooling equilibrium where *Sender* uses signal $\mathtt{m}_3$ both for type $\mathtt{t}_2$ and $\mathtt{t}_3$. When observing signal $\mathtt{m}_3$, *Receiver* can only guess what is the true type of *Sender*. Assuming that all types are equally likely, the signaling success rate (and the expected payoff) for the agents is $^2/_3$: type $\mathtt{t}_1$ always yields success, while type $\mathtt{t}_2$ and $\mathtt{t}_3$ yield success only half of the time. This state is an equilibrium,



(a) A separating equilibrium. For each type t_i *Sender* sends a distinct signal m_i.

(b) A partial pooling equilibrium. For both types $\mathtt{t}_2$ and $\mathtt{t}_3$ *Sender* sends the same signal $\mathtt{m}_3$.

Figure 3.2: Agent strategies for a Lewis signaling game with three types ($n = 3$). *Sender*'s strategies map types t to signals m and *Receiver*'s strategies map signals m to responses r. A solid line represents a probability of 1. A dashed line represents a probability of $^1/_2$.

since, neither *Sender* nor *Receiver* can change his strategy to increase his payoff. *Sender*'s strategy is a best response to *Receiver*'s strategy, and vice versa, *Receiver*'s strategy is a best response to *Sender*'s strategy. It is the existence of many such suboptimal equilibria that makes it hard to find an optimal one.

Another difficulty arises when the type distribution π is non-uniform, such as $\pi = (^3/_4, ^1/_4)$ instead of $\pi = (^1/_2, ^1/_2)$. For example, if one of the types has a probability of 90%, *Receiver* can simply ignore the signals, can always pick the action corresponding to the most frequent type, and will already be successful 90% of the time.

As briefly mentioned in Section 3.1 and more extensively discussed in Section 3.6, many adaptive heuristics may get stuck in these suboptimal equilibria and never reach a Pareto optimal one. Others are guaranteed to find a Pareto optimal equilibrium in theory, but require too much time in practice. In the next section, I introduce WSLI and show it overcomes both difficulties and finds a Pareto optimal equilibrium fast.

3.3 Win-Stay/Lose-Inaction (WSLI)

I define the new adaptive heuristic WSLI as follows:

- Initially, play random.
- Repeat forever the first action that yields success.

So, WSLI never changes its behavior after a failure, whether this is due to an action that was randomly chosen or one that was previously successful.

3.3.1 WSLI in Lewis signaling games

As an example, consider how WSLI reaches a separating equilibrium in a Lewis signaling game with two types ($n = 2$). Figure 3.3a shows the initial strategies for both *Sender* (mapping types t to signals m) and *Receiver* (mapping signals m to responses r). Their behavior is random and will remain this way until the first successful interaction.

After the first success, one can relabel the successful type, signal, response as $\mathtt{t}_1$, $\mathtt{m}_1$, and $\mathtt{r}_1$, respectively, without loss of generality. From now on, *Sender* will always use signal $\mathtt{m}_1$ when observing type $\mathtt{t}_1$ and *Receiver* will always respond with $\mathtt{r}_1$ to signal $\mathtt{m}_1$. We say a path $\mathtt{t}_1 \to \mathtt{m}_1 \to \mathtt{r}_1$ is learned for type $\mathtt{t}_1$. Behavior for other types and signals remains random (Figure 3.3b). Several things can happen now.

1. *Nature* draws type $\mathtt{t}_1$. This will trigger *Sender* to send signal $\mathtt{m}_1$, and consequently, *Receiver* will respond with $\mathtt{r}_1$. This always leads to success, and the path $\mathtt{t}_1 \to \mathtt{m}_1 \to \mathtt{r}_1$ persists. In general, whenever a type occurs for which a path was already learned, the interaction is successful and the agents do not change their behavior.
2. *Nature* can also draw $\mathtt{t}_2$ (or, more generally, a type for which no path is yet learned). *Sender* will signal at random and there are again two possibilities:
   (a) If he picks signal $\mathtt{m}_1$ (or, more generally, a signal which is already used in a learned path), then *Receiver* will definitely respond

with $\mathtt{r}_1$ and the game fails. The agents do not update their behavior in that case.

(b) If he picks signal $\mathtt{m}_2$ (or, more generally, a signal which is not yet used in a learned path), then *Receiver* either guesses the incorrect response or the correct response. In the former case, the behavior of the agents remains the same. In the latter case, they update their behavior and a new path $\mathtt{t}_2 \rightarrow \mathtt{m}_2 \rightarrow \mathtt{r}_2$ is learned.

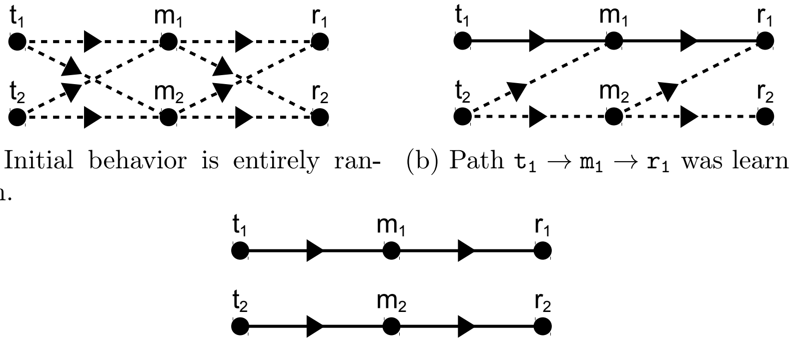
The agents' strategies are in a separating equilibrium and each interaction will be successful (Figure 3.3c). It is straightforward to generalize the reasoning above to all Lewis signaling games: strategies change only if

- *Nature* selects a type t that does not yet belong to a learned path,
- *Sender* selects a signal m that does not yet belong to a learned path, and
- *Receiver* selects the correct response r.

In this case, a new path $t \rightarrow m \rightarrow r$ is learned and interactions for type t will always succeed. When a path is learned for all types, agents are in a separating equilibrium. Due to symmetry, each of the $n!$ separating equilibria is reached with equal probability.

3.3.2 The learning process modeled by a Markov chain

Given how WSLI behaves in Lewis signaling games, one can prove that it always reaches a separating equilibrium and predict the average number of iterations needed. The learning process can be modeled by a Markov chain that has a state for each possible subset of the set of types $\mathcal{T}$. A state represents the types for which a path is already learned and interactions are always successful. Figure 3.4a shows the Markov chain for Lewis signaling games with three types ($\mathcal{T} = \{\mathtt{t}_1, \mathtt{t}_2, \mathtt{t}_3\}$). Since WSLI can never forget a learned path, the Markov chain can never get into a state with fewer learned paths than the current state. It either remains in the same state, or it goes



(a) Initial behavior is entirely random.

(b) Path $\mathtt{t}_1 \to \mathtt{m}_1 \to \mathtt{r}_1$ was learned.

(c) Final strategies form a separating equilibrium.

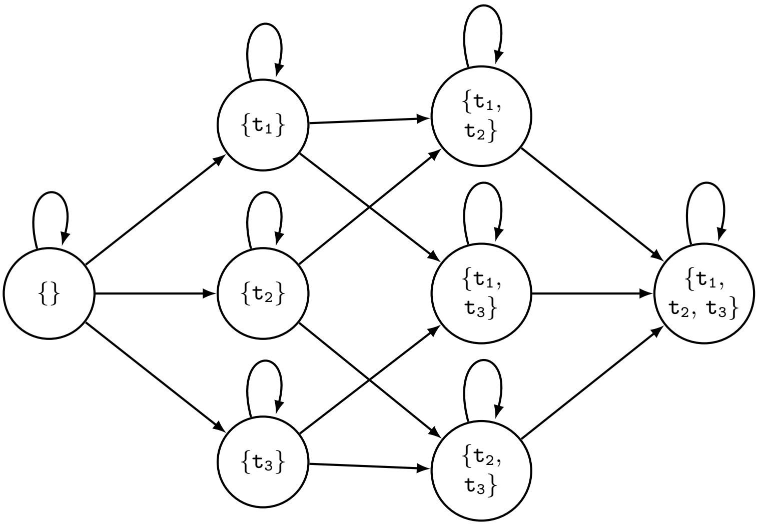
Figure 3.3: Emergence of a separating equilibrium in a Lewis signaling game with two types ($n = 2$) when both *Sender* and *Receiver* apply WSLI. Each figure shows *Sender*'s and *Receiver*'s strategy (mapping types t to signals m and signals m to responses r, respectively). A solid line represents the probability 1. A dashed line represents the probability $1/n = 1/2$.

to a state where one extra path is learned. There are 2^n states in total: one initial state with no types ($\binom{n}{0} = 1$), n states with one type ($\binom{n}{1} = n$), and so on until one final state with all types ($\binom{n}{n} = 1$).

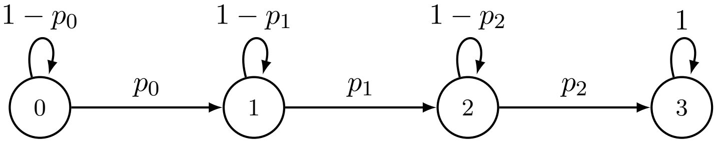
The probability to go from state $\mathcal{L}$ to $\mathcal{L}'$, where $\mathcal{L}'$ contains all types in $\mathcal{L}$ and one other type t that is not in $\mathcal{L}$, equals the probability that

- *Nature* selects type t,
- *Sender* selects a signal that is not yet used in a path, and
- *Receiver* selects the correct response.

The probability that *Nature* select type t is simply π_t. The probability that *Sender* selects an unused signal is $\frac{n-l}{n}$, where n is the number of signals and l is the number of learned paths and used signals. The probability of selecting the correct response is $1/n$, where n is the number of responses. The product of these probabilities is the probability of learning a path for



(a) General type distributions. Each state of the Markov chain is represented by the set of types for which the game is always successful.



(b) Uniform type distributions. Each state of the Markov chain is represented by the number of types for which the game is always successful. The transition probabilities p_l for $l = 0$, 1, and 2 are given by Equation (3.1).

Figure 3.4: The Markov chains for Lewis signaling games with three types ($n = 3$).

type t:

$$\Pr(\mathcal{L} \to t \cup \mathcal{L}) = \pi_t \frac{n-l}{n} \frac{1}{n},$$

where $l = |\mathcal{L}|$ and $t \notin \mathcal{L}$. The probability that the process remains in the same state is

$$\Pr(\mathcal{L} \to \mathcal{L}) = 1 - \sum_{t \notin \mathcal{L}} \Pr(\mathcal{L} \to t \cup \mathcal{L}).$$

Proving that WSLI always reaches a separating equilibrium is trivial.

Theorem 1. *For any Lewis signaling game, WSLI always reaches a separating equilibrium.*

Proof. For any Lewis signaling game with a probability distribution π over its typeset $\mathcal{T}$ of size n, the learning process generated by WSLI can be modeled by a Markov chain as described above.

For all states $\mathcal{L}$ of the Markov chain that are a strict subset of $\mathcal{T}$, the probability to go to a state with one more type is greater than zero:

$$\text{for all } \mathcal{L} \subset \mathcal{T}, t \notin \mathcal{L} : \ \Pr(\mathcal{L} \to t \cup \mathcal{L}) > 0.$$

So, these states are not absorbing and, by induction, the probability to go from any of these states to the state with all types $\mathcal{T}$ is greater than zero.

The state that contains all types $\mathcal{T}$ is an absorbing state and it is the only one:

$$\Pr(\mathcal{L} \to \mathcal{L}) = 1 - \sum_{t \notin \mathcal{L}} \Pr(\mathcal{L} \to t \cup \mathcal{L}) = 1 \text{ if and only if } \mathcal{L} = \mathcal{T}.$$

From the above follows that the process is always absorbed in the state that contains all types $\mathcal{T}$ and since this state represents the separating equilibria, the process always ends up in a separating equilibrium. □

3.3.3 Expected time until equilibrium

For any given Lewis signaling game, the method described in Appendix A calculates the expected number of iterations needed to reach a separating equilibrium. Here, I derive a general formula for Lewis signaling games with uniform type distributions and a lower and upper bound for Lewis signaling games with non-uniform type distributions.

Uniform type distributions

For Lewis signaling games with uniform type distributions, the Markov chain can be simplified, since all types have the same probability. For n types, $\pi_t = {}^1\!/_n$ for all types t. The *number* of learned paths is sufficient to discriminate all states $l = 0, 1, \ldots, n$ of the Markov chain (Figure 3.4b), where n is the total number of types and l the number of learned paths. The probability p_l to go from state l to state $l+1$ equals the probability of learning a new type-signal-response path which is the probability of selecting a type for which no path yet exists ($\frac{n-l}{n}$ because all types t occur with equal probability), selecting an unused signal ($\frac{n-l}{n}$), and selecting the correct response (${}^1\!/_n$):

$$p_l = \frac{n-l}{n}\frac{n-l}{n}\frac{1}{n} = \frac{(n-l)^2}{n^3}. \tag{3.1}$$

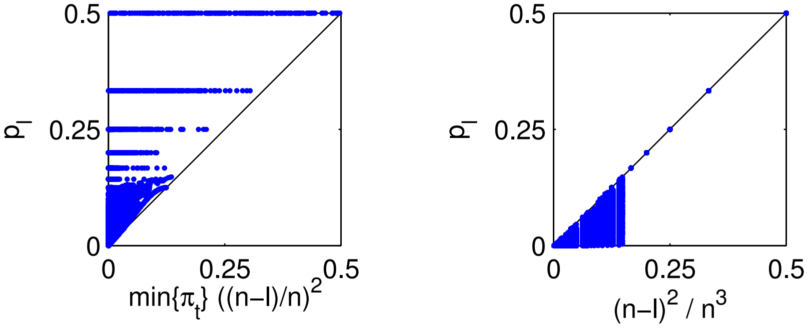
The process remains in the same state l with probability $1 - p_l$.

If, at each iteration, a new path is learned with probability p, then the expected number of iterations to learn a new path is ${}^1\!/_p$. The number of iterations is distributed according to a geometric distribution with mean ${}^1\!/_p$. The expected number of iterations $\mathbb{E}[T_c]$ to learn a path for all n types (a separating equilibrium) is the sum of the expected number of iterations needed for each new path:

$$\mathbb{E}[T_c] = \sum_{l=0}^{n-1} \frac{1}{p_l}. \tag{3.2}$$

Substituting p_l in Equation (3.2) by Equation (3.1) yields

$$\mathbb{E}[T_c] = \sum_{l=0}^{n-1} \frac{n^3}{(n-l)^2} = n^3 \sum_{i=1}^{n} \frac{1}{i^2}. \tag{3.3}$$



(a) $p_l \geq \frac{(n-l)^2}{n^2} \min_t\{\pi_t\}$ (b) $p_l \leq \frac{(n-l)^2}{n^3}$

Figure 3.5: The probability p_l that WSLI learns a path for a new type given that it has already learned l paths is bounded from above and below (Equation (3.4)). I verified this numerically for 10,000 randomly generated samples. The method is explained in the text.

The expected number of iterations is polynomial in the number of types n, $\mathbb{E}[T_c] = \mathcal{O}(n^3)$, because the sum $\sum_{i=1}^{n} {}^{1}\!/\!{}_{i^2}$ quickly converges to ${}^{\pi^2}\!/\!{}_{6} \approx 1.64$ while n goes to infinity (Daners, 2012).

Non-uniform type distributions

For Lewis signaling games with non-uniform type distributions I cannot provide an exact formula. Deriving an exact formula is hard (or maybe impossible) since there are many paths through the Markov chain. The expected number of iterations to traverse a path varies from path to path and so does the probability of taking each path.

It is possible to bound the probability p_l to go from a state with l types to a state with $l+1$ types given that the process is currently in a state with l types:

$$\frac{(n-l)^2}{n^2} \min_t\{\pi_t\} \leq p_l \leq \frac{(n-l)^2}{n^3}, \tag{3.4}$$

where $\min_t\{\pi_t\}$ is the probability of the rarest type. Figure 3.5 shows experimental evidence that these bounds hold. I generated 10,000 samples by calculating p_l for randomly chosen type distributions and different values of l. The probability p_l for any given type distribution and given l can be derived from the transition matrix of the corresponding Markov chain (as explained above) and the expected time spent in each state of the Markov chain (Appendix A.3).

Using the same reasoning as for uniform type distributions, the expected time $\mathbb{E}[T_c]$ is given by substituting Equation (3.4) in Equation (3.2):

$$n^3 \sum_{i=1}^{n} \frac{1}{i^2} \leq \mathbb{E}[T_c] \leq \frac{n^2}{\min_t\{\pi_t\}} \sum_{i=1}^{n} \frac{1}{i^2}. \tag{3.5}$$

For WSLI, Lewis signaling games with uniform type distributions are the easiest one. It needs less time, on average, to reach a separating equilibrium when the type distribution is uniform than when it is non-uniform. The good news is that the time needed for games with non-uniform type distributions, is still polynomial in the size of the game n.

3.4 Reinforcement Learning

In this section, I show how to implement WSLI with three well-known reinforcement learning algorithms thereby explaining why they perform well in Lewis signaling games. In the next section, I show that these algorithms not only perform as well as WSLI but are also robust to errors.

3.4.1 Implementing WSLI

The three reinforcement learning rules from Section 2.3.2—Learning Automata, Roth-Erev learning, and Q-learning—implement WSLI in Lewis signaling games for certain parameters (Table 3.1).

Learning automata do so when the reward factor $\alpha = 1$ and penalty factor $\beta = 0$ (also known as 'Linear-Reward-Inaction'). First, initial behavior is random. Second, when an interaction fails (payoff $u = 0$), the

algorithm	parameters		
Learning automata	$\alpha = 1$	$\beta = 0$	
Q-learning	$0 < \alpha < 1$	$\epsilon = 0$	$Q_0 = 0$
Roth-Erev learning	$0 < \lambda \leq 1$	$Q_0 = 0$	

Table 3.1: The parameters for which the three reinforcement learning algorithms implement WSLI.

probabilities over the actions do not change. This can be verified with the update rule:

$$p_{s,a} \leftarrow \begin{cases} p_{s,a} + \alpha\, u\, (1 - p_{s,a}) - \beta\, (1-u)\, p_{s,a} & \text{if action } a \text{ was taken,} \\ p_{s,a} - \alpha\, u\, p_{s,a} + \beta\, (1-u) \left(\frac{1}{n-1} - p_{s,a}\right) & \text{otherwise,} \end{cases}$$

Third, when an interaction succeeds (payoff $u = 1$), the probability of the action taken becomes 1 and all others 0, independent of the current probabilities. The first successful action is thus repeated forever. See again the update rule.

Roth-Erev learning implements WSLI when the initial action values $Q_0 = 0$ and the discount factor $0 < \lambda \leq 1$. Q-learning implements WSLI when the initial action values $Q_0 = 0$, the learning rate $0 < \alpha < 1$, and the exploration rate $\epsilon = 0$ (also known as 'greedy'). For Roth-Erev and Q-learning, it is somewhat harder to see this than for learning automata, because the action values do *change on failure* (the action value of the action resulting in failure is slightly decreased), but it would take an infinite number of failures before the probability distribution over the actions also changes.

The number of iterations WSLI needs, in theory, to reach a separating equilibrium (Equation (3.3)) matches the experimental results for the three reinforcement learning rules in different Lewis signaling games. Figure 3.6 shows this for learning automata (with reward factor $\alpha = 1$ and penalty factor $\beta = 0$) in Lewis signaling games with uniform type distributions π.

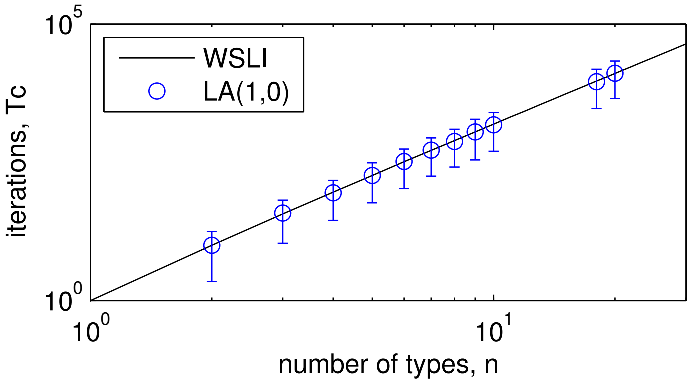


Figure 3.6: The theoretically expected number of iterations needed to find a separating equilibrium in Lewis signaling games with uniform type distributions for WSLI (solid black line, WSLI) matches the results for learning automata with reward factor $\alpha = 1$ and penalty factor $\beta = 0$ (LA(1,0)). The blue circles show the average number of iterations and the error bars show the standard deviation over 1,000 simulations.

3.4.2 Experiments

I now turn to some experimental results with these algorithms to find other parameter configurations where they perform well and concentrate on two performance criteria:

- whether or not a separating equilibrium is reached and
- the number of iterations needed to reach a separating equilibrium.

Figure 3.7 shows the results for various configurations of the three algorithms and the Lewis signaling game with type distribution $\pi = (\frac{1}{36}, \frac{2}{36}, \frac{3}{36}, \frac{4}{36}, \frac{5}{36}, \frac{6}{36}, \frac{7}{36}, \frac{8}{36})$. This Lewis signaling game is hard enough to show a clear difference between well and badly performing parameter configurations. Each experiment consisted of 1,000 runs. Per run, I recorded how many iterations were needed to reach a separating equilibrium. If, after 100,000 iterations, still no separating equilibrium was reached, the run was terminated and counted as a failure.

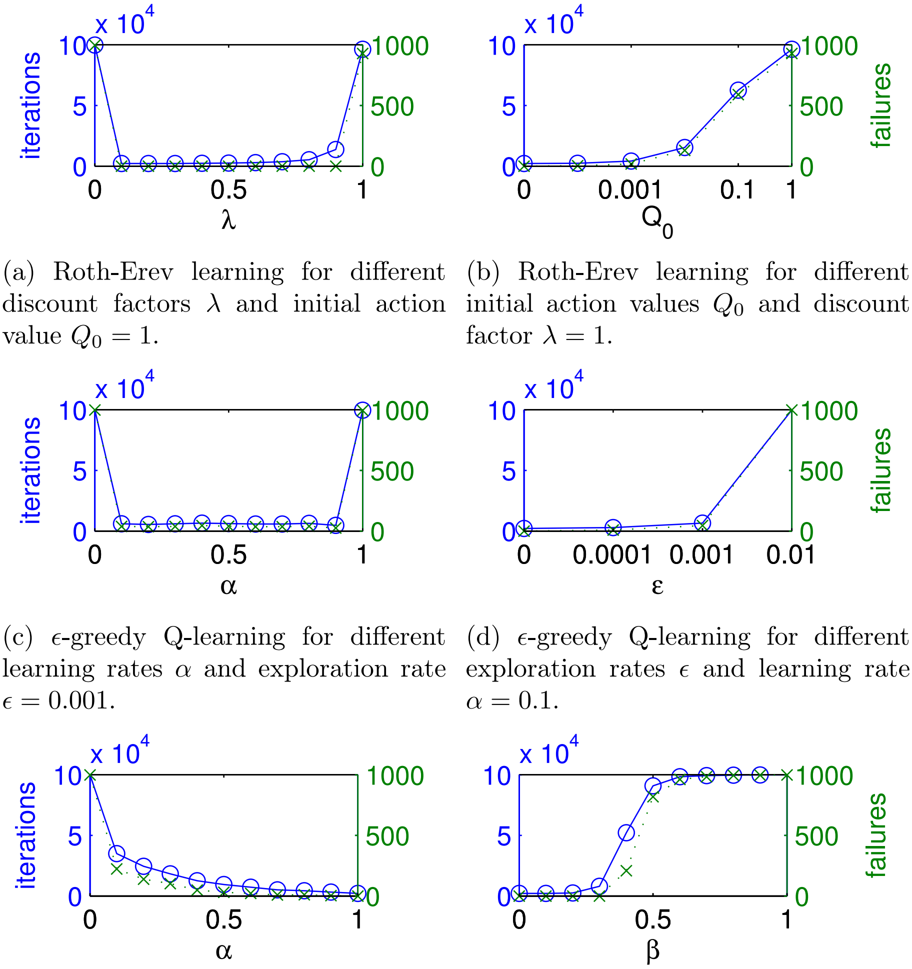
A separating equilibrium is reached when the signaling success rate—the probability that the next game will be successful—is above some *threshold* θ. A threshold $\theta = 1$ would be too strict, since some strategies cannot reach this level even though they can perform nearly optimal. For example, ϵ-greedy Q-learning can never achieve a signaling success rate of 1, unless the exploration rate $\epsilon = 0$. Therefore, I chose to set the threshold θ halfway between the optimal value, which is 1, and the signaling success rate of the best suboptimal equilibrium (a Pareto dominated partial pooling equilibrium). This allows to clearly distinguish between runs which do not end up in a separating equilibrium but are very near and runs that are closer to a partial pooling equilibrium than a separating one.

The *best suboptimal equilibrium* has a signaling success rate of 1 minus the probability of the rarest type: $1 - \min_t(\pi_t)$. The best suboptimal equilibrium is a partial pooling equilibrium, such as the one in Figure 3.2b, where *Sender* uses the same signal $\mathtt{m}_3$ for different types. When seeing this signal $\mathtt{m}_3$, *Receiver* cannot distinguish between the types $\mathtt{t}_2$ and $\mathtt{t}_3$, and hence can do no better than assuming the most frequent type. So, whenever *the other* type occurs, the game fails. In all other cases, the game succeeds. Thus, the signaling success rate in the *best* partial pooling equilibrium is determined by the frequency of the *rarest* type. For the example in Figure 3.7, the rarest type is $\mathtt{t}_1 = \min_t\{\pi_t\}$ with probability $\pi_{\mathtt{t}_1} = 1/36$ and the threshold is at $\theta = 71/72$, halfway between 1 and $1 - \min_t\{\pi_t\} = 35/36$.

The experiments revealed the following:

- Roth-Erev learning performs best with a small but positive discount factor ($0 < \lambda \ll 1$, Figure 3.7a). It performs better when initial action values are very small and performs best when they are zero ($Q_0 = 0$, Figure 3.7b). This was also reported by Skyrms (2010, p 97).

- Q-learning performs well at any learning rate $0 < \alpha < 1$ (Figure 3.7c), and best when playing greedy (exploration rate $\epsilon = 0$, Figure 3.7d). Combining Q-learning with other action selection strategies also revealed that playing greedy works best. Q-learning with softmax action selection performs well at low temperatures τ. Both ϵ-greedy and softmax Q-learning performs well with decreasing exploration rates ϵ and



(a) Roth-Erev learning for different discount factors λ and initial action value $Q_0 = 1$.

(b) Roth-Erev learning for different initial action values Q_0 and discount factor $\lambda = 1$.

(c) ϵ-greedy Q-learning for different learning rates α and exploration rate $\epsilon = 0.001$.

(d) ϵ-greedy Q-learning for different exploration rates ϵ and learning rate $\alpha = 0.1$.

(e) learning automata for different reward factors α and penalty factor $\beta = 0$.

(f) learning automata for different penalty factors β and reward factor $\alpha = 1$.

Figure 3.7: In each figure, the blue circles, connected with a solid blue line, show (on the left y-axis) the number of iterations needed to learn a separating equilibrium averaged over 1,000 and the green crosses, connected by a dotted green line, show (on the right y-axis) the number of runs out of 1,000 that failed to find a separating equilibrium in less than 100,000 iterations for the Lewis signaling game with type distribution $\pi = \left(\frac{1}{36}, \frac{2}{36}, \frac{3}{36}, \frac{4}{36}, \frac{5}{36}, \frac{6}{36}, \frac{7}{36}, \frac{8}{36}\right)$.

temperatures τ. Decreasing these parameters fast works better than decreasing them slowly. The rest of the text therefore only discusses ϵ-greedy action selection with a constant exploration rate ϵ.

- Learning automata perform better for high reward factors α (Figure 3.7e) and low penalty factors β (Figure 3.7f). They performs best when the $\alpha = 1$ and $\beta = 0$.

- Finally, all three algorithms perform equally well in the best case: they need slightly more than 2,000 iterations on average and always find a separating equilibrium in less than 100,000 iterations. For other Lewis signaling games we found similar results. As you may expect, more types and non-uniform distributions require more iterations.

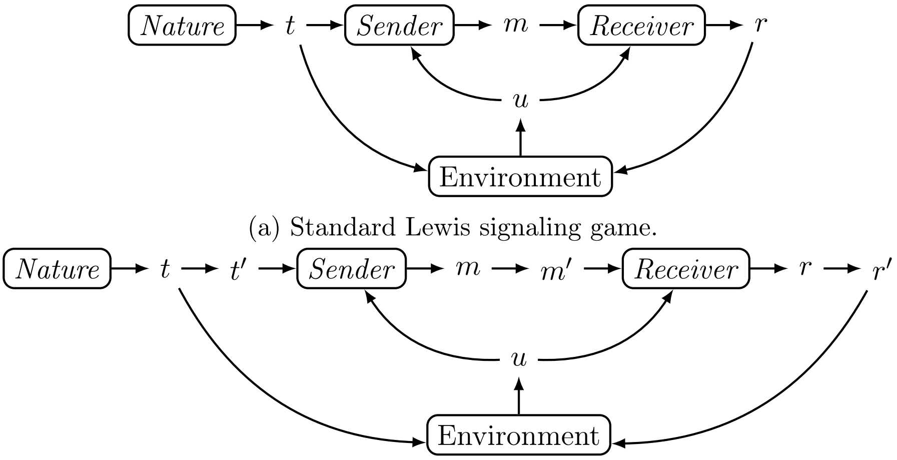
All three algorithms perform best when they mimic WSLI. Performance remains optimal for (slight) deviations of some parameters. Roth-Erev learning, for example, is also optimal for non-zero initial action values ($Q_0 \neq 0$) if the discount factor is small enough ($0 < \lambda \ll 1$). In the next section, I show how this allows the three algorithms to be robust to errors while WSLI cannot.

For completeness, I mention some algorithms that yielded unsatisfactory results. UCB1 (Auer et al., 2002) always found a separating equilibrium but is a factor slower than WSLI. Some algorithms sometimes, or always, failed to find a separating equilibrium. These are EXP3, EXP3.1, EXP3.S (Auer et al., 2003), and the Reinforcement Comparison and Pursuit method from (Sutton and Barto, 1998, ch. 2). In the remainder of the text, I focus on Roth-Erev learning, ϵ-greedy Q-learning, and learning automata.

3.5 Robustness

The results of the experiments in this section show that the reinforcement learning algorithms are robust to errors in the Lewis signaling game. Three types of errors may occur:

1. the type t' observed by *Sender* may be different from the true type t,



(a) Standard Lewis signaling game.

(b) Lewis signaling game with errors.

Figure 3.8: Flow diagrams representing Lewis signaling games. See Section 3.2 for a full explanation of the game.

2. the signal m' observed by *Receiver* may be different from *Sender*'s intended signal m, and

3. the true response r' may be different from *Receiver*'s intended response r.

The flow diagrams in Figure 3.8 illustrate the difference between the original game and the game with errors. Per interaction, at most three errors can occur. Each type of error occurs with a (small) fixed probability p_e, called the '*error rate.*' The probability that no error occurs is $(1 - p_e)^3$.

An algorithm which is robust to errors should avoid getting locked into suboptimal behavior which it cannot 'unlearn.' Unfortunately, this is a key characteristic of WSLI and hence it is not robust to errors. Whenever an error occurs and the interaction fails, no harm is done since WSLI does not update its behavior on failure. But whenever an error occurs and the interaction succeeds, chances are WSLI learns the wrong thing and will forever repeat its unsuccessful behavior. Consider the following example

algorithm	parameters		
Learning automata	$\alpha = 1$	$0 < \beta \ll 1$	
Q-learning	$0 < \alpha < 1$	$\epsilon = 0$	$Q_0 > 0$
Roth-Erev learning	$0 < \lambda < 1$	$Q_0 > 0$	

Table 3.2: The parameters for which the three reinforcement learning algorithms are robust to errors but still perform close to their optimal in the original Lewis signaling game.

that leads to success, but where one error occurs. *Nature* selects $\mathtt{t}_1$ as *Sender*'s true type but *Sender* wrongfully observes $\mathtt{t}_2$. Next, *Sender* selects signal $\mathtt{m}_2$ and *Receiver* responds with $\mathtt{r}_1$. Since the true type $\mathtt{t}_1$ and the true response $\mathtt{r}_1$ match, the interaction succeeds. The agents have now learned the path $\mathtt{t}_2 \to \mathtt{m}_2 \to \mathtt{r}_1$ which fails under normal circumstances. WSLI can also learn correct behavior even though an error occurs, namely when errors 'cancel out.'

WSLI is only an idealized version of the reinforcement learning algorithms. Slight modifications of the most successful parameter configurations avoid getting locked in and make the algorithms robust to errors. A positive initial action value ($Q_0 > 0$) does that for Roth-Erev learning and Q-learning. A positive penalty factor ($\beta > 0$) does that for learning automata. Figure 3.9 shows the performance of these robust algorithms in Lewis signaling games with uniform type distributions and error rate $p_e = {}^1\!/_{100}$. On the one hand, Q-learning performs slightly worse than the other two algorithms (a factor, not bigger than two and decreasing with the size of the game n). On the other hand, Q-learning is more reliable (its standard deviation is only half of that of the other two algorithms).

The learning algorithms are now capable of unlearning previous experience but this does not decrease their performance in Lewis signaling games without errors provided that the discount factor of Roth-Erev learning $\lambda < 1$ and the penalty factor of learning automata $\beta \ll 1$. Table 3.2 summarizes these parameters. Previously successful behavior is only forgotten after some consecutive failures. For small error rates, truly successful behavior will rarely lead to failure and even more rarely to a long enough sequence of

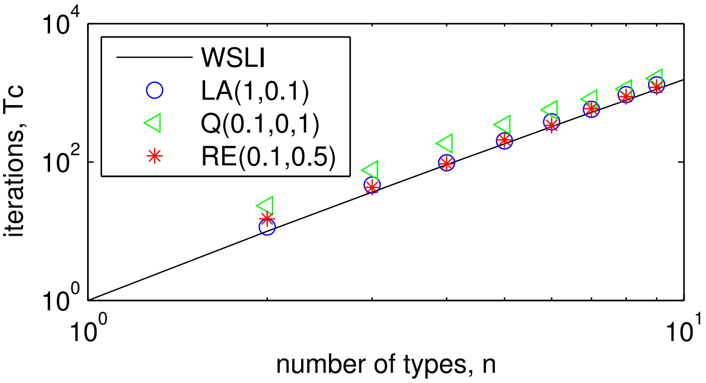


Figure 3.9: With error rate $p_e = {}^1\!/_{100}$, the robust algorithms perform almost as well as theory predicts for WSLI without errors (WSLI, black solid line). The data show the number of iterations, averaged over 100 runs, needed to reach a separating equilibrium in Lewis signaling games with uniform type distributions of different sizes $n = 2, 3, \ldots, 9$. The algorithms are learning automata with reward factor $\alpha = 1$ and penalty factor $\beta = 0.1$ (LA(1,0.1), blue circles); Q-learning with learning rate $\alpha = 0.1$, exploration rate $\epsilon = 0$, and initial action value $Q_0 = 1$ (Q(0.1,0,1), green triangles); and Roth-Erev learning with discount factor $\lambda = 0.1$ and initial action value $Q_0 = 0.5$ (RE(0.1,0.5), red asterisk).

failures, and thus will never be forgotten. Behavior that was only successful due to errors will soon yield a long sequence of failures and be forgotten.

3.6 Related Work

I briefly discuss the other work applying the fixed player model in Lewis signaling games (without errors).

Argiento et al. (2009) have proven that basic *Roth-Erev learning* (with initial action values $Q_0 = 1$ and discount factor $\lambda = 1$) always converges to a separating equilibrium in the Lewis signaling game with two equiprobable types ($\pi = \left(\frac{1}{2}, \frac{1}{2}\right)$). Basic Roth-Erev learning fails for games with more than two types ($n > 2$) and for games with non-uniform type distributions π (Barrett, 2006; Catteeuw et al., 2011; Huttegger, 2007). Skyrms (2010, p 97) reported that smaller initial action values ($Q_0 < 1$) increase the probability of reaching a separating equilibrium, even when the number of types is larger than two ($n > 2$) and the probability distribution over the types π is non-uniform. Here, I was able to explain this.

Barrett and Zollman (2009) apply *win-stay/lose-randomize* to Lewis signaling games and prove that it always reaches a separating equilibrium. Win-stay/lose-randomize repeats what is successful but chooses a random action after a failure. In the Lewis signaling game, it is equal to Roth-Erev learning when the discount factor $\lambda = 0$. Although this is theoretically very interesting, experiments show that the number of interactions needed to reach a separating equilibrium increases exponentially with the size of the game (Catteeuw et al., 2011). Catteeuw et al. (2011) show experimentally how Roth-Erev learning with a discount factor $0 < \lambda < 1$ always reaches a separating equilibrium and is much faster than win-stay/lose-randomize. Barrett (2006) studied two other variations of Roth-Erev learning for signaling games with an arbitrary number of types ($n \geq 2$) but uniform type distributions. One variation allows for negative rewards, the other randomizes action values. Both variations seem to help reaching a separating equilibrium, but do not guarantee it.

Barrett and Zollman (2009) also discuss *softmax Q-learning* (but call it ‘smoothed reinforcement learning’) and a more complex learning rule called ‘*ARP*’ (Bereby-Meyer and Erev, 1998). They conclude that forgetting old experience increases chances of finding a separating equilibrium. I found that this is unnecessary except for Lewis signaling games where errors occur.

Win-stay/lose-shift (which Skyrms (2010) calls ‘best response’) does not work in Lewis signaling games with two types ($n = 2$), since the process may get into endless loops. When there are more than two types ($n > 2$), the process must be redefined. For example, when loosing, it could pick any of the alternatives at random with equal probability. Skyrms (2010, pp 103-105) calls this process ‘best response for all we know.’ Still, this process cannot handle the case for $n = 2$ types, and Zollman proposes to add inertia: only now and then apply the best response rule. The rest of the time behavior is not updated.

3.7 Conclusion

In order to gain more insight into the emergence of signaling under common interest, I studied the Lewis signaling game in the fixed player model with different adaptive heuristics. There are three main results.

1. I introduced the new adaptive process WSLI. Markov chain analysis proves that it always reaches a separating equilibrium in all Lewis signaling games and predicts that the expected number of iterations needed is polynomial in the number of types: $\mathbb{E}[T_c] = \mathcal{O}(n^3)$.

2. Three reinforcement learning algorithms mimic WSLI in Lewis signaling games: learning automata with reward factor $\alpha = 1$ and penalty factor $\beta = 0$; greedy Q-learning with initial action value $Q_0 = 0$ and learning rate $0 < \alpha < 1$; and Roth-Erev learning with initial action value $Q_0 = 0$ and discount factor $0 < \lambda \leq 1$.

3. Slight adaptations render these algorithms robust without decreasing their performance in the original Lewis signaling game. The resulting configurations are: learning automata with reward factor $\alpha = 1$ and

small positive penalty factor $0 < \beta \ll 1$; greedy Q-learning with positive initial action value $Q_0 > 0$ and learning rate $0 < \alpha < 1$; and Roth-Erev learning with positive initial action value $Q_0 > 0$ and discount factor $0 < \lambda < 1$.

These results tell us that *under very weak assumptions signaling can emerge by chance and can do this reasonably fast.* Though the number of interactions needed to reach a separating equilibrium is polynomial in the number types n, there is already some successful signaling while learning. Admittedly, for large communication systems, one would prefer time grows only linear or even sublinear with the number of types n, and so other solutions are necessary. For example, if an agent could know which signal he uses for which type, he could avoid using that same signal for other types. This would definitely speed up learning, but it imposes extra requirements on the agents' cognitive capabilities. This is exactly what researchers in the domain of language games are doing (De Beule et al., 2006; Steels, 1999, 2001).

Chapter 4

Costly Signals

The previous chapter studied the emergence of signaling under common interest, but most interactions are characterized by conflicting interests. This chapter examines when costly signals allow the emergence of signaling if interests conflict. It is based on the publications (Catteeuw and Manderick, in press; Catteeuw et al., 2013). The next chapter studies another alternative: punishment.

Contents

4.1 Introduction

In economics, Spence's job market model (Spence, 1973) (Example 1.5) shows that a university degree can work as an honest signal when applying for a job since there is a cost of acquiring that degree. More importantly, the degree is increasingly more costly to acquire for less skilled employees. As such, higher skilled employees invest in a higher degree than lower skilled employees and the employer, who is unable to directly observe the employees' abilities, has good reasons to believe that job candidates with higher degrees have higher abilities.

Honest signaling is important in many other economic applications with private information, such as product advertisement where the seller does and the buyer does not know the quality of the product. The seller can invest in costly advertisement to signal the quality of his product. Riley (2001) provides an overview of signaling in economics.

Zahavi (1975, 1977) discovered the same principle independently and named it the '*handicap principle*.' He claims that male characteristics used for sexual selection, such as the peacock's tail, the extra large antlers of deer, or the colorful plumage of male birds, are honest signals of the males' quality because they are a handicap. The peacock's tail (Example 1.3), for example, makes it harder for the peacock to escape from predators. Since only the fittest can afford the largest tails, females can reliably infer which males would make better mates from the size of their tails. Maynard Smith and Harper (2003) provide an overview of signaling in animals.

The Philip Sidney game (Section 4.2) is the classic game theoretic model for signaling when interests conflict. While Grafen (1990) proved that the handicap principle can work, Maynard Smith (1991) introduced the Philip Sidney game as a simplification of Grafen's model that still captures all necessary details to illustrate that concept (Maynard Smith and Harper, 2003, ch. 2).

Until now, both economists and biologists have almost exclusively relied on static analyses of honest signaling. In a static analysis, one merely verifies stability according to some equilibrium concept. Economists are mostly concerned with the necessary requirements for which honest signaling is a unique Nash equilibrium or a refinement thereof. See Appendix B for more information on equilibria in signaling games. Similarly, biologists show when honest signaling can or cannot be an equilibrium (in this case, an evolutionarily stable strategy) in many variations of the Philip Sidney game. See for example (Bergstrom and Lachmann, 1997, 1998; Brilot and Johnstone, 2003; Grafen, 1990; Lachmann and Bergstrom, 1998; Maynard Smith, 1991).

But stability is not sufficient for honest signaling to emerge, it merely says that *if* it emerges it will persist (Lachmann and Bergstrom, 1998). The same critique has been formulated by Huttegger and Zollman (2010). Instead of investigating the evolutionary stability of honest signaling in the Philip Sidney game, they focus on the evolutionary dynamics of the game. They employed the replicator dynamics (Hofbauer and Sigmund, 1998; Maynard Smith, 1982) which describes how strategies may spread in an infinite, well-mixed population under the influence of natural selection. They discovered that in some cases honest signaling is less likely to evolve in the replicator dynamics than is otherwise suggested by the analysis of evolutionarily stable strategies (ESSs). Section 4.5 discusses this related work in more detail.

This chapter shows that honest signaling can emerge from initially random behavior through adaptive processes even when there is a conflict of interest. I consider two adaptive processes: individual learning dynamics (Section 4.3) and evolutionary dynamics in finite populations (Section 4.4). Each time, I show under which conditions honest signaling is an equilibrium but not the result of a dynamical process. When the signal cost or the degree of common interest is too high, the cost of signaling outweighs its benefit, so non-signaling equilibria emerge. The opposite also occurs: (partial) signaling emerges in settings where it cannot be stable because signals are too cheap.

Contributions

Section 4.3 contrasts individual learning (ϵ-greedy Q-learning in the random matching model) with the Nash equilibrium. Section 4.4 contrasts evolutionary dynamics in finite populations with evolutionary stability in finite and infinite populations.

- In individual learning dynamics, honest signaling emerges only if it is a Pareto optimal Nash equilibrium. This is possible when the signal cost and the degree of common interest are relatively low, otherwise the cost of signaling outweighs its benefits and non-signaling equilibria emerge.
- In individual learning dynamics, when signals are too cheap for honest signaling to be stable, partial signaling—where agents sometimes signal or respond honestly and sometimes dishonestly—emerged.
- I compare the evolutionary stability of honest signaling in finite and infinite populations and show the effect of population size and selection pressure. For high signal cost and low selection pressure, honest signaling may be evolutionarily stable in infinite but not in finite populations. In large populations with high selection pressure, honest signaling is stable under the same circumstances.
- Honest signaling is observed under the same circumstances in the evolutionary dynamics as in the individual learning dynamics. It is possible that honest signaling is not stable but still the most frequent strategy in evolutionary dynamics.

4.2 The Philip Sidney Game

In biology, the standard model of costly signaling is the Philip Sidney game (Maynard Smith, 1991). It is a signaling game (Section 2.2.4) and shown in Figure 4.1. *Sender* has two types: `healthy` and `needy` with probabilities p and $1-p$, respectively, so the probability distribution over the types is $\pi = (p, 1-p)$. In both cases, he can either signal at some cost c or be quiet at no cost. *Receiver* does not know *Sender*'s type, but he observes whether or not *Sender* signals. He has a resource and must decide whether or not to donate it to *Sender*.

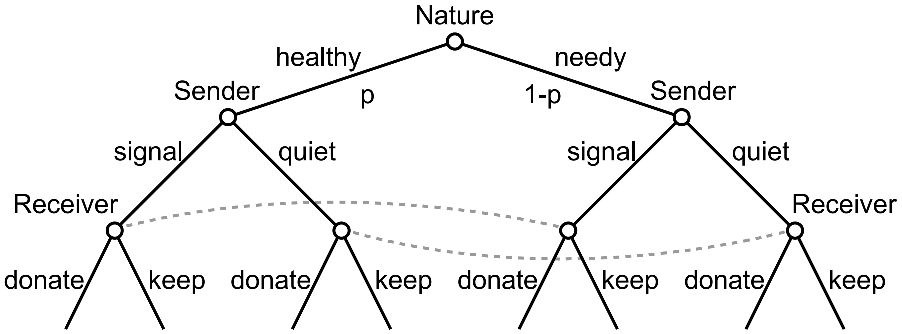


Figure 4.1: The Philip Sidney game is a signaling game with two types, two signals, and two responses. See Table 4.1 for the players' payoffs at each outcome.

To vary between conflicting and common interest, each player's payoff (Table 4.1) is the sum of his own survival probability plus a fraction R of the other player's survival probability. R is the players' *degree of common interest*. If $R = 1$, both players receive the same payoffs and interests are perfectly aligned. If $R = 0$, the players' interests are conflicting: *Sender* prefers to receive the resource while *Receiver* prefers to keep it.

In biology, R could represent the players' *coefficient of relatedness*—the fraction of genes that two organisms share through descent. The principle of *inclusive fitness* (Hamilton, 1964) states that genetically related organisms have some common interest: they benefit from each other's survival because they have more genes in common than two random organisms in the same population.

Remember that an outcome of a game in extensive form is a path from the root of the tree to a leaf node. For the Philip Sidney game these are 3-tuples (t, m, r) and consist of a type $t \in \{\texttt{healthy}, \texttt{needy}\}$, a signal $m \in \{\texttt{signal}, \texttt{quiet}\}$, and a response $r \in \{\texttt{donate}, \texttt{keep}\}$.

The survival probabilities are normalized and depend on the outcome of the game. *Receiver* is sure to survive if he keeps the resource, but if he donates the resource he survives with probability $S < 1$. If *Sender* receives the resource he is sure to survive, otherwise his survival probability depends on his type: if he is needy, he will die; if he is healthy, he will survive with

Table 4.1: *Sender*'s and *Receiver*'s payoffs for all possible outcomes of the Philip Sidney game. To vary the degree of common interest, a player's payoff is the sum of his own survival probability plus R times the other player's survival probability. See the text for an explanation of the players' survival probabilities. The parameters are listed below.

		`signal`		`quiet`	
		Sender	*Receiver*	*Sender*	*Receiver*
`healthy`	`donate`	$1-c+RS$	$S+R(1-c)$	$1+RS$	$S+R$
	`keep`	$V(1-c)+R$	$1+RV(1-c)$	$V+R$	$1+RV$
`needy`	`donate`	$1-c+RS$	$S+R(1-c)$	$1+RS$	$S+R$
	`keep`	$0+R$	$1+R0$	$0+R$	$1+R0$

parameter	meaning
$0<p<1$	probability that *Sender* is healthy
$0 \leq R \leq 1$	degree of common interest
$0<S<1$	*Receiver*'s survival probability without the resource
$0<V<1$	*Sender*'s survival probability when healthy without the resource
$0 \leq c \leq 1$	signal cost

probability $V < 1$. As already mentioned, signaling is costly: *Sender*'s survival probability is decreased by a factor $(1-c)$ if he signals. Table 4.1 shows all payoffs depending on the outcome and lists all parameters and their domain and meaning.

In the Philip Sidney game, both players have four pure strategies (Table 4.2). *Sender*'s pure strategies are $\texttt{S}_{\texttt{A}}$ (always signal), $\texttt{S}_{\emptyset}$ (never signal), $\texttt{S}_{\texttt{H}}$ (signal only when healthy), and $\texttt{S}_{\texttt{N}}$ (signal only when needy). *Receiver*'s strategies are $\texttt{D}_{\texttt{A}}$ (always donate), $\texttt{D}_{\emptyset}$ (never donate), $\texttt{D}_{\texttt{S}}$ (donate only when signal), and $\texttt{D}_{\texttt{Q}}$ (donate only when quiet). The strategic form of the Philip Sidney game can be derived from these pure strategies (Definition 2.1). Its payoff table is given in Appendix C.

Table 4.2: In the Philip Sidney game, both players have four pure strategies.

	strategy	meaning
Sender	$\mathtt{S}_{\mathtt{A}}$	always signal
	$\mathtt{S}_{\emptyset}$	never signal
	$\mathtt{S}_{\mathtt{N}}$	signal only when needy
	$\mathtt{S}_{\mathtt{H}}$	signal only when healthy
Receiver	$\mathtt{D}_{\mathtt{A}}$	always donate
	$\mathtt{D}_{\emptyset}$	never donate
	$\mathtt{D}_{\mathtt{S}}$	donate only when signal
	$\mathtt{D}_{\mathtt{Q}}$	donate only when quiet

The symmetric Philip Sidney game, which is the symmetric version of the Philip Sidney game in strategic form (Definition 2.1 and 2.2), has sixteen pure strategies. They are represented by tuples (X, Y). The first component determines what to do when in the role of *Sender* and the second one determines what to do when in the role of *Receiver*. For example, the honest signaling strategy is $(\mathtt{S}_{\mathtt{N}}, \mathtt{D}_{\mathtt{S}})$ (signal only when needy, donate only when signal). The payoffs of these pure strategies are calculated as explained at the end of Section 2.2.2.

There exist other variants of the Philip Sidney game, for example with continuous types, signals, or responses, but these yield qualitatively the same result (Maynard Smith and Harper, 2003, ch. 3).

The next sections, study the Philip Sidney game in more detail: the link with the handicap principle, the circumstances that make honestly signaling an evolutionarily stable strategy, and a classification of different types of conflict.

4.2.1 Evolutionary stability

On the one hand, the Philip Sidney game is a good model for the handicap principle since signals can be costly and *Sender* benefits more from receiving the resource when he is needy than when he is healthy. On the other hand, the cost of signaling does not depend on the state of *Sender*, but

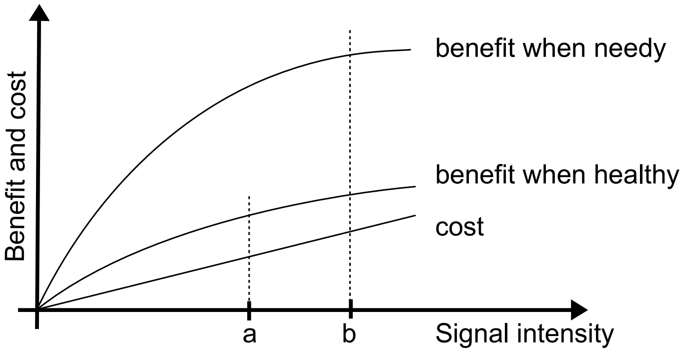
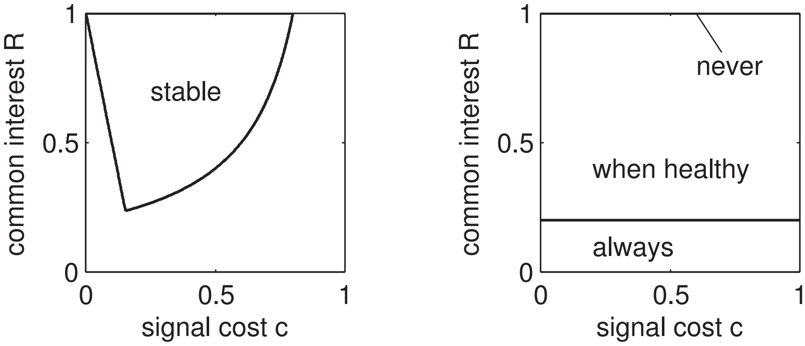


Figure 4.2: Visual representation of the handicap principle where signal cost and benefits increase with signal intensity. The benefit for a needy individual is higher than for a healthy one such that the optimal signaling intensity for needy individuals is higher than for healthy individuals: $a < b$.

only on the signal's intensity. Figure 4.2 visualizes this (assuming there is a continuum of possible signals, which is not the case for the Philip Sidney game, but the reasoning remains the same). Godfray (1991) and Johnstone and Grafen (1992b) use similar models where the signal indicates a level of need and depends only on its intensity, but *Sender* benefits' vary depending on his type. A typical example is a chick begging his mother for food to indicate it is hungry. In (Grafen, 1990) and (Johnstone and Grafen, 1992a) the signal indicates a quality and the same signal is more costly for low quality types than for high quality types. Another example is the peacock which signals his quality as a mate and parent with his tail (Example 1.3). In both scenarios, there exists an optimal signal intensity where the benefits maximally outweigh the costs (a and b in Figure 4.2). Depending on the parameters of the Philip Sidney game the optimal signal intensity for needy individuals may be higher than for healthy individuals ($a < b$ as in Figure 4.2), so that healthy individuals cannot profit from being dishonest: honest signaling is a Nash equilibrium.

The handicap principle only indicates that there is a *possibility* that honest signaling is stable. To verify its stability biologists usually rely on



(a) Stability.

(b) Conflict regions.

Figure 4.3: (a) Stability of honest signaling and (b) regions of conflict for all combinations of the signal cost c and the degree of common interest R, where the players' survival probability without the resource $S = V = {}^4\!/_5$. Both figures are independent of the probability p that *Sender* is healthy.

the concept of evolutionarily stable strategies (Definition 2.6). For two-player asymmetric games, the evolutionarily stable strategies coincide with the strict Nash equilibria (Section 2.3.1), so honest signaling ($\mathtt{S_N}, \mathtt{D_S}$) is evolutionarily stable if *Sender*'s strategy $\mathtt{S_N}$ is the best response to *Receiver*'s strategy $\mathtt{D_S}$ and *Receiver*'s strategy $\mathtt{D_S}$ is the best response to *Sender*'s strategy $\mathtt{S_N}$ (Maynard Smith, 1991).

Straightforward algebra shows that $\mathtt{S_N}$ is the best response to $\mathtt{D_S}$ whenever $R < 1 - c + RS$ and $1 - c + RS < V + R$. Similarly, $\mathtt{D_S}$ is the best response to $\mathtt{S_N}$ whenever $1 + RV > S + R$ and $S + R(1 - c) > 1$ (Maynard Smith, 1991). For the Philip Sidney game where $S = V = {}^4\!/_5$, Figure 4.3a shows for which combinations of the signal cost c and the degree of common interest R honest signaling is evolutionarily stable. The probability p that *Sender* is healthy has no influence.

4.2.2 Conflict and costly signals

Let us now verify when there is a conflict of interest. That is, when would *Sender* prefer a different outcome than *Receiver*. When *Sender* is needy, he prefers to receive the resource if $R(1-S) < 1$ (which is always satisfied) and *Receiver* prefers to keep it if $S + R < 1$ or $R < 1 - S$. The case where *Sender* does not want the resource, but *Receiver* wants to donate it is impossible. When *Sender* is healthy, he prefers to receive the resource if $R(1-S) < 1-V$ and *Receiver* prefers to keep it if $R(1-V) < 1-S$. The effect of the degree of common interest R is as expected. For $R = 0$ these conditions always hold, so there is a conflict. For $R = 1$ the conditions cannot hold at the same time, so there is no conflict. For the Philip Sidney game where $S = V = {}^{4}\!/_{5}$, Figure 4.3b shows there is a conflict when *Sender* is needy when $R < {}^{1}\!/_{5}$ and there is a conflict when *Sender* is healthy when $R < 1$. The latter holds whenever $S = V$. Just as the stability of honest signaling, conflicts are not influenced by the probability p that *Sender* is healthy.

According to the handicap principle, if interest conflict, honest signaling can only be stable when signals are costly. Maynard Smith (1991) further shows that this holds in the Philip Sidney game. For the current parameters ($S = V = {}^{4}\!/_{5}$) there is a conflict if $R < 1$. The only case where honest signaling is stable but signals are cost-free is when $R = 1$.

4.3 Individual Learning

This section randomly matches *Sender*s and *Receiver*s to play the Philip Sidney game: the random matching model (Section 2.3.2). The agents use ϵ-greedy Q-learning (Section 2.3.2). This algorithm is ideal for learning in populations with conflicting interests for several reasons:

- It poses no constraints on the agents' rewards, so it can be directly applied to the Philip Sidney game whose payoffs $0 \leq u < 2$. Learning automata would require the rewards to be scaled to the closed range $[0, 1]$.

- Agents stay adaptive. When an action's expected payoff changes and that action is used, the agent will adjust the action's value at the same fixed rate α no matter how long the agent has been learning. A constant exploration rate ϵ allows agents to (re)discover actions whose expected payoff has improved.

- Agents explore forever at a fixed but small exploration rate ϵ. This ensures that each possible information set is reached and forces the opponent(s) to optimize at each information set, not just those along the usual path of play. This implements sequential rationality (Appendix B).

- It is possible to increase the amount of exploration at the beginning of the learning process by optimistically initializing the Q-values (Sutton and Barto, 1998, p 40). In these experiments, I set the initial Q-value $Q_0 = 2$ which is the highest possible payoff in any Philip Sidney game.

Staying adaptive and exploring forever is a necessary requirement in competitive environments as they tend to be non-stationary. Whereas in a stationary environment exploration can be ignored once enough information has been collected, in a non-stationary environment the agent has to continue exploring in order to track changes in the environment.

4.3.1 Experiments and results

I used the following parameters for all experiments reported in this section: the learning rate $\alpha = {}^{1}/_{10}$, the exploration rate $\epsilon = {}^{1}/_{100}$, the initial action value $Q_0 = 2$, and the population size $N = 100$. At each of the 10^6 iterations, I randomly selected two agents to play the game (on average, each agent was selected 10^4 times as *Sender* and 10^4 times as *Receiver*). After these 10^6 iterations, I recorded the outcome of 100 games and computed the frequency of each possible outcome. Each outcome is a type $t \in \{\texttt{healthy}, \texttt{needy}\}$, a signal $m \in \{\texttt{signal}, \texttt{quiet}\}$, and a response $r \in \{\texttt{donate}, \texttt{keep}\}$. Finally, these results were averaged over 100 simulations per experiment. Figure 4.4 shows an example of an experiment where the signal cost $c = {}^{1}/_{10}$, the degree of common interest $R = {}^{1}/_{4}$, the players' survival probability without the resource $S = V = {}^{4}/_{5}$, and the probability that *Sender* is healthy $p = {}^{1}/_{2}$. It shows the evolution of the frequencies of each of the outcomes over time.

Figure 4.5 shows for which combinations of signal cost c and degree of common interest R honest signaling evolves. Honest signaling is the strategy pair $(\mathtt{S_N}, \mathtt{D_S})$, so the frequency of honest signaling is the sum of the frequencies of the outcomes $(\mathtt{healthy}, \mathtt{quiet}, \mathtt{keep})$ and $(\mathtt{needy}, \mathtt{signal}, \mathtt{donate})$. In particular, dark red in the figure indicates honest signaling was always observed, dark blue indicates that honest signaling was never observed.

Three things are remarkable:

- First, it is particularly surprising that in a large part of the area where honest signaling is an equilibrium (the area enclosed by the solid black line), honest signaling does not evolve! It only does so near the lower tip of that region at $(c, R) = (0.153, 0.236)$.
- Second, there is a region where honest signaling evolves but is not an equilibrium (red/orange/yellow region to the left of the lower tip).
- Finally, the green area in the lower part of the figure seems to indicate honest signaling is observed 50% of the time, though this is just an artifact.

I now explain each of these observations.

When honest signaling is stable

In the region where honest signaling is stable, it is observed only near the lower tip (Figure 4.5). No signaling was observed when the common interest R was too high ($R > {}^{1}/_{2}$). Honest signaling not necessarily emerges when it is an equilibrium since an other equilibrium may dominate it. For example the strategies $(\mathtt{S_\emptyset}, \lambda\mathtt{D_A} + (1-\lambda)\mathtt{D_Q})$ for all $\lambda \in [0,1]$. In this example (Philip Sidney games with parameters $S = V = {}^{4}/_{5}$ and $p = {}^{1}/_{2}$), this set of weak Nash equilibria Pareto dominates honest signaling $(\mathtt{S_N}, \mathtt{D_S})$ when $R \geq \frac{1}{1+5c}$.

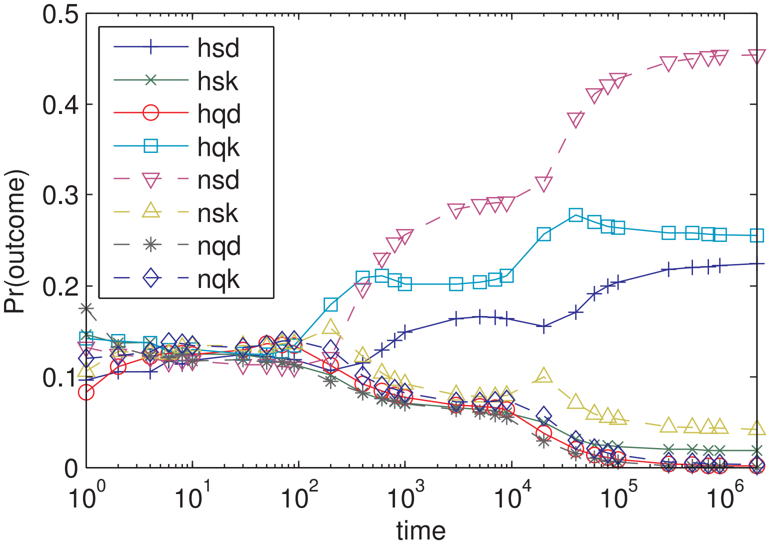


Figure 4.4: The evolution of the outcomes in the Philip Sidney game with the signal cost $c = {}^1\!/_{10}$, the degree of common interest $R = {}^1\!/_4$, the players' survival probability without the resource $S = V = {}^4\!/_5$, and the probability that *Sender* is healthy $p = {}^1\!/_2$. Outcomes are abbreviated by the first letter of each component, so 'nsd' denotes ($\texttt{needy}, \texttt{signal}, \texttt{donate}$), 'hqk' denotes ($\texttt{healthy}, \texttt{quiet}, \texttt{keep}$), etc. Each outcome occurred at most half of the time, since the type $t \in \{\texttt{healthy}, \texttt{needy}\}$ is included in the outcome and the type distribution was fixed at $(p, 1-p) = ({}^1\!/_2, {}^1\!/_2)$. At the end of this experiment, when *Sender* was needy he almost always signaled and *Receiver* mostly donated ('nsd'). When *Sender* was healthy, about half of the time he remained quiet and *Receiver* kept the resource ('hqk'). The other half of the time *Sender* lied and got the resource ('hsd').

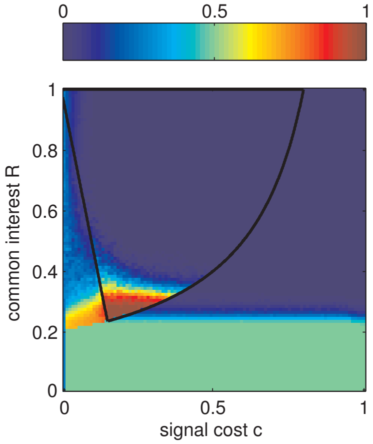


Figure 4.5: Frequency of honest signaling (by summing the frequencies of (`healthy`, `quiet`, `keep`) and (`needy`, `signal`, `donate`)) for all combinations of signal cost c and degree of common interest R. The probability that *Sender* is healthy $p = {}^1\!/_2$ and the players' survival probability without the resource $S = V = {}^4\!/_5$.

This was also observed in the experiments. More than 90% of the time *Sender* did not signal and *Receiver* donated in that region (Figure 4.6).

Figure 4.7 shows for which Philip Sidney games honest signaling is Pareto dominated by an other Nash equilibrium. In the region where honest signaling is the unique Pareto optimal Nash equilibrium, it always emerged. In the region where honest signaling is a Pareto optimal Nash equilibrium that is not unique, honest signaling emerged although not exclusively. In the region where honest signaling is a Pareto dominated Nash equilibrium, it never emerged. Whether or not honest signaling is a Nash equilibrium does not depend on the probability p that *Sender* is healthy (Section 4.2.1 and Figure 4.3a), but Pareto optimality of the equilibria does depend on probability p. Still, the same effects were found for other values of probability p.

Clearly, other equilibria may be more important and must be taken into account in order to predict the outcome of a game. By considering honest

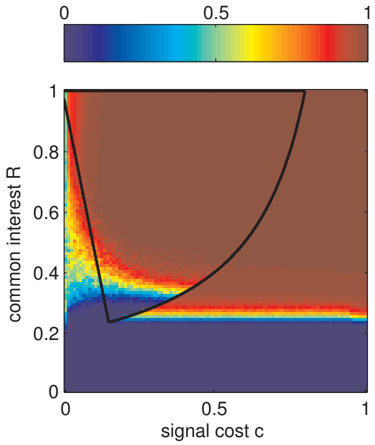


Figure 4.6: Sum of the frequency of outcomes ($\texttt{healthy}, \texttt{quiet}, \texttt{donate}$) and ($\texttt{needy}, \texttt{quiet}, \texttt{donate}$) for all combinations of signal cost c and degree of common interest R. The probability that *Sender* is healthy $p = {}^1\!/\!_2$ and the players' survival probability without the resource $S = V = {}^4\!/\!_5$.

signaling alone and determining whether it is an evolutionarily stable strategy (ESS) or not, one overlooks other equilibria. Another problem is that strategies like $(\texttt{S}_\emptyset, x\texttt{D}_\texttt{A} + (1-x)\texttt{D}_\texttt{Q})$ are not even ESS, because they are weak Nash equilibria in an asymmetric two-player game (Section 2.3.1). Even after considering all ESSs, one would still conclude that honest signaling is the only equilibrium and will necessarily emerge.

An alternative is to find all Nash equilibria and determine their importance by means of Pareto optimality. Unfortunately, it is possible that multiple Pareto optimal Nash equilibria exist in which case multiple equilibria may emerge.

Partial communication

There are cases where signaling took place but honest signaling is not an equilibrium. The region with the signal cost $c < (1-R)/5$ and the degree of common interest R just above $1/(5-5c)$ (Figure 4.5) showed the following

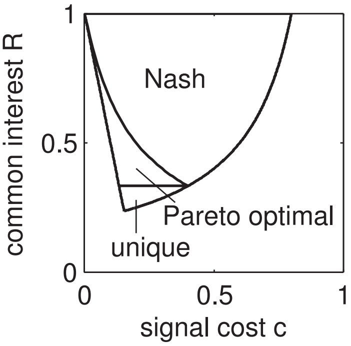


Figure 4.7: For which combinations of signal cost c and common interest R honest signaling is a Nash equilibrium. The probability that *Sender* is healthy $p = {}^{1}\!/_{2}$ and the players' survival probability without the resource $S = V = {}^{4}\!/_{5}$. In the biggest part of that area honest signaling is a Nash equilibrium Pareto dominated by another equilibrium ('Nash'). In the region below it, it is Pareto optimal (not Pareto dominated), but there is an other Nash equilibrium that is also Pareto optimal ('Pareto optimal'). Finally in the lower tip, honest signaling is a unique Pareto optimal Nash equilibrium ('unique').

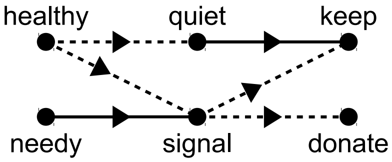
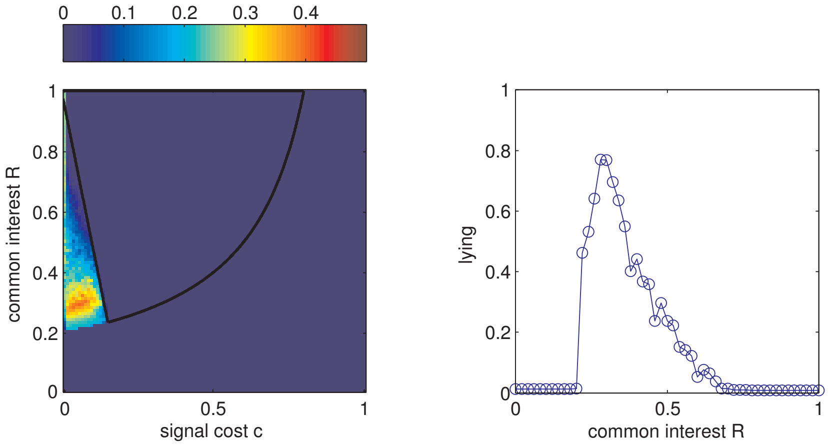


Figure 4.8: Mixed strategies representing partial signaling. *Sender* always signals when needy. *Receiver* always keeps the resource when *Sender* is quiet. When healthy, *Sender* is sometimes quiet (honest) and sometimes signals (dishonest). When *Sender* signals *Receiver* sometimes donates (honest) and sometimes keeps the resource (dishonest).



(a) For varying signal cost c. (b) For signal cost $c = 0.05$.

Figure 4.9: Frequency of lying where the probability that *Sender* is healthy $p = {}^1\!/_2$ and the players' survival probability without the resource $S = V = {}^4\!/_5$.

behavior (Figure 4.8). When needy, *Sender* always signaled; when healthy, he was sometimes quiet and sometimes signaled—he lied. When *Sender* was quiet, *Receiver* always kept the resource; when *Sender* signaled, he sometimes donated the resource and sometimes kept it. The experiment shown in Figure 4.4 is an example.

The behavior observed in that experiment was $(0.48\,\mathtt{S_A} + 0.52\,\mathtt{S_N}, 0.92\,\mathtt{D_S} + 0.8\,\mathtt{D_\emptyset})$ which is close to one of the Nash equilibria of that game ($p = {}^1\!/_2$, $S = V = {}^4\!/_5$, $c = {}^1\!/_{10}$, and $R = {}^1\!/_4$): $(0.17\,\mathtt{S_A} + 0.83\,\mathtt{S_N}, 0.62\,\mathtt{D_S} + 0.38\,\mathtt{D_\emptyset})$. These strategy profiles may not seem close, but they are close in terms of payoffs: *Sender* cannot improve his payoff by switching from the observed behavior to the equilibrium behavior and *Receiver* can improve his payoff merely 0.7% when switching from the observed behavior to the equilibrium behavior.

Figure 4.9 shows the frequency at which *Sender* lied. It peaked for a degree of common interest about $R = {}^3\!/_{10}$. For weaker conflicts ($R > {}^3\!/_{10}$) lying was less frequent until, at some point, signaling does not pay

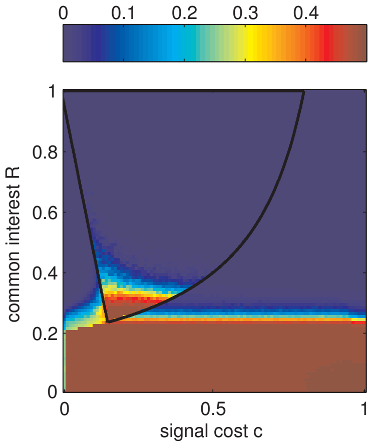


Figure 4.10: Frequency of outcome (`healthy`, `quiet`, `keep`) for all combinations of signal cost c and degree of common interest R. The probability that *Sender* is healthy $p = {}^1\!/_2$ and the players' survival probability without the resource $S = V = {}^4\!/_5$.

off at all because *Receiver* prefers to donate no matter *Sender*'s type. For stronger conflicts ($R < {}^3\!/_{10}$) lying was less frequent while honest signaling increased until signaling no longer pays off because *Receiver* prefers to keep the resource no matter *Sender*'s type ($R < 1/(5 - 5c)$). Intensifying the conflict (decreasing R from ${}^3\!/_{10}$ to ${}^1\!/_5$) decreased lying and at the same time increased honest signaling. Lying almost always payed off, which means the players learned to lie at the optimal rate, above which *Receiver* would no longer trust the signal.

While ESS analysis predicts that a minimal signal cost is required for honest signaling ($c > 1 - 5R$), these experiments predict that signaling is still possible. Although lying emerges, *Receiver*s still benefit by trusting the signal most of the time.

Artifact

The seemingly 50% of honest signaling observed in the Philip Sidney games in the lower part of Figure 4.5 (common interest $R < 1/4$) is an artifact and is entirely due to the outcome ($\texttt{healthy}, \texttt{quiet}, \texttt{keep}$) (Figure 4.10). In that region, it does not pay off *Receiver* to donate no matter *Sender*'s type, so it does not pay off *Sender* to signal. The observed strategy in that region is ($\texttt{S}_\emptyset, \texttt{D}_\emptyset$) (never signal, never donate) which partially overlaps with honest signaling ($\texttt{S}_\texttt{N}, \texttt{D}_\texttt{S}$) but is in fact a pooling equilibrium: *Sender* always uses the same signal, so there is no communication.

Meaning reversed

The Philip Sidney game has two separating equilibria: ($\texttt{S}_\texttt{N}, \texttt{D}_\texttt{S}$) and ($\texttt{S}_\texttt{H}, \texttt{D}_\texttt{Q}$). These are equilibria where *Sender* uses a different signal when he is healthy and when he is needy, so that *Receiver* can perfectly infer *Sender*'s type. Intuitively, ($\texttt{S}_\texttt{H}, \texttt{D}_\texttt{Q}$) (signal when healthy, donate when quiet) seems weird, but is also honest signaling. It simply has the meaning of the signals reversed: `quiet` means you are in need, `signal` means you are healthy. When signals are cost-free, as was the case in Chapter 3, both separating equilibria are equivalent and the meaning of the signal is purely conventional. In the experiments both equilibria were equally likely to emerge. When `signal` costs more than `quiet`, the asymmetry creates an obvious, or natural, meaning and an obscure one. The second equilibrium was not observed in the experiments and the conditions under which it is an equilibrium rarely hold. For example, it cannot be an equilibrium when *Sender*'s survival probability when healthy but without the resource is less than *Receiver*'s survival probability without the resource ($V \leq S$). Never signal ($\texttt{S}_\emptyset$) is then a better response to donate when quiet ($\texttt{D}_\texttt{Q}$) than signal when healthy ($\texttt{S}_\texttt{H}$).

4.3.2 Summary

Randomly matching *Sender*s and *Receiver*s that learn individually predicts honest signaling differently than ESS in two ways:

- When the degree of common interest is too high, signaling is no longer beneficial since *Receiver* donates his resource no matter *Sender*'s type.
- When the signal cost is below the minimum needed for honest signaling to be an equilibrium, signaling does not abruptly break down, but some lying emerges.

The model discussed here is based on individual learning. In the next section, the model is based on social learning and evolution.

4.4 Evolution in Finite Populations

This section analyzes evolutionary stability and dynamics in finite populations (Section 2.3.1) for the Philip Sidney game. Section 4.4.1 and 4.4.2 study the effects of the selection pressure and the population size, respectively, on the emergence of honest signaling. Section 4.4.3 analyzes the dynamics and compares its long term outcome to predictions of evolutionary stability.

4.4.1 Effect of selection pressure

Recall that a strategy `W` is an evolutionarily stable in finite populations ($\mathrm{ESS_N}$) (Definition 2.7) if natural selection (or social learning)

- opposes any mutant invading the population of `W`'s, and
- opposes any mutant strategy replacing an entire population of `W`'s.

Strategy `W` resists invasion if the fitness of the single mutant `M` is lower than the fitness of the wild type `W` in a population of $N-1$ `W`'s and one mutant `M`: $f_{\mathtt{M}}(1) < f_{\mathtt{W}}(1)$ for all $\mathtt{M} \neq \mathtt{W}$ (Equation (2.9)). While the first condition is independent of the selection pressure β, the second one is not. Strategy `W` is resistant to replacement if the fixation probability of any mutant strategy `M` is less than the fixation probability under random drift: $\rho_{\mathtt{W}\to\mathtt{M}} < {}^1\!/\!_N$. For low selection pressure $\beta \ll 1$, the condition simplifies to Equation (2.10), but in general each case must be numerically verified. Figure 4.11 shows

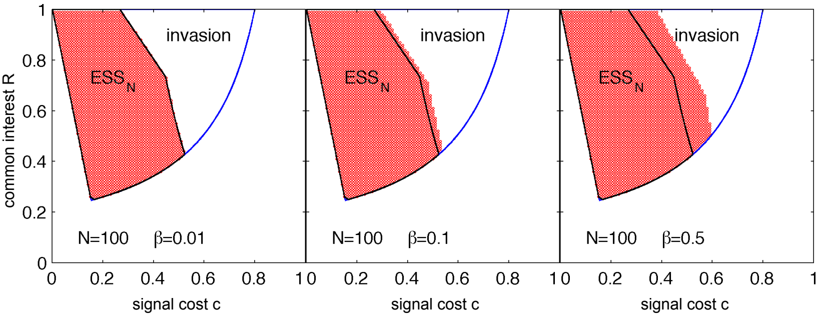


Figure 4.11: The effect of selection pressure β on the evolutionary stability of honest signaling ($\mathrm{ESS_N}$) in the Philip Sidney game when the probability that *Sender* is healthy $p = 1/2$ and the players' survival probabilities without the resource $S = V = 4/5$. The population size $N = 100$. For increasing selection pressure ($\beta = 0.01$, 0.1 and 0.5) honest signaling resists replacement and is evolutionarily stable (red shaded area) for larger signal costs c. There is a large set of cases where honest signaling is resistant to invasion (area labeled 'invasion'), but not to replacement by any mutant. If, by chance, a few mutants manage to survive in a population of honest signalers, they have good chances to take over the population.

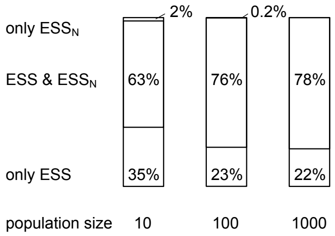


Figure 4.12: The Philip Sidney games where honest signaling is evolutionary stable in infinite populations (ESS) and/or in finite populations (ESS$_{\mathrm{N}}$). These games represent about 7% of 10^6 games randomly selected from the entire space of Philip Sidney games. In 22% to 35% of these games, honest signaling is only evolutionarily stable under the assumption of infinite populations (ESS). For very small populations, there are some games where honest signaling is ESS$_{\mathrm{N}}$ but not ESS.

the result for the Philip Sidney games where the probability that *Sender* is healthy $p = 1/2$ and the players' survival probabilities without the resource $S = V = 4/5$. The higher the selection pressure, the more cases where honest signaling resists replacement by any mutant. Since resistance to replacement is the most restrictive condition for ESS$_{\mathrm{N}}$ (Figure 4.11), higher selection pressure favors the evolutionary stability of honest signaling. Since ESS$_{\mathrm{N}}$ is a subset of ESS for sufficiently large populations (Nowak et al., 2004), ESS$_{\mathrm{N}}$ approaches ESS for increasing selection pressure β.

4.4.2 Effect of population size

Similarly to the methods described in (Gokhale and Traulsen, 2010; Han et al., 2012), I randomly sampled 10^6 parameter configurations from the game's entire parameter space (see Table 4.1 for the game's parameters) and found that honest signaling $(\mathtt{S_N}, \mathtt{D_S})$ is evolutionarily stable in slightly less than 7% of the games.

In finite populations, however, in an important part (22% to 35%) of the games where honest signaling is ESS it is not ESS$_{\mathrm{N}}$ (evolutionarily stable in finite populations), see Figure 4.12. For small populations there are, on the one hand, less cases where honest signaling is ESS$_{\mathrm{N}}$ than ESS, and on the

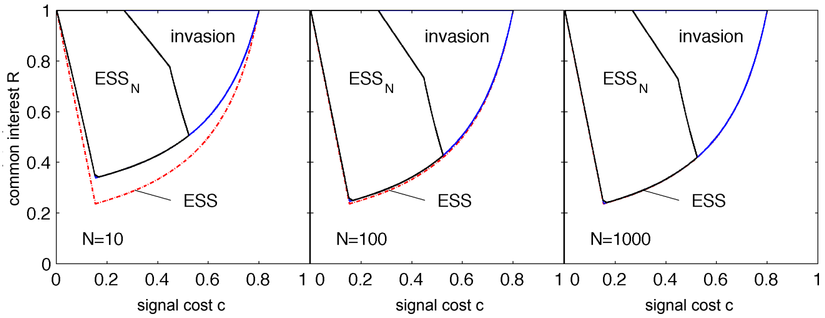


Figure 4.13: The effect of population size N on the evolutionary stability of honest signaling ($\mathrm{ESS_N}$) in the Philip Sidney game when the probability that *Sender* is healthy $p = 1/2$ and the players' survival probabilities without the resource $S = V = 4/5$. As the population size increases ($N = 10, 100$, and 1000) there are more cases where honest signaling is $\mathrm{ESS_N}$ (area surrounded by a black, solid line and labeled '$\mathrm{ESS_N}$'). But they remain a subset of the cases where it is ESS (area surrounded by a red, dashed line and labeled 'ESS'). There is a large set of cases where honest signaling is resistant to invasion (area surrounded by a blue, solid line and labeled 'invasion'), but not to replacement by any mutant.

other hand, a marginal fraction of cases appear where honest signaling is $\mathrm{ESS_N}$ but not ESS. For reasonable population sizes ($N \geq 100$), the number of such cases is negligible. The net effect is that, for finite populations, there are a lot less cases where honest signaling is evolutionarily stable than what the infinite population model suggests.

For a slice of the space of Philip Sidney games ($p = 1/2$, $S = V = 4/5$) Figure 4.13 shows the regions where honest signaling is ESS and $\mathrm{ESS_N}$ for population sizes $N = 10, 100$, and 1000. These results are in accordance with the theory (Section 2.3.1). For large populations, every strategy that is $\mathrm{ESS_N}$ is also ESS. For small populations, the ESS conditions are neither necessary nor sufficient for $\mathrm{ESS_N}$. These experiments show to what extent this holds for honest signaling in the Philip Sidney game.

These results have two implications. First of all, they suggest that honest signaling is not as widespread as was previously suggested by the ESS concept. Second, given that there is still an important overlap between

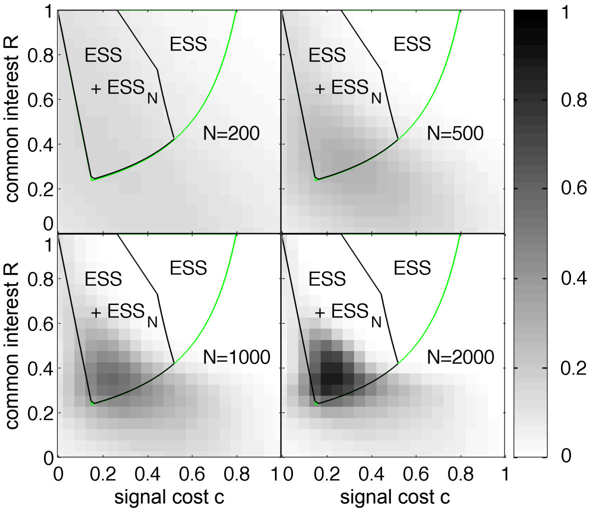


Figure 4.14: Frequency of honest signaling (white means the entire population never adopts honest signaling, while black means the entire population always uses honest signaling) in the Philip Sidney game when the probability that *Sender* is healthy $p = 1/2$ and the players' survival probabilities $S = V = 4/5$. The selection pressure $\beta = 1/10$ and the population size varies ($N = 200, 500, 1000$, and 2000).

games where honest signaling is evolutionarily stable in finite populations ($\mathrm{ESS_N}$) and where it is evolutionarily stable in infinite populations (ESS), there is increased evidence that there are indeed settings where honest signaling is a viable strategy. Figure 4.13 shows that especially in games with higher signal cost c honest signaling is no longer an $\mathrm{ESS_N}$ and that the maximum signal cost c decreases with increased common interest R. For smaller populations, the minimum signal cost c and the minimum degree of common interest R at which honest signaling is stable increase slightly.

4.4.3 Stationary distribution vs. evolutionary stability

Finally, I show that there is an important difference between the stability of honest signaling and whether or not it is the most prevalent strategy in the population. Using the methods described in Section 2.3.1, I compute the time the entire population adopts each of the sixteen strategies described above (Section 4.2) and analyze whether and when honest signaling is the most frequent strategy. Two interesting and important observations follow.

First, with increasing population size N, the fitness of each strategy becomes more important. The population size N has an effect similar to the selection pressure β. For smaller populations, the frequency of the strategies approaches that of neutral selection: the fitness of a strategy is irrelevant and all strategies occur with equal frequency, which is $1/16$. Figure 4.14 illustrates this as the frequency of honest signaling for most games is closer to $1/16$ when the population size $N = 200$ than when $N = 2000$.

Second, there is a considerable difference between the cases where honest signaling is evolutionarily stable and where it is the most prevalent strategy. On the one hand, there are cases where honest signaling is evolutionarily stable but not viable (it has frequency 0 in the stationary distribution). In Figure 4.14, for population size $N = 2000$, the games in a wide area around signal cost $c = 1/2$ and common interest $R = 7/10$ are such cases. Interestingly, almost all of the games in which honest signaling is ESS but not $\mathrm{ESS_N}$ are such cases. On the other hand, even though honest signaling is neither $\mathrm{ESS_N}$ nor ESS, it may still be the most frequent strategy. This is illustrated by the gray areas in Figure 4.14 outside the 'ESS' region. An extreme example is the game with $c = 1/4$ and $R = 1/4$ ($N = 2000$) where honest signaling is the most frequent strategy. (It has frequency above 50%, but note that a strategy can be the most frequent strategy even though its frequency is less than 50% since there are more than two strategies.)

4.4.4 Summary

The evolutionary stability of honest signaling in finite populations ($\mathrm{ESS_N}$) differs from evolutionary stability in infinite populations (ESS) as follows:

- $\mathrm{ESS_N}$ does not predict honest signaling at high signal costs where ESS does.

- For smaller populations, the minimum signal cost c and minimum degree of common interest R for honest signaling to be $\mathrm{ESS_N}$ increases slightly. For larger populations, the minima are the same as those for ESS.

- For higher selection pressure, honest signaling becomes resistant to replacement at higher signal costs and $\mathrm{ESS_N}$ approaches ESS since at reasonably large populations $\mathrm{ESS_N}$ is a subset of ESS.

The evolutionary dynamics yield the same qualitative results as those of the individual learning (Section 4.3). Honest signaling has high frequencies in both models in similar regions. Again, cases occurred where honest signaling is stable but does not evolve, or where honest signaling is not stable but still evolves and even is the most frequent strategy.

The predictions of these dynamics resemble those of $\mathrm{ESS_N}$ more than those of ESS. The dynamics and $\mathrm{ESS_N}$ predict honest signaling is not a viable strategy when the signal cost is high whereas ESS does.

4.5 Related Work

Bergstrom and Lachmann (1997) show that honest signaling can be Pareto dominated by pooling equilibria in both the discrete and continuous Philip Sidney game because the minimal signal cost needed at equilibrium is too high.

Lachmann and Bergstrom (1998) suggested that dynamic analyses are crucial for predicting the emergence and hence the existence of honest signaling. As far as I know, the first dynamic analysis of the Philip Sidney game is the work of Huttegger and Zollman (2010). They apply the replicator dynamics—evolutionary dynamics assuming well-mixed infinite populations and no mutations—and contrast their results to those obtained by calculating the evolutionarily stable strategies, which is a static equilibrium analysis. They find that in many cases honest signaling has far smaller basins of attraction than other equilibria and is thus less likely to emerge.

Whereas 100% reliable signaling is too costly, partially reliable signaling turns out to be much less costly. Examples are the so-called ‘partial pooling equilibria' where the same signal is used by similar types (Lachmann and Bergstrom, 1998) and mixed equilibria where signals may sometimes be truthful and sometimes be deceptive (Huttegger and Zollman, 2010; Zollman et al., 2013). In the replicator dynamics, the basins of attraction of these mixed equilibria are equally large as those of honest signaling and are thus equally likely to emerge.

4.6 Conclusion

This chapter examined whether honest signaling can emerge if interests conflict but signals are costly. Traditionally, biologists tested when honest signaling is an evolutionarily stable strategy in the Philip Sidney game. Since, stability analyses do not reveal whether an equilibrium can emerge, I compared their results with dynamic models: individual learning in the random matching model in Section 4.3 and evolutionary stability and dynamics in finite populations in Section 4.4.

I can draw three main conclusions regarding honest signaling.

1. When honest signaling is stable, it will not necessarily evolve either because the signals cost too much or the common interest is too high. Both the individual learning in Section 4.3 as the evolutionary dynamics in Section 4.4 showed this. Individual learning also revealed that when the signal cost or degree of common interest is too high, honest signaling is Pareto dominated by an other equilibrium. The evolutionary dynamics showed similar results. The evolutionary stability in finite populations (ESS_N) differs from the one in infinite populations (ESS) in that it does not predict honest signaling to be stable at high signal costs. Although for large populations and high selection pressure, ESS_N approaches ESS. Whereas Chapter 3 showed how common interest was beneficial to the emergence of honest signaling, this chapter revealed how common interest may render it useless. When the degree of common interest is high, not signaling Pareto dominates honest signaling because, no matter what, *Receiver* prefers to donate his resource to *Sender*.

2. When honest signaling is not stable, there may still be signaling. In Section 4.3 partially reliable signaling emerged in the form of a mixed equilibrium where *Sender* is sometimes honest and sometimes dishonest. Similarly to honest signaling this equilibrium does not emerge for high common interest, but contrary to honest signaling this equilibrium emerges for very cheap signals.

3. Even unstable strategies may be the most likely to emerge. Section 4.4 showed an example.

I can also draw some conclusions regarding dynamic versus static analyses:

- When a strategy is an ESS, it may still be evolutionarily less likely to emerge. For example, because it is Pareto dominated by another equilibrium. You should thus consider all ESSs, not just the strategy you are interested in.

- Some equilibria may be Pareto optimal even when they are not ESS. These equilibria may be evolutionarily more likely to emerge than the ESS it dominates. You should thus consider all Nash equilibria and verify whether or not it dominates the strategy of interest.

- A strategy may not be an equilibrium at all, but it can still be evolutionarily important. Only dynamic analyses can reveal such phenomena.

Chapter 5

Costly, Social Punishment

The previous chapter studied how costly signals allow honest signaling to emerge when interests conflict. This chapter, based on the publication (Catteeuw et al., 2014a), studies an alternate explanation: punishment. Can punishment of dishonesty allow the emergence of honest signaling if interests conflict even when punishment is costly for the punisher?

Contents

5.1 Introduction

In the previous chapter, an honest signal was costly. It is also possible, to make dishonest signals costly and honest ones cost-free. For example, Lachmann et al. (2001) study cost functions such as "it is free to signal below quality but lethal to signal above quality," where 'quality' refers to *Sender*'s type. For example, the peacock's quality or type is his quality as a peahen's mating partner (Example 1.3). Such cost functions have the advantage that the cost is paid for dishonest, out-of-equilibrium behavior and honest signaling is a cost-free equilibrium. They could arise due to punishment.

People and animals are willing to pay a cost in order to punish those that infringed their interests (Clutton-Brock and Parker, 1995; Fehr and Gächter, 2002; Guala, 2012). Here are three examples of punishment among animals in signaling contexts:

- Some species, like house sparrows (Moller, 1987) and paper wasps (Tibbetts and Izzo, 2010), punish liars. They wear colored patches, called 'badges,' that indicate fighting ability and are used to resolve small conflicts without a costly fight, because weaker animals can avoid stronger ones. Only conflicts between equally ranked individuals often escalate. Moller (1987) and Tibbetts and Izzo (2010) experimentally changed badges and/or fighting ability and discovered that if a conflict does escalate and one individual lied or bluffed by exaggerating his fighting ability, his opponent reacted extra aggressively. A liar willing to retreat when his opponent charges still gets attacked and suffers a punishment. (See a more extensive discussion on signals in animal contests and punishment in Section 5.4).

- Another example is seen in Rhesus macaques where individuals are punished for the lack of signaling. Animals that do not send food calls (they signal `quiet` instead of `food`) and are caught with food are often punished (Hauser and Marler, 1993).

- The signal's receiver can also be punished. In many social primates, females that do not responds to a male's attempt to mate (a signal) are often attacked and such punishment is effective (Nadler and Miller, 1982; Smuts and Smuts, 1993).

Punishment promotes cooperation, for example in the prisoner's dilemma and the public goods game (Hauert et al., 2007; Hilbe and Traulsen, 2012; Sigmund et al., 2010), but the effect of punishment in signaling contexts is merely studied implicitly by assuming cost functions like the one mentioned above (for example (Lachmann et al., 2001)).

In Section 5.2, I define a new game that is based on the Philip Sidney game and explicitly includes punishment of the four types of dishonest behavior that can occur in signaling contexts (Table 5.1): lying (signal when healthy), timid (quiet when needy), greedy (keep when signal), and worried behavior (donate when quiet).

I applied evolutionary dynamics in finite populations (Section 2.3.1) to study the effects of punishment on the evolution of honest signaling. In Section 5.3, I present the results and compare them with those from Section 4.4.

Contributions

- I define a new game based on the Philip Sidney game to explicitly study the effect of punishment on the emergence of honest signaling (in both *Sender*s and *Receiver*s).
- Evolutionary dynamics in finite populations shows that: punishing liars increases the emergence of honest signaling for cheap and cost-free signals; punishing greedy individuals increases the emergence of honest signaling for low common interest and costly signals; but punishing worried and timid individuals does not.

5.2 Philip Sidney Game with Explicit Punishment

The model is based on the symmetric Philip Sidney game (Section 4.2) which has sixteen strategies (Table 4.2) but adds new honest signaling

Table 5.1: There are four ways to deviate from honest signaling. This table lists the name of such behavior, by which player it is performed, to what behavior and strategies it corresponds in the Philip Sidney game, and by which strategy it is punished.

	name	behavior and strategies	punishing strategy
Sender	lying	signal when healthy: $\mathtt{S_H}$, $\mathtt{S_A}$	$\mathtt{P_L}$
	timid	quiet when needy: $\mathtt{S_H}$, $\mathtt{S_\emptyset}$	$\mathtt{P_T}$
Receiver	greedy	keep when signal: $\mathtt{D_Q}$, $\mathtt{D_\emptyset}$	$\mathtt{P_G}$
	worried	donate when quiet: $\mathtt{D_Q}$, $\mathtt{D_A}$	$\mathtt{P_W}$

strategies that punish dishonest behavior. I distinguish four different strategies (Table 5.1) that target each of the possible deviations from honest signaling: punishment of lying $\mathtt{P_L}$, of greedy $\mathtt{P_G}$, of timid $\mathtt{P_T}$, and of worried opponents $\mathtt{P_W}$. An individual is lying (or simply a liar) if he signals when healthy; greedy if he keeps when *Sender* signals; timid if he remains quiet when needy; and worried if he donates when *Sender* is quiet.

For simplicity, only honest signalers can punish but some agents can be punished on several occasions. For example, the strategy $(\mathtt{S_H}, \mathtt{D_Q})$ is lying, greedy, timid, and worried simultaneously, albeit on different occasions. When adopting this strategy, a healthy *Sender* is deemed a liar, while a needy *Sender* is deemed a timid.

In the new game, a player's survival probability is decreased by c' if he punishes his opponent and decreased by c'' if he is punished. As before, a player's payoff includes a fraction R of his opponent's survival probability. Punishing your opponent always lowers your own payoff even if the cost to punish $c' = 0$ due to the degree of common interest with your opponent ($R > 0$) and decreasing his survival probability decreases your payoff. For example, if *Sender* punishes *Receiver*, *Sender* earns $u_S = (v_S - c') + R(v_R - c'')$ and *Receiver* $u_R = (v_R - c'') + R(v_S - c')$, where *Sender*'s and *Receiver*'s survival probability, v_S and v_R, depend on the game's outcome as shown in Table 4.1. If *Receiver* punishes *Sender*, the costs c' and c'' are swapped. If no one punishes, the costs c' and c'' are left out (as in the original Philip Sidney game).

Table 5.2: Parameters of the Philip Sidney game with punishment. The new parameters are p', c', and c''. The others are the same as in the Philip Sidney game (Section 4.2 and Table 4.1).

new	parameter	meaning
	$0 < p < 1$	probability that *Sender* is healthy
✓	$0 < p' < 1$	probability that *Sender*'s type is revealed
	$0 \leq R \leq 1$	degree of common interest
	$0 < S < 1$	*Receiver*'s survival probability without the resource
	$0 < V < 1$	*Sender*'s survival probability when healthy without the resource
	$0 \leq c \leq 1$	signal cost
✓	$0 \leq c' \leq 1$	cost payed to punish the opponent
✓	$0 \leq c'' \leq 1$	cost incurred when punished

While greedy and worried behavior is always noticed and hence always punished by strategies such as $\mathtt{P_G}$ and $\mathtt{P_W}$, lying and timid behavior may go unnoticed. To discover lying and timid agents, their type must be revealed. This happens with probability $p' < 1$. With probability $1 - p'$ their type remains private and they cannot be punished. Table 5.2 shows the parameters of this new game including those of the original Philip Sidney game (Section 4.2 and Table 4.1).

To study the effect of punishing dishonest behavior on the evolution of honest signaling, I considered a finite population of the sixteen pure strategies of the symmetric Philip Sidney game together with one of the punishment strategies described above. For different parameter configurations, I numerically computed the stationary distributions of the strategies using the evolutionary dynamics described in Section 2.3.1.

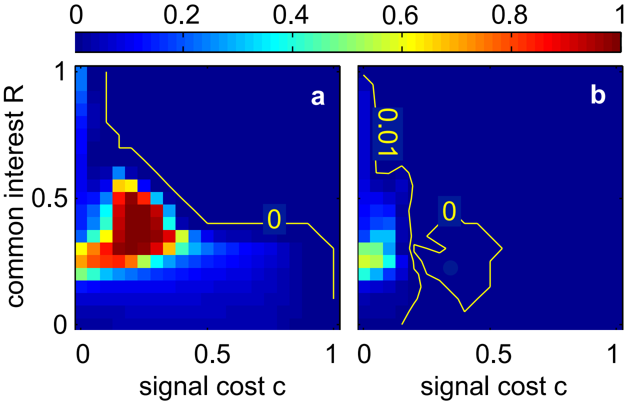


Figure 5.1: *a)* Total frequency of honest signaling in the Philip Sidney game with punishment of the liars ($\mathtt{P_L}$). *b)* Increase in frequency of honest signaling when $\mathtt{P_L}$ is present compared to when it is absent (being replaced with another pure honest signaling strategy). For small signal costs there is a clear increase in the frequency of honest signaling. There is also a region where punishment has a slightly negative effect ($\approx -10^{-3}$ inside the 0-contour line). Both figures were obtained for population size $N = 100$, selection pressure $\beta = 5$, and varying signal cost c and common interest R. The other parameters of the game are $p = p' = {}^1/_2$, $S = V = {}^4/_5$, $c' = {}^1/_2$, and $c'' = 1$. See Table 5.2 for their meaning.

5.3 Experiments and Results

5.3.1 Punishing liars boosts honesty

First, consider $\mathtt{P_L}$: the honest signaling strategy that punishes liars. Figure 5.1a shows the frequency of honest signaling (that is, the sum of the frequencies of $\mathtt{P_L}$ and of the pure honest signaling strategy $(\mathtt{S_N}, \mathtt{D_S})$) as a function of the signal cost c and the common interest R. There is a high level of honest signaling for a wide range of c and R, and even for low values of c. This was not the case in the original Philip Sidney game (Chapter 4). So, punishment of liars can provide an alternative for the handicap principle (Section 4.1): honest signaling can emerge for cheap and cost-free signals if

Sender's type may get revealed and he runs the risk of getting punished.

The effect of punishing liars on the level of honest signaling is clearer in Figure 5.1b. It shows the increase in frequency of honest signaling when $\mathtt{P_L}$ is present in the population compared to the case where $\mathtt{P_L}$ is replaced by another pure honest signaling strategy. Irrespective of the common interest R, when signal cost c is small enough, the presence of $\mathtt{P_L}$ increases the level of honest signaling.

For higher signal costs c, the cost of punishment and honesty may not outweigh the benefits. Figure 5.1b shows a region where punishment has no effect and even slightly decreases the level of honest signaling (approximately -10^{-3}).

The increase in honest signaling can be explained as follows. $\mathtt{P_L}$ is effective for small signal costs c since it is resistant to invasion by any mutant in that region whereas honest signaling without punishment can be invaded by mutants that always signal ($\mathtt{S_A}$) (Figure 5.2). In the original Philip Sidney game, honest signaling can be stable only if there is a non-zero signal cost ($c > 0$). In the new model, honest signaling with punishment cannot be stable, since it can always be replaced by honest signaling without punishment through random drift. But, except for the pure honest signaling strategy, punishment is stable even for cost-free signals ($c = 0$).

Moreover, punishment directly affects liars and indirectly affects other strategies that can replace liars. Figure 5.3 shows the Markov chain and the transition probabilities of the Philip Sidney game with and without $\mathtt{P_L}$ (ignoring the state $\mathtt{P_L}$ and the arrows connected to it). The strategies $(\mathtt{S_A}, \mathtt{D_S})$, and $(\mathtt{S_A}, \mathtt{D_\emptyset})$ can replace honest signaling, but not honest signaling with punishment of liars. The latter can only be replaced through random drift by honest signaling itself. Lying is less frequent when the punisher $\mathtt{P_L}$ is present. For example, the frequency of $(\mathtt{S_A}, \mathtt{D_\emptyset})$ and $(\mathtt{S_H}, \mathtt{D_Q})$ dropped below the average frequency (which is $^1\!/n$ if n is the number of strategies). Since liars are less frequent, other strategies that replace liars are also affected. For example, the frequency of $(\mathtt{S_\emptyset}, \mathtt{D_\emptyset})$ also dropped below the averaged frequency.

Here, I call $(\mathtt{S_H}, \mathtt{D_Q})$ a liar, but it is also honest in some sense. As mentioned at the end of Section 4.3.1, this strategy simply has the meaning of

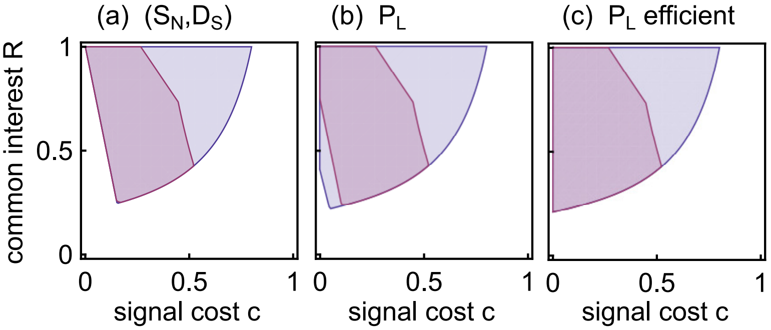


Figure 5.2: Evolutionary stability in finite populations ($\mathrm{ESS_N}$, Definition 2.7) in the Philip Sidney game with punishment for varying signal cost c and common interest R. The purple shaded areas indicate $\mathrm{ESS_N}$ while the blue shaded areas indicate resistance to invasion. The figures were obtained for population size $N = 100$ and selection pressure $\beta \to 0$. The probability that *Sender* is healthy $p = 1/2$ and the players' survival probabilities without the resource $S = V = 4/5$. (a) Honest signaling $(\mathtt{S_N}, \mathtt{D_S})$ can be $\mathrm{ESS_N}$ only if the signal cost $c \geq (1 - R)/5$. (b) Punishment of liars $\mathtt{P_L}$ is $\mathrm{ESS_N}$ for smaller signal costs. Here, the probability that Sender's type is revealed $p' = 1/10$, the cost to punish $c' = 1/2$, and the cost of being punished $c'' = 1$. (c) The $\mathrm{ESS_N}$ region extends more towards the point $(c, r) = (0, 1/5)$, if punishment is more efficient. Here, $p' = 1/2$ instead of $1/10$. Honest signaling and punishment of liars cannot be evolutionarily stable against each other because they behave identical when they interact with each other.

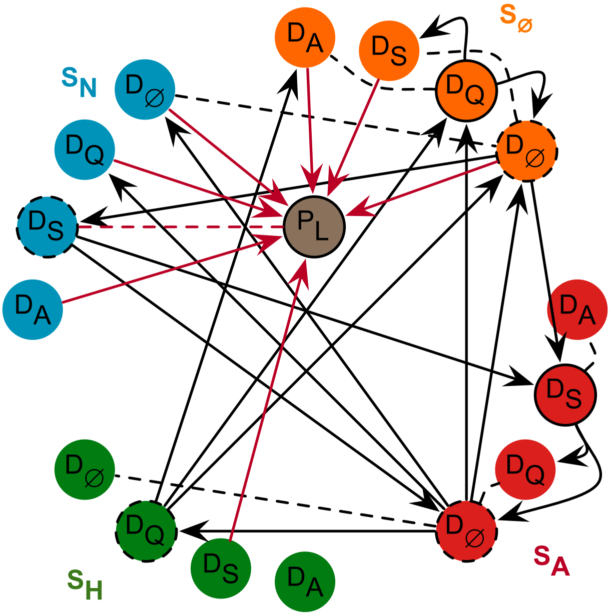


Figure 5.3: Markov chain of transitions between monomorphic populations (Section 2.3.1) for all strategies (Table 4.2) of the Philip Sidney game with and without punishment of liars. Punishment of liars $\mathtt{P_L}$ can only be replaced by honest signaling ($\mathtt{S_N}, \mathtt{D_S}$) through random drift. Black borders indicate strategies which have above average frequency ($> 1/n$). Dashed black borders indicate strategies drops below average after the introduction of the punishment strategy $\mathtt{P_L}$. Dashed lines indicate random drift. For clarity, I only drew transitions departing from strategies with above average frequency and transitions to the punishment strategy $\mathtt{P_L}$. The figure was obtained for signal cost $c = 1/100$ and the common interest $R = 3/10$. The other parameters are the same as in Figure 5.1.

the signals `signal` and `quiet` reversed. When the signal cost c is very small (for example $c = {}^1/_{100}$ as in Figure 5.3), this strategy is only slightly worse than honest signaling. When the signal cost $c = 0$, both strategies behave symmetrically and perform equally well. When signals have different costs, individuals have a common preference of when to signal what and end up using the same signal under the same circumstances: the cost creates an obvious, or natural, meaning and an obscure one. When signals are cost-free there is no a priori common preference for one signal or the other. This does not only make lying cheap, but also creates two equally valid honest signaling strategies: one where `signal` means `needy` and `quiet` means `healthy`, and the other where `signal` means `healthy` and `quiet` means `needy`. Punishment deters lying, but can also 'teach' others the preferred meanings.

These results show that costly, social punishment is indeed an alternative explanation for honest signaling. Signals do not need to be costly as the handicap principle suggests: the cost may be paid by liars and punishers. This cost deters liars and so, is rarely paid.

5.3.2 Punishing timid senders is ineffective

The effect of $\mathtt{P_T}$, the honest signaling strategy that punishes timid behavior, on the frequency of honest signaling is mostly neutral and sometimes negative (Figure 5.4 and 5.6). This means that the cost of punishment does not outweigh its benefits. For more efficient punishment (larger ${}^{c''}/_{c'}$ and probability that *Sender*'s type is revealed p'), I observed a small improvement for high signal cost c and low common interest R.

5.3.3 Punishing greedy and worried receivers

Receiver can deviate from honest signaling in two ways (Table 5.1): he can be greedy (keep when *Sender* signals) or worried (donate when *Sender* is quiet). *Sender* can always detect such behavior. The strategy $\mathtt{P_G}$ punishes greedy behavior and the strategy $\mathtt{P_W}$ punishes worried behavior.

Punishing greedy individuals improves honest signaling for low common

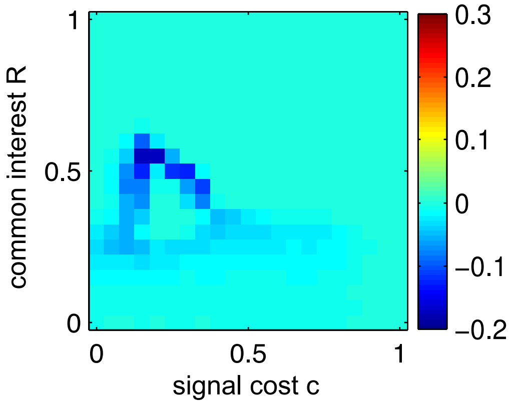


Figure 5.4: Increase in frequency of honest signaling when adding punishment of timid individuals $\mathtt{P_T}$. The parameters are the same as in Figure 5.1.

interest if signals are costly ($R < {}^3\!/_{10}$ and $c > {}^1\!/_5$ in Figure 5.5a). This region characterizes high conflicts of interest, where greed ($\mathtt{D}_\emptyset$) pays off in the original Philip Sidney game (Chapter 4). It also decreases the frequency of honest signaling for small signal costs (blue region near $(c, R) = ({}^1\!/_{10}, {}^3\!/_{10})$ in Figure 5.5a).

Punishing worried individuals slightly improves honest signaling if there is a full conflict or no conflict at all. The improvement is too small to be visible in Figures 5.5b and 5.6. Its effect is mostly neutral and even negative when there is only a conflict when *Sender* is healthy (blue region where ${}^1\!/_5 < R < 1 - c$ in Figure 5.5b).

Greedy and worried behavior can always be detected, while lying and timid behavior can only be detected with probability p'. This is the probability that *Sender*'s type is revealed. In the experiments reported here, $p' = {}^1\!/_2$. But punishing greedy and worried behavior is not necessarily more effective than punishing lying and timid behavior.

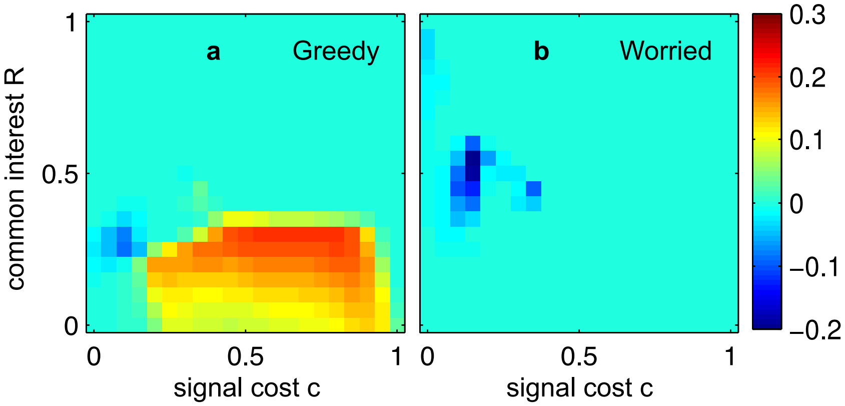


Figure 5.5: Increase in frequency of honest signaling when punishing *a)* greedy individuals $\mathtt{P_G}$ and *b)* worried individuals $\mathtt{P_W}$. The parameters are the same as in Figure 5.1.

5.3.4 Overall effects of punishment

To better understand the overall effects of different forms of punishment, I measured the average total frequency of honest signaling in different regions of conflict (Section 4.2). Figure 5.6 shows for varying signal cost c and common interest R the result for the region with a conflict only when *Sender* is healthy and the region of full conflict. The baseline mode replaces the punishment strategy with another pure honest signaling strategy and is labeled 'none.'

Punishing lying individuals is most effective when there is a conflict only when *Sender* is healthy and slightly effective when there is a full conflict. Punishing greedy individuals is very effective when there is a full conflict. Punishing worried individuals is ineffective and punishing timid individuals is, on average, even counterproductive for the evolution of honest signaling. Additional analysis shows that these results are robust for varying probability p' of detecting lying and timid behavior, effectiveness of punishment c''/c', and selection pressure β.

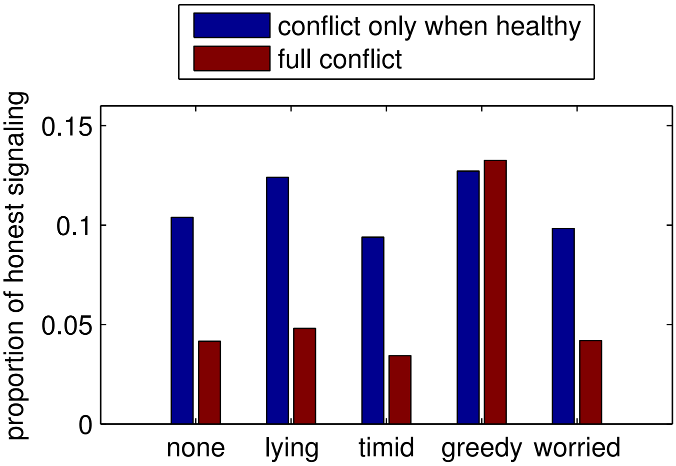


Figure 5.6: Frequency of honest signaling for different forms of punishment averaged over different regions of conflict as described in Section 4.2. The baseline model (column 'none') includes two honest signaling strategies without punishment. Punishing lying individuals is most effective when there is a conflict only when *Sender* is healthy. Punishing lying individuals (column 'lying') is most effective when there is a conflict only when *Sender* is healthy and slightly effective when there is a full conflict. Punishing timid individuals (column 'timid') is, on average, even counterproductive for the evolution of honest signaling. Punishing greedy individuals (column 'greedy') is very effective when there is a full conflict and punishing worried individuals (column 'worried') is ineffective. The parameters are the same as in Figure 5.1.

5.4 Related Work

The only related work is the study of signaling in animal contests (Enquist, 1985; Hurd, 1997; Hurd and Enquist, 1998). Such contests may lead to costly escalations which can be avoided by signaling (Maynard Smith and Price, 1973). House sparrows (Moller, 1987) and paper wasps (Tibbetts et al., 2010), for example, use badges as cost-free signals of fighting ability and avoid risky escalations. Enquist (1985) and Hurd (1997) constructed game theoretic models of animal contest which show that cost-free signaling can be an evolutionarily stable strategy since it is risky to exaggerate your fighting ability.

Many authors, including Maynard Smith and Harper (2003), interpret this as a form of punishment, though I find this disputable. Punishment assumes that one individual thinks or knows that the other cheated and responds aggressively towards the cheater. In the example of house sparrows and paper wasps (Section 5.1), when one animal cheats its opponent does not know that and there is no reason why it should think the other cheated since, more often than not, the signals are honest. An animal will attack its opponent because, for itself, this is the best response to a signal of high fighting ability and perhaps because it wants to find out whether or not his opponent signals truthfully. It cannot attack to punish his opponent for signaling dishonestly since it has yet to find out.

I believe individuals do not lie in these scenarios because of the uncertain effect of a dishonest signal. Since players are uncertain about their opponent's fighting ability they do not know whether a dishonest signal will scare off the opponent or encourage an escalation. Furthermore, the stronger animals run less risk by exaggerating their strength by the same amount than a weaker one.

In animal contests both players have private information: they are both *Sender* and *Receiver* at the same time. This makes it difficult to interpret what happens. In my extension of the Philip Sidney game, only one player has private information and punishment is modeled explicitly. It allows to clearly distinguish between punishment and *Receiver*'s response to a signal, and between different types of punishment (punishment of dishonest *Sender* by *Receiver*s and vice versa).

5.5 Conclusion

I studied whether or not punishment increases the emergence of honest signaling. In my extension of the Philip Sidney game, I distinguished four different forms of dishonest behavior: lying, timid, greedy, and worried behavior. Punishing lying individuals increases the frequency of honest signaling when the signals are cheap and even cost-free. This suggests an alternative for the handicap principle which is the most influential explanation for the evolution of honest signaling when interests conflict. Punishing greedy individuals increases the frequency of honest signaling when common interest is low and signals are sufficiently costly. Punishing timid or worried individuals is mostly counterproductive. They do not lead to any clear improvement in general and even result in an overall decrease of the frequency of honest signaling.

This chapter did not take antisocial punishment (punishing honest signalers) and spiteful punishment (punishing everyone) into account (Hilbe and Traulsen, 2012). While social punishment may promote the evolution of cooperation (Hauert et al., 2007; Hilbe and Traulsen, 2012; Sigmund et al., 2010), antisocial and spiteful punishment may destroy these benefits (Hilbe and Traulsen, 2012; Rand and Nowak, 2011). But antisocial and spiteful punishment may be avoided if reputation effects are taken into account (Hilbe and Traulsen, 2012) or prior agreements are made (Han et al., 2013). Future work could analyze whether these mechanisms can still deal with antisocial and spiteful punishment in the context honest signaling.

In general, social punishment (punishing those that defect or do not contribute to the public good) promotes the evolution of cooperation (Hauert et al., 2007; Hilbe and Traulsen, 2012; Sigmund et al., 2010). In my extension of the Philip Sidney game, this is not the case. Punishment is more complex, showing diverse possibilities which result in different outcomes. Greedy individuals in the Philip Sidney game are similar to defectors in a public goods game or the prisoners' dilemma, but the other types of dishonest behavior (lying, timid and worried) are not present in these games.

In short, this chapter demonstrates that punishing dishonest behavior promotes the evolution of honest signaling in several situations. Signaling provides a richer and more complex framework for the study of the evolutionary roles of punishment than the context of cooperation. More effort is required to clarify the role of antisocial punishment in the evolution of honest signaling.

Chapter 6

Conclusion

This chapter gives a brief summary, some critique, and three directions for future work.

Contents

6.1 Summary

This thesis studied the emergence of honest signaling, a problem which is related to philosophy, linguistics, economics, and biology. It consists of two questions: “How do signals emerge?” and “Why are signals honest?”.

Throughout the thesis, I provide evidence that signals emerge by chance. The meaning of signals is the outcome of stochastic processes that simulate individual learning, social learning, or evolution. In the Lewis signaling game all separating equilibria—states of perfect communication—are equally preferred and emerge with equal probability. In the Philip Sidney game, there are two separating equilibria, but the signal cost renders one, honest signaling, more preferable than the other, inverse honest signaling. Still, both emerge, honest signaling simply has higher probability than the other separating equilibrium.

The thesis supports several answers to the second question (“Why are signals honest?”). The first answer is that signaling can emerge when agents have common interests. This was demonstrated in Chapter 3. First, I invented a new behavioral rule: win-stay/lose-inaction (WSLI) and proved that two individuals repeatedly interacting in any Lewis signaling game always reach a separating equilibrium if they apply WSLI. Moreover, the number of iterations they need is only polynomial in the number of signals of the game: $\mathcal{O}(n^3)$. Second, I gave some reinforcement learning algorithms that perform as well as WSLI but can also cope with errors (in observation or execution).

The second answer, given in Chapter 4, is that signaling can emerge when agents have conflicting interest, provided that the signals are costly: the handicap principle. Most biologists study this by evaluating whether or not honest signaling is an equilibrium of the Philip Sidney game or any of its variants. Because static equilibrium analyses do not reveal whether signaling can emerge, I applied learning and evolutionary dynamics. This lead to two insights. First, independent of the dynamics, there are many cases where honest signaling is an equilibrium but where the dynamics does not lead to it. Learning dynamics lead to honest signaling only when it was a Pareto optimal Nash equilibrium. Evolutionary dynamics in finite populations lead to honest signaling in a region quite similar to that of learning dynamics, and very different from the region where honest signaling is an equilibrium. Second, both dynamics revealed some cases where honest signaling is not an equilibrium but where some signaling still emerged. Learning dynamics revealed that for cheap signals, individuals were sometimes honest and sometimes dishonest. The evolutionary dynamics revealed that honest signaling can still be the most frequent strategy.

The third answer, given in Chapter 5, is that signaling can emerge if honest agents can punish dishonest ones even though they have conflicting interests and signals are cheap or cost-free. I invented a new game to model this: the Philip Sidney game with punishment. Evolutionary dynamics in finite populations revealed that punishing liars is effective when signals are cheap hence showing that punishment is indeed an alternative for the handicap principle. Punishing greedy individuals is effective when common interest is low but signals are costly. Punishing worried or timid individuals is not effective.

6.2 Critique

In Chapter 3, where two agents applied WSLI and interacted in the Lewis signaling game, I did not discuss how this generalizes to games with more or less signals than the number of types or to populations of agents that are randomly matched. The generalization from Lewis signaling games to signaling games with more signals than types and responses is straightforward: the excess signals will remain unused. The generalization to signaling games with less signals than types and responses is as follows. WSLI will lead to a Nash equilibrium where only as many types as signals are successfully communicated. The most frequent types have the highest probability to be successfully communicated. So, for uniform type distributions, WSLI will still reach one of the Pareto optimal Nash equilibria, but for non-uniform type distributions this is no longer guaranteed. Though, the more equilibria an equilibrium dominates, the higher its chances. Alternately, it is possible to equip the agents with the capacity to invent signals (Skyrms, 2010, ch.

10).

For several dynamics, the generalization from two agents to populations leads to dialects (Zollman, 2005). Mechanism, such as social structure (Wagner, 2009) or combining evolution and learning (Zollman and Smead, 2010), may counter this effect. Preliminary experiments show that this is also the case for WSLI and its robust implementations.

Though I mentioned how signaling is important for understanding the origins of language and many phenomena in economics and biology, I did not give any applications or implications for computer science. I think this research is still valuable for computer science or engineering, especially for the design of multiagent systems. The designer of a multiagent system usually implements a shared coordination protocol which allows the agents to coordinate (Tambe, 1997). Agents who do not share a common coordination protocol, may still coordinate if they share a common language (Barrett et al., 2013). If the agents can learn or bootstrap a language, the system designer does not even need to design the language beforehand—he may simply provide the agents with the necessary learning capabilities. A language invented by the agents themselves may be more efficient than what a system designer can come up with, since the language will be adapted to the specific coordination problem the agents face and may even coevolve with the problem.

Servin and Kudenko (2008) implemented a multiagent system to detect intrusions into a computer network. At different locations in the network, a ‘sensor’ agent monitors traffic. Some sensor agent are on routers, others on end user computers. They detect different features that may or may not indicate an attack. Detecting an attack by monitoring only one node is unreliable, so the sensor agents send signals to a ‘response’ agent. The latter decides whether or not to warn the network administrator of a possible attack. In this example, the system designer did not know in advance what information should be transmitted, the system had to learn what to communicate when. My work shows what type of algorithm could work well for such a scenario: learning to signal in a cooperative setting. Servin and Kudenko (2008) chose Q-learning with decreasing softmax which corresponds to what I proposed.

Another example is ‘vehicular ad hoc networks’ or VANETs where cars exchange information in case of natural disasters (Camara et al., 2010), road accidents (Xu et al., 2004), or special road conditions (Yang et al., 2004). Users must be able to fully trust this information. Haddadou and Rachedi (2013) propose that sending messages is costly but when a message is validated the sender is rewarded. My research warns that to calculate the necessary cost, you cannot simply rely on equilibrium analysis. You should use multiple static and dynamic models, to increase the evidence that no cheating will emerge. It also shows that you could punish and use cheap(er) signals. Punishing those that exaggerate a situation (liars) is probably efficient while punishing those that underestimate a situation (timid individuals) is not.

6.3 Future Work

This work can be extended in many ways. I discuss three of them.

From signaling games to more complex signaling problems

Lewis signaling games can be extended to more complex communication problems: games with multiple senders or receivers; dialogues, where the game consists of two stages and the players’ roles are reversed in the second stage; signaling chains, where a player’s private information is the signal sent by the previous player; etc. (Skyrms, 2010, ch. 11 and 13). Future work could investigate whether WSLI still leads to efficient signaling in these more complex problems and what happens if the players have an incentive to cheat like in the Philip Sidney game.

Everyone can punish everyone

Chapter 5 was restricted to social punishment: only honest individuals could punish and only dishonest individuals could be punished. Future work could allow everyone to punish everyone. It is unclear how this will affect honest signaling though the effect on the evolution of cooperation was studied before. Antisocial (punishing cooperators) and spiteful punishment (punishing everyone) destroys the benefits that social punishment provides for the evolution of cooperation (Hilbe and Traulsen, 2012; Rand and Nowak, 2011) but mechanisms like reputation (Hilbe and Traulsen, 2012) and prior agreement (Han et al., 2013) counter these effects. Future work could analyze whether these mechanisms can still deal with antisocial and spiteful punishment in the context honest signaling.

Other mechanisms that promote honest signaling

Future work could investigate other mechanisms to promote honest signaling between agents with conflicting interests than those studied in this thesis: costly signals and punishment. (Számadó, 2011; Zollman, 2013) proposed many mechanisms all of which, I think, can be categorized in two groups: cost based mechanisms and correlation based mechanisms. Cost based mechanisms like costly signals and punishment alter the payoffs in a way that benefits honest signaling. A proximity risk is an other example: animals often use threat displays to settle conflicts but these are only reliable when performed within 'striking distance,' close enough to the opponent such that the animal runs a risk that his opponent strikes. This risk, like punishment, is a potential cost to cheating. Correlation based mechanisms alter the individuals with which players interact. For example, players could avoid individuals they do not trust or the social network structure could protect clusters of honest individuals.

Appendix A

Finite Markov Chains

This appendix gives some definitions and properties of finite Markov chains. You can find these in any classic textbook on the subject (for example Meyn and Tweedie (1993)), but they are mentioned here for convenience.

Finite Markov chains occur in several places in this thesis. In Section 3.3.2, I model the learning process of win-stay/lose-inaction in Lewis signaling games with an absorbing Markov chain (Appendix A.3). The evolutionary dynamics in finite populations of Section 2.3.1 uses a Markov chain twice. Once to calculate the probability that one mutant takes over an entire population, also known as the 'fixation probability' (Appendix A.4) and once to calculate the relative amount of time each strategy is used by the entire population (Appendix A.2).

A.1 Definition

A finite Markov chain is a *stochastic process* that is discrete in time and state space and is memoryless: the future only depends on the current state and not on the history.

Let the random variable X_n be the state of the process at time $n \in \mathbb{N}$ and take values $s_n \in \mathcal{S} = \{1, 2, ..., m\}$.

Definition A.1. The sequence of random variables $X_0, X_1, X_2, X_3, \ldots$ is a *Markov chain* if the probability distribution over X_{n+1} only depends on the current state X_n and not on the history $X_{n-1}, X_{n-2}, \ldots, X_0$:

$$\Pr\left(X_{n+1} \mid X_n = s_n, X_{n-1} = s_{n-1}, \ldots, X_0 = s_0\right) = \Pr\left(X_{n+1} \mid X_n = s_n\right).$$

Definition A.2. A Markov chain is *time-homogeneous* if the conditional probabilities $\Pr(X_{n+1} \mid X_n = s_n)$ are independent of time n.

From hereon, I use 'Markov chain' to refer to finite, time-homogeneous Markov chains. The statements below do not necessarily hold for infinite Markov chains or Markov chains that are not time-homogeneous.

A Markov chain with m states is defined by an $m \times m$ *transition matrix* P. Each element $P_{i,j}$ of that matrix represents the probability that the process transitions from state i to j given that it is in state i:

$$P_{i,j} = \Pr(X_{n+1} = j \mid X_n = i) \text{ for any } n.$$

P is a *row-stochastic* matrix: on each row of the transition matrix the elements sum up to 1 and no element is negative. The probability to go from state i to j in k steps is denoted $(P^k)_{i,j}$, it is the element in the ith row and jth column of matrix P^k (the transition matrix P to the power k).

Given the current probability distribution $\Pr(X_n)$ over all states $\mathcal{S}$, the next distribution is $\Pr(X_{n+1}) = \Pr(X_n)P$. The probability distribution $\Pr(X_{n+k})$ over the states after k steps is $\Pr(X_n)P^k$.

Each Markov chain has a corresponding directed graph that has a node for each state and an edge from node i to j if the Markov chain can make the transition from state i to j in one step: $P_{i,j} > 0$.

A.2 Stationary Distribution

Definition A.3. A *stationary distribution* π of a Markov chain with transition matrix P is a probability distribution over all states such that $\pi P = \pi$. Once the process reaches a stationary distribution, it remains there.

According to the Perron-Frobenius theorem, every stochastic matrix P that has only positive elements ($P_{i,j} > 0$) has an eigenvalue 1 and the corresponding eigenvector π has only positive components ($\pi_i > 0$). The vector π is unique up to a constant factor, so it must be normalized to be a valid probability distribution ($\sum_i \pi_i = 1$).

The stationary distribution π is a long-term prediction of the process independent of the initial distribution and also gives the fraction of time spent in each state.

A.3 Absorbing Markov Chains

A special class of Markov chains are those that end up in a state from which the process can no longer escape. These are called 'absorbing Markov chains' and the states from which the process cannot escape are called 'absorbing states.' Given an initial state, it is possible to calculate the probability of ending up in any of the absorbing states and also the number of steps before reaching an absorbing state.

First some definitions.

Definition A.4. State i is *accessible* from state j if there is a path in the Markov chain's graph from state j to i.

Definition A.5. An *absorbing state* i is a state of the Markov chain from which it is impossible to escape: $P_{i,i} = 1$ and thus $P_{i,j} = 0$ for all $j \neq i$.

Definition A.6. An *absorbing Markov chain* is a Markov chain with at least one absorbing state and a path in the Markov chain's graph from every state to an absorbing state.

Definition A.7. In an absorbing Markov chain, a state which is not absorbing is a *transient state.*

Now, consider the *canonical form* P' of the transition matrix P. It has the states of transition matrix P reordered such that first t states are transient and the last r states are absorbing:

$$P' = \begin{pmatrix} Q & R \\ Z & I \end{pmatrix}.$$

Here, I denotes the $r \times r$ identity matrix. Z the $r \times t$-matrix that contains only zeros. R is the $t \times r$-matrix with the transition probabilities from every transient state to every absorbing state. At least some elements in R are nonzero. Q is the $t \times t$-matrix with the transition probabilities between all transient states.

Finally, consider the $t \times t$ *fundamental matrix* F. Each element $F_{i,j}$ is the number of times the process is in transient state j when it started from the transient state i:

$$\begin{aligned} F &= \sum_{k=0}^{\infty} Q^k \\ &= I + Q + Q^2 + Q^3 + \ldots \\ &= (I - Q)^{-1} \end{aligned}$$

It is now possible to calculate the absorption probabilities and the expected time till absorption. Given the fundamental matrix F the probability that transient state i leads to absorbing state j is $B_{i,j}$ where $B = FR$ and the number of steps until transient state i leads to any absorbing state is the i'th component of $F\vec{1}$, where $\vec{1}$ is the column vector whose elements are all 1.

A.4 Absorption Probability in a Moran Process

A Moran process (Moran, 1958) is a stochastic process that can be represented by a Markov chain with $N+1$ states $\mathcal{S} = \{0, 1, \ldots, N\}$ and transition matrix P (Figure A.1). From every state i, only states $i-1$, i, and $i+1$ are accessible in one step: $P_{i,j} > 0$ implies $j = i-1, i,$ or $i+1$. States 0 and N are absorbing: $P_{0,0} = P_{N,N} = 1$.

Theorem 2. *Assume that the Moran process with transition matrix P (Figure A.1) starts in state 1. The probability x_1 that the process ends up in state N is*

$$x_1 = \frac{1}{1 + \sum_{i=1}^{N-1} \prod_{j=1}^{i} \frac{P_{j,j-1}}{P_{j,j+1}}}.$$

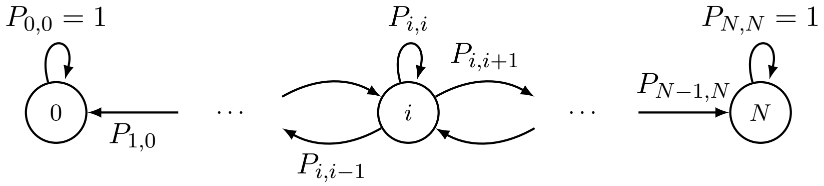


Figure A.1: A Markov chain with transition matrix P and states $i = 0, 1, \dots, N$. From every state i, only states $i-1$, i, and $i+1$ are accessible in one step. States 0 and N are absorbing.

Proof. Let p_i be the probability to reach state N from state i. Because states 0 and N are absorbing, the probability to reach state N from state 0 is 0 ($p_0 = 0$) and the probability to reach state N from state N itself is 1 ($p_N = 1$). The probability to reach state N from state $i = 1, 2, \dots, N-1$ is

$$\begin{aligned} p_i &= P_{i,i-1}\, p_{i-1} + P_{i,i}\, p_i + P_{i,i+1}\, p_{i+1} \\ &= P_{i,i-1}\, p_{i-1} + (1 - P_{i,i-1} - P_{i,i+1})\, p_i + P_{i,i+1}\, p_{i+1} \end{aligned}$$

This can be rewritten by grouping $P_{i,i+1}$ on the left and $P_{i,i-1}$ on the right hand side of the equation:

$$P_{i,i+1}(p_{i+1} - p_i) = P_{i,i-1}(p_i - p_{i-1})$$

or

$$(p_{i+1} - p_i) = \frac{P_{i,i-1}}{P_{i,i+1}}(p_i - p_{i-1}).$$

By working out the recursion above, $p_{i+1} - p_i$ can be written as a function of p_1, because $p_0 = 0$ and so $p_1 - p_0 = p_1$:

$$p_{i+1} - p_i = p_1 \prod_{j=1}^{i} \frac{P_{j,j-1}}{P_{j,j+1}}. \tag{A.1}$$

The sum over $p_{i+1} - p_i$ from 0 to $N-1$ is 1, because $p_0 = 0$ and $p_N = 1$:

$$\begin{aligned}\sum_{i=0}^{N-1}(p_{i+1} - p_i) &= (p_1 - p_0) + (p_2 - p_1) + \ldots + (p_N - p_{N-1})\\ &= p_N - p_0\\ &= 1. \end{aligned} \tag{A.2}$$

The probability p_1 to reach state N from state 1 can be expressed as a function of the transition matrix P by combining Equation (A.1) and (A.2):

$$\begin{aligned}1 = \sum_{i=0}^{N-1}(p_{i+1} - p_i) &= \sum_{i=0}^{N-1}(p_1 \prod_{j=1}^{i} \frac{P_{j,j-1}}{P_{j,j+1}})\\ &= p_1 \sum_{i=0}^{N-1} \prod_{j=1}^{i} \frac{P_{j,j-1}}{P_{j,j+1}}\end{aligned}$$

so

$$\begin{aligned}p_1 &= \frac{1}{\sum_{i=0}^{N-1} \prod_{j=1}^{i} \frac{P_{j,j-1}}{P_{j,j+1}}}\\ &= \frac{1}{1 + \sum_{i=1}^{N-1} \prod_{j=1}^{i} \frac{P_{j,j-1}}{P_{j,j+1}}}.\end{aligned}$$

□

When $P_{i,i+1} = P_{i,i-1}$ for all states i, the probability p_1 only depends on the number of states N:

$$\begin{aligned}p_1 &= \frac{1}{1 + \sum_{i=1}^{N-1} \prod_{j=1}^{i} \frac{P_{j,j-1}}{P_{j,j+1}}}\\ &= \frac{1}{1 + \sum_{i=1}^{N-1} \prod_{j=1}^{i} 1}\\ &= \frac{1}{N}.\end{aligned}$$

Appendix B

The Nash Equilibrium and its Refinements in Signaling Games

The *Nash equilibria* (Nash, 1950a) of a signaling game (Definition 2.5) are the Nash equilibria of its corresponding strategic form—just as for all other finite extensive form games. Since different mixed strategies—probability distributions over the set of pure strategies—may cause the same behavior and correspond to the same behavioral strategy, it makes more sense to define equilibria in terms of behavioral strategies (Gintis, 2000). Remember that a behavioral strategy for a player is a probability distribution over all actions at each of his information sets. *Sender*'s behavioral strategy β_1 defines a probability distribution over the signals $m \in \mathcal{M}$ for all types $t \in \mathcal{T}$ and *Receiver*'s behavioral strategy β_2 defines a probability distribution over the responses $r \in \mathcal{R}$ for all signals $m \in \mathcal{M}$. For example, $\beta_1(m|t)$ is the probability that *Sender* uses signal m when observing type t. The strategy profile (β_1, β_2) is a Nash equilibrium of the signaling game $(\mathcal{T}, \mathcal{M}, \mathcal{R}, \pi, u)$ when:

- For each type t, *Sender* only uses signals m^* that maximize his payoff against *Receiver*'s strategy β_2:

$$\text{For all types } t \in \mathcal{T}:$$

$$\beta_1(m^*|t) > 0 \text{ implies } m^* = \arg\max_{m} \sum_{r} \beta_2(r|m)\, u_1(t, m, r). \quad \text{(B.1)}$$

- For each signal m used in equilibrium, *Receiver* only uses responses r^* that maximize his payoff against *Sender*'s strategy $\beta_1(m|t)$ and the

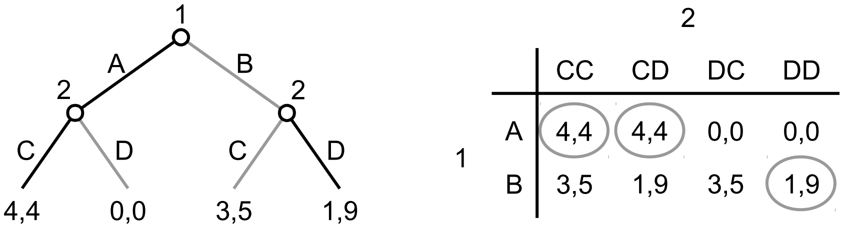


Figure B.1: *Left*: An extensive form game. At each decision node, the action maximizing payoff is drawn in black and the other actions are drawn in gray. The subgame perfect equilibrium outcome is $(\mathtt{A}, \mathtt{C})$. *Right*: The corresponding strategic form which has three pure strategy Nash equilibria: $(\mathtt{A}, \mathtt{CC})$, $(\mathtt{A}, \mathtt{CD})$, and $(\mathtt{B}, \mathtt{DD})$. From the point of view of the extensive form, only the second one seems rational. The others consist of a so-called incredible threat by player 2.

type distribution π:

$$\text{For all signals } m \in \mathcal{M} \text{ such that } \sum_t \beta_1(m|t)\,\pi_t > 0:$$
$$\beta_2(r^*|m) > 0 \text{ implies } r^* = \arg\max_r \sum_t \Pr(t|m)\,u_2(t,m,r), \quad \text{(B.2)}$$

where $\Pr(t|m)$ denotes the *Receiver*'s belief that *Sender* has type t when sending signal m. These posterior beliefs must be consistent with *Sender*'s strategy and the prior type distribution according to Bayes' rule:

$$\Pr(t|m) = \frac{\beta_1(m|t)\,\pi_t}{\sum_{t' \in \mathcal{T}} \beta_1(m|t')\,\pi_{t'}}. \quad \text{(B.3)}$$

Signaling games have many Nash equilibria and game theory cannot predict what players will or should do. To solve this equilibrium selection problem refinements try to restrict the set of equilibria.

The *subgame perfect equilibrium* (Selten, 1965) is a refinement for games in extensive form (Section 2.2.1). It requires equilibrium strategies to be optimal in every subgame—a subtree that does not split up any information set—and not only on the path of equilibrium play. It excludes Nash equilibria that involve so-called 'incredible threats.' For example, the pure Nash equilibria of the extensive form game in Figure B.1 are the same as those of its corresponding strategic form: $(\mathtt{A}, \mathtt{CC})$, $(\mathtt{A}, \mathtt{CD})$, and $(\mathtt{B}, \mathtt{DD})$. The last one is based on player 2's threat to play `D` after `A` which is incredible since, once `A` is played the best player 2 can do is to play `C`, not `D`. Similarly, the Nash equilibrium $(\mathtt{A}, \mathtt{CC})$ is based on an incredible threat. The only subgame perfect equilibrium is $(\mathtt{A}, \mathtt{CD})$.

The subgame perfect equilibrium has no effect on extensive form games with incomplete information (Section 2.2.4) because they have no proper subgames. The only subgame is the entire game itself because the information sets of uninformed players contain a decision node which have the root node as their only common predecessor. Any Nash equilibrium of an incomplete information game is subgame perfect.

The *perfect Bayesian equilibrium* (Fudenberg and Tirole, 1991a) is similar to subgame perfection but can refine Nash equilibria in incomplete information games. Each player must play optimal at each information set as opposed to each subgame. For signaling games, it is easily defined in terms of Nash equilibria in behavioral strategies: it is a Nash equilibrium where *Receiver*'s strategy is optimal for all signals, including those observed with zero probability in equilibrium. In other words, Equation (B.2) should hold for all signals $m \in \mathcal{M}$, not just for those used in equilibrium. *Receiver*'s belief $\Pr(t|m)$ for unobserved signals m is arbitrary. This refinement does not rule out strict Nash equilibria or Nash equilibria where every signal is used. It only rules out some equilibria where multiple best replies are available (Binmore, 2007, ch. 14). For incomplete information games, the perfect Bayesian equilibrium is stronger than the subgame perfect equilibrium.

Another refinement for extensive form games is the *sequential equilibrium* (Kreps and Wilson, 1982). For incomplete information games with at most two types or at most two stages the set of sequential equilibria coincides with the set of perfect Bayesian equilibria (Fudenberg and Tirole, 1991b). Since signaling games are two-stage games (first *Sender* plays, then *Receiver*) all perfect Bayesian equilibria are sequential equilibria and vice versa.

Other refinements further restrict the possible beliefs of *Receiver* by imposing extra requirements on the players' rationality. The research was started by Cho and Kreps (1987) and lead to a whole range of refinements for signaling games, none of which are generally applicable or agreed upon. Examples of such refinements are Condition D1, Divinity, and the Intuitive Criterion (Sobel, 2009). They require players to be 'unrealistically' rational and are still a source of much debate. See for example (Riley, 2001) and (Binmore, 2007, ch. 14).

Appendix C

The Philip Sidney Game in Strategic Form

This appendix gives the payoff table of the Philip Sidney game in strategic form.

	$\mathtt{D_A}$	$\mathtt{D_\emptyset}$	$\mathtt{D_S}$	$\mathtt{D_Q}$
$\mathtt{S_A}$	$d+RS$ $S+Rd$	$pVd+R$ $1+pRVd$	$d+RS$ $S+Rd$	$pVd+R$ $1+pRVd$
$\mathtt{S_\emptyset}$	$1+RS$ $S+R$	$pV+R$ $1+pRV$	$pV+R$ $1+pRV$	$1+RS$ $S+R$
$\mathtt{S_N}$	$d+RS+pc$ $S+Rd+pRc$	$pV+R$ $1+pRV$	$p(V+R)+q(d+RS)$ $p(1+RV)+q(S+Rd)$	$p(1+RS)+qR$ $p(S+R)+q$
$\mathtt{S_H}$	$1+RS-pc$ $S+R(1-pc)$	$R+pVd$ $1+pRVd$	$p(d+RS)+qR$ $p(S+Rd)+q$	$p(Vd+R)+q(1+RS)$ $p(1+RVd)+q(S+R)$

Table C.1: Payoff table for the Philip Sidney game in strategic form. Each entry gives *Sender*'s payoff on the first line and *Receiver*'s payoff on the second line. To be able to fit everything on one page, I substituted $1-p$ by q and $1-c$ by d. The meaning of the parameters is given in Table 4.1 and the meaning of the strategies is given in Table 4.2.

Index

Bibliography

Akerlof, George A. (1970). "The Market for "Lemons:" Quality Uncertainty and the Market Mechanism." In: *Quarterly Journal of Economics* 84.3, pp. 488–500.

Argiento, Raffaele, Robin Pemantle, Brian Skyrms, and Stanislav Volkov (2009). "Learning to signal: Analysis of a micro-level reinforcement model." In: *Stochastic Processes and their Applications* 119.2, pp. 373–390.

Arthur, W. Brian (1993). "On designing economic agents that behave like human agents." In: *Journal of Evolutionary Economics* 3.1, pp. 1–22.

Auer, Peter, Nicolò Cesa-Bianchi, and Paul Fischer (2002). "Finite-time Analysis of the Multiarmed Bandit Problem." In: *Machine Learning* 47.2, pp. 235–256.

Auer, Peter, Nicolò Cesa-Bianchi, Yoav Freund, and Robert E. Schapire (2003). "The Nonstochastic Multiarmed Bandit Problem." In: *SIAM Journal on Computing* 32.1, pp. 48–77.

Aumann, Robert and Adam Brandenburger (1995). "Epistemic conditions for Nash equilibrium." In: *Econometrica* 63.5, pp. 1161–1180.

Barrett, Jeffrey A. (2006). *Numerical Simulations of the Lewis Signaling Game: Learning Strategies, Pooling Equilibria, and the Evolution of Grammar*. Tech. rep. September. Irvine, CA, USA: University of California, Irvine: Institute for Mathematical Behavioral Science.

Barrett, Jeffrey A. and Kevin J. S. Zollman (2009). "The role of forgetting in the evolution and learning of language." In: *Journal of Experimental & Theoretical Artificial Intelligence* 21.4, pp. 293–309.

Barrett, Samuel, Noa Agmon, Noam Hazon, et al. (2013). "Communicating with Unknown Teammates." In: *Proceedings of the 13th Adaptive and Learning Agents workshop*. Ed. by Sam Devlin, Daniel Hennes, and Enda Howly. Saint Paul, MN, USA, pp. 46–52.

Beggs, Alan W. (2005). "On the convergence of reinforcement learning." In: *Journal of Economic Theory* 122.1, pp. 1–36.

Bereby-Meyer, Y. and Ido Erev (1998). "On Learning To Become a Successful Loser: A Comparison of Alternative Abstractions of Learning Processes in the Loss Domain." In: *Journal of mathematical psychology* 42.2/3, pp. 266–286.

Bergstrom, Carl T. and Michael Lachmann (1997). “Signalling among relatives. I. Is costly signalling too costly?” In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 352.1353, pp. 609–617.

— (1998). “Signaling among relatives. III. Talk is cheap.” In: *Proceedings of the National Academy of Sciences of the United States of America* 95.9, pp. 5100–5105.

Binmore, Ken G. (2007). *Playing for Real: a text on game theory.* New York: Oxford University Press.

Boyd, Robert and Peter J. Richerson (1992). “Punishment Allows the Evolution of Cooperation (or Anything Else) in Sizable Groups.” In: *Ethology and Sociobiology* 195.13, pp. 171–195.

Brilot, Ben O. and Rufus A. Johnstone (2003). “The limits to cost-free signalling of need between relatives.” In: *Proceedings of the Royal Society B: Biological Sciences* 270.1519, pp. 1055–1060.

Camara, Daniel, Christian Bonnet, and Fethi Filali (2010). “Propagation of Public Safety Warning Messages: A Delay Tolerant Network Approach.” In: *IEEE Wireless Communication and Networking Conference.* Sydney, Australia: IEEE, pp. 1–6.

Catteeuw, David (2014). “The Emergence of Honest Signaling.” In: *European Conference on Complex Systems.* Lucca, Italy, p. 1.

Catteeuw, David and Bernard Manderick (2009). “Learning in the Time-Dependent Minority Game.” In: *Proceedings of the 11th annual conference on Genetic and Evolutionary Computation.* Montréal, Canada: ACM Press, pp. 2011–2016.

— (2011a). “Heterogeneous Populations of Learning Agents in Minority Games.” In: *Proceedings of the 11th Adaptive and Learning Agents workshop.* Ed. by Peter Vrancx, Matt Knudson, and Marek Grzes. Taipei, Taiwan, pp. 15–20.

— (2011b). “Heterogeneous Populations of Learning Agents in the Minority Game.” In: *Lecture Notes in Computer Science, Adaptive and Learning Agents* 7113, pp. 100–113.

— (2011c). “Learning in Minority Games with Multiple Resources.” In: *Lecture Notes in Computer Science, Advances in Artificial Life* 5778. Ed. by George Kampis, István Karsai, and Eörs Szathmáry, pp. 326–333.

— (2012a). “Emergence of Honest Signaling through Reinforcement Learning.” In: *Proceedings of the 12th Adaptive and Learning Agents workshop.* Ed. by Enda Howley, Peter Vrancx, and Matt Knudson. Valencia, Spain, pp. 81–86.

— (2012b). “Honest Signaling: Learning Dynamics versus Evolutionary Stability.” In: *Proceedings of the 21st Belgian-Dutch Conference on Machine Learn-*

ing. Ed. by Bernard De Baets, Bernard Manderick, Michael Rademaker, and Willem Waegeman. Ghent, Belgium, pp. 1–6.

— (2013). "The Limits of Reinforcement Learning in Lewis Signaling Games." In: *Proceedings of the 13th Adaptive and Learning Agents workshop*. Ed. by Sam Devlin, Daniel Hennes, and Enda Howly. Saint Paul, MN, USA, pp. 22–30.

— (2014). "The Limits and Robustness of Reinforcement Learning in Lewis Signaling Games." In: *Connection Science* 26.2, pp. 161–177.

— (in press). "Honesty and deception in populations of selfish, adaptive individuals." In: *The Knowledge Engineering Review* 31.2.

Catteeuw, David, Joachim De Beule, and Bernard Manderick (2011). "Roth-Erev Learning in Signaling and Language Games." In: *Proceedings of the 23rd Benelux Conference on Artificial Intelligence*. Ed. by Patrick De Causmaecker, Joris Maervoet, Tommy Messelis, et al. Ghent, Belgium, pp. 65–74.

Catteeuw, David, Bernard Manderick, and The Anh Han (2013). "Evolutionary Stability of Honest Signaling in Finite Populations." In: *Proceedings of the IEEE Congress on Evolutionary Computation*. Ed. by Luis Gerardo de la Fraga and Carlos A. Coello Coello. Cancun, Mexico: IEEE Computer Society, pp. 2864–2870.

Catteeuw, David, The Anh Han, and Bernard Manderick (2014a). "Evolution of Honest Signaling by Social Punishment." In: *Proceedings of the 2014 Genetic and Evolutionary Computation Conference*. Ed. by Christian Igel and Dirk V. Arnold. Vancouver, BC, Canada: ACM Press, pp. 153–160.

Catteeuw, David, Madalina M. Drugan, and Bernard Manderick (2014b). "'Guided' Restarts Hill-Climbing." In: *In Search of Synergies between Reinforcement learning and Evolutionary Computation, Workshop at the 13th International Conference on Parallel Problem Solving from Nature*. Ed. by Madalina M. Drugan and Bernard Manderick. Ljubljana, Slovenia, pp. 1–4.

— (2014c). "'Guided' Restarts Hill-Climbing." In: *International Conference on Metaheuristics and Nature Inspired Computing*. Marrakech, Morocco, pp. 1–2.

Chicago Mercantile Exchange (2006). *An Introduction to Futures and Options*.

Cho, In-Koo and David M. Kreps (1987). "Signaling Games and Stable Equilibria." In: *The Quarterly Journal of Economics* 102.2, pp. 179–221.

Claus, Caroline and Craig Boutilier (1998). "The Dynamics of Reinforcement Learning in Cooperative Multiagent Systems." In: *Proceedings of the National Conference on Artificial Intelligence*. AAAI Press, pp. 746–752.

Clutton-Brock, TH and GA Parker (1995). "Punishment in animal societies." In: *Nature* 373, pp. 209–216.

Cooper, David J and John B Van Huyck (2003). “Evidence on the equivalence of the strategic and extensive form representation of games.” In: *Journal of Economic Theory* 110.2, pp. 290–308.

Daners, Daniel (2012). “A Short Elementary Proof of Σ 1/kˆ2 = πˆ2/6.” In: *Mathematics Magazine* 85.5, pp. 361–364.

De Beule, Joachim, Bart De Vylder, and Tony Belpaeme (2006). “A cross-situational learning algorithm for damping homonymy in the guessing game.” In: *Proceedings of the 10th International Conference on the Simulation and Synthesis of Living Systems.* Ed. by Luis M. Rocha, Larry S. Yaeger, Mark A. Bedeau, et al. Bloomington, IN, USA: MIT Press, pp. 466–472.

Ekman, Paul (1992). “Facial Expressions of Emotion: New Findings, New Questions.” In: *Psychological Science* 3.1, pp. 34–38.

Enquist, Magnus (1985). “Communication during aggressive interactions with particular reference to variation in choice of behaviour.” In: *Animal Behaviour* 33.4, pp. 1152–1161.

Fehr, Ernst and Simon Gächter (2002). “Altruistic punishment in humans.” In: *Nature* 415.6868, pp. 137–40.

Fudenberg, Drew and Lorens A. Imhof (2006). “Imitation Processes with Small Mutations.” In: *Journal of Economic Theory* 131.1, pp. 251–262.

Fudenberg, Drew and David K. Levine (1998a). “Learning in games.” In: *European Economic Review* 42.3-5, pp. 631–639.

— (1998b). *The Theory Of Learning In Games.* Cambridge, MA: The MIT Press.

Fudenberg, Drew and Jean Tirole (1991a). *Game Theory.* Cambridge, MA: MIT Press.

— (1991b). “Perfect Bayesian equilibrium and sequential equilibrium.” In: *Journal of Economic Theory* 53.2, pp. 236–260.

Gintis, Herbert (2000). *Game Theory Evolving.* Princeton University Press, p. 531.

Godfray, H. C. J. (1991). “Signalling of need by offspring to their parents.” In: *Nature* 352.6333, pp. 328–330.

Gokhale, Chaitanya S. and Arne Traulsen (2010). “Evolutionary games in the multiverse.” In: *Proceedings of the National Academy of Sciences* 107.12, pp. 5500–5504.

Grafen, Alan (1990). “Biological signals as handicaps.” In: *Journal of Theoretical Biology* 144.4, pp. 517–546.

Guala, Francesco (2012). “Reciprocity: weak or strong? What punishment experiments do (and do not) demonstrate.” In: *The Behavioral and brain sciences* 35.1, pp. 1–15.

Haddadou, Nadia and Abderrezak Rachedi (2013). "DTM2: Adapting job market signaling for distributed trust management in vehicular ad hoc networks." In: *IEEE International Conference on Communications*. Budapest, Hungary, pp. 1827–1832.

Hamilton, W. D. (1964). "The genetical evolution of social behaviour." In: *Journal of Theoretical Biology* 7.1, pp. 1–52.

Han, The Anh, Arne Traulsen, and Chaitanya S. Gokhale (2012). "On equilibrium properties of evolutionary multi-player games with random payoff matrices." In: *Theoretical population biology* 81.4, pp. 264–272.

Han, The Anh, Luís Moniz Pereira, Francisco C. Santos, and Tom Lenaerts (2013). "Good agreements make good friends." In: *Scientific reports* 3, p. 2695.

Harford, Tim (2006). *The Undercover Economist*. Londen: Abacus.

Harsanyi, John C. (1967). "Games with Incomplete Information Played by Bayesian Players, Parts I, II and III." In: *Behavioral Science* 14, pp. 159–182, 320–334, 486–502.

Harsanyi, John C. and Reinhard Selten (1988). *A General Theory of Equilibrium Selection in Games*. Cambridge, MA, USA: MIT Press.

Hart, Sergiu (2005). "Adaptive heuristics." In: *Econometrica* 73.5, pp. 1401–1430.

Hart, Sergiu and Yishay Mansour (2010). "How long to equilibrium? The communication complexity of uncoupled equilibrium procedures." In: *Games and Economic Behavior* 69.1, pp. 107–126.

Hart, Sergiu and Andreu Mas-Colell (2003). "Uncoupled Dynamics Do Not Lead to Nash Equilibrium." In: *The American Economic Review* 93.5, pp. 1830–1836.

— (2006). "Stochastic uncoupled dynamics and Nash equilibrium." In: *Games and Economic Behavior* 57.2, pp. 286–303.

Hasson, Oren (1991). "Pursuit-deterrent signals: communication between prey and predator." In: *Trends in ecology & evolution* 6.10, pp. 325–329.

Hauert, Christoph, Arne Traulsen, Hannelore Brandt, et al. (2007). "Via freedom to coercion: the emergence of costly punishment." In: *Science* 316.5833, pp. 1905–7.

Hauser, Marc D. and Peter Marler (1993). "Food-associated calls in rhesus macaques (Macaca mulatta): II. Costs and benefits of call production and suppression." In: *Behavioral Ecology* 4.3, pp. 206–212.

Hilbe, Christian and Arne Traulsen (2012). "Emergence of responsible sanctions without second order free riders, antisocial punishment or spite." In: *Scientific reports* 2, p. 458.

Hofbauer, Josef and Karl Sigmund (1998). *Evolutionary games and population dynamics*. Cambridge University Press.

Hurd, Peter L. (1997). “Is Signalling of Fighting Ability Costlier for Weaker Individuals?” In: *Journal of Theoretical Biology* 184.1, pp. 83–88.

Hurd, Peter L. and Magnus Enquist (1998). “Conventional Signalling in Aggressive Interactions: the Importance of Temporal Structure.” In: *Journal of Theoretical Biology* 192.2, pp. 197–211.

Huttegger, Simon M. (2007). “Evolution and the Explanation of Meaning.” In: *Philosophy of Science* 74.1, pp. 1–27.

Huttegger, Simon M. and Kevin J. S. Zollman (2010). “Dynamic stability and basins of attraction in the Sir Philip Sidney game.” In: *Proceedings of the Royal Society B: Biological Sciences* 277.1689, pp. 1915–1922.

Imhof, Lorens A., Drew Fudenberg, and Martin A. Nowak (2005). “Evolutionary cycles of cooperation and defection.” In: *Proceedings of the National Academy of Sciences* 102.31, pp. 10797–10800.

Johnstone, R A and A Grafen (1992a). “Error-prone signalling.” In: *Proceedings of the Royal Society B: Biological Sciences* 248.1323, pp. 229–233.

Johnstone, Rufus A. and Alan Grafen (1992b). “The Continuous Sir Philip Sidney Game: A Simple Model of Biological Signalling.” In: *Journal of Theoretical Biology* 156.2, pp. 215–234.

Jüppsche (2011). *Bee dance.* [Computer Drawing.] Retrieved on October 2, 2014 from Wikimedia Commons’s website http://commons.wikimedia.org/wiki/File:Bee_dance.svg.

Kreps, David M. and Robert Wilson (1982). “Sequential Equilibria.” In: *Econometrica* 50.4, pp. 863–894.

Lachmann, Michael and Carl T. Bergstrom (1998). “Signalling among Relatives II. Beyond the Tower of Babel.” In: *Theoretical Population Biology* 54, pp. 146–160.

Lachmann, Michael, Szabolcs Számadó, and Carl T. Bergstrom (2001). “Cost and conflict in animal signals and human language.” In: *Proceedings of the National Academy of Sciences of the United States of America* 98.23, pp. 13189–13194.

Lewis, David K. (1969). *Convention: A Philosophical Study.* Cambridge, MA, USA: Harvard University Press.

Mangasarian, O. L. (1964). “Equilibrium Points of Bimatrix Games.” In: *Journal of the Society for Industrial and Applied Mathematics* 12.4, pp. 778–780.

Martinez, Yailen, Bert Van Vreckem, and David Catteeuw (2009). “Multi-Stage Scheduling Problem with Parallel Machines.” In: *Book of Abstracts of the 14th Belgian-French-German Conference on Optimization.* Leuven, Belgium: Katholieke Universiteit Leuven, p. 162.

Martinez, Yailen, Bert Van Vreckem, David Catteeuw, and Ann Nowé (2010). "Application of Learning Automata for Stochastic Online Scheduling." In: *Recent Advances in Optimization and its Applications in Engineering, Postproceedings of the 14th Belgian-French-German Conference on Optimization.* Ed. by Moritz Diehl, Francois Glineur, Elias Jarlebring, and Wim Michiels. Springer-Verlag, pp. 491–498.

Maynard Smith, John (1982). *Evolution and the Theory of Games.* Oxford, UK: Cambridge University Press.

— (1991). "Honest signalling: The Philip Sidney game." In: *Animal Behaviour* 42, pp. 1034–1035.

Maynard Smith, John and David Harper (2003). *Animal signals.* Oxford, UK: Oxford University Press.

Maynard Smith, John and George R. Price (1973). "The Logic of Animal Conflict." In: *Nature* 246.5427, pp. 15–18.

McKelvey, Richard D., Andrew M. McLennan, and Theodore L. Turocy (2014). *Gambit: Software Tools for Game Theory.*

Meyn, S.P. and R.L. Tweedie (1993). *Markov chains and stochastic stability.* 2005th ed. Londen: Springer-Verlag.

Mihaylov, Mihail (2012). "Decentralized Coordination in Multi-Agent Systems." PhD Thesis. Brussels, Belgium: Vrije Universiteit Brussel.

Moller, Anders Pape (1987). "Social control of deception among signalling house sparrows Passer domesticus." In: *Behavioral Ecology and Sociobiology* 20.5, pp. 307–311.

Moran, Patrick A. P. (1958). "Random processes in genetics." English. In: *Mathematical Proceedings of the Cambridge Philosophical Society* 54.01, pp. 60–71.

Nadler, R.D. and L.C. Miller (1982). "Influence of Male Aggression on Mating of Gorillas in the Laboratory." In: *Folia Primatologica* 38.3-4, pp. 233–239.

Narendra, Kumpati S and Mandayam A L Thathachar (1974). "Learning Automata-A Survey." In: *IEEE Transactions on Systems, Man, and Cybernetics* 4.4, pp. 323–334.

— (1989). *Learning automata: an introduction.* Upper Saddle River, NJ, USA: Prentice-Hall.

Nash, John Forbes (1950a). "Equilibrium Points in n-Person Games." In: *Proceedings of the National Academy of Sciences of the United States of America* 36.1, pp. 48–49.

— (1950b). "Non-Cooperative Games." PhD thesis. Princeton, New Jersey, USA: Princeton University.

Neumann, John von and Oskar Morgenstern (1944). *Theory of Games and Economic Behavior*. Princeton University Press.

Nowak, Martin A. and Karl Sigmund (2004). "Evolutionary dynamics of biological games." In: *Science* 303.5659, pp. 793–799.

Nowak, Martin A., Akira Sasaki, Christine Taylor, and Drew Fudenberg (2004). "Emergence of cooperation and evolutionary stability in finite populations." In: *Nature* 428.4, pp. 646–650.

Ohta, Tomoko (2002). "Near-neutrality in evolution of genes and gene regulation." In: *Proceedings of the National Academy of Sciences of the United States of America* 99.25, pp. 16134–7.

Pentland, Alex (2010). *Honest Signals: How They Shape Our World*. MIT Press.

Petrie, M., T. Halliday, and C. Sanders (1991). "Peahens prefer peacocks with elaborate trains." In: *Animal behaviour* 41, pp. 323–331.

Rand, David G. and Martin A. Nowak (2011). "The evolution of antisocial punishment in optional public goods games." In: *Nature communications* 2, p. 434.

Recreational Scuba Training Council (2005). *Common Hand Signals for Recreational Scuba Diving*.

Riley, J. R., U. Greggers, A. D. Smith, et al. (2005). "The flight paths of honeybees recruited by the waggle dance." In: *Nature* 435.7039, pp. 205–207.

Riley, John G. (2001). "Silver signals: Twenty-five years of screening and signaling." In: *Journal of Economic Literature* 39.2, pp. 432–478.

Roth, Alvin E. and Ido Erev (1995). "Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term." In: *Games and Economic Behavior* 8.1, pp. 164–212.

Schelling, Thomas (1960). *The Strategy of Conflict*. First. Cambridge, MA, USA: Harvard University Press.

Seidenfeld, Teddy (1994). "When Normal and Extensive Form Decisions Differ." In: *Logic, Methodology and Philosophy of Science IX*. Ed. by D. Prawitz, Brian Skyrms, and D. Westerstahl. Uppsala, Sweden: Elsevier Science, pp. 451–463.

Selten, Reinhard (1965). "Spieltheoretische Behandlung eines Oligopolmodells mit Nachfrageträgheit: Teil I: Bestimmung des Dynamischen Preisgleichgewichts." In: *Zeitschrift für die gesamte Staatswissenschaft* 121.2, pp. 301–324.

— (1980). "A note on evolutionarily stable strategies in asymmetric animal conflicts." In: *Journal of Theoretical Biology* 84.1, pp. 93–101.

Servin, Arturo Lev and Daniel Kudenko (2008). "Multi-Agent Reinforcement Learning for Intrusion Detection: A Case Study and Evaluation." In: *Lecture Notes in Artificial Intelligence, 6th German Conference on Multi-Agent System Technologies* 5244, pp. 159–170.

Seyfarth, Robert M., Dorothy L. Cheney, and Peter Marler (1980). "Vervet monkey alarm calls: Semantic communication in a free-ranging primate." In: *Animal Behaviour* 28.4, pp. 1070–1094.

Shannon, Claude E. (1948). "A Mathematical Theory of Communication." In: *Bell System Technical Journal* 27, pp. 379–423, 623–656.

Shapley, Lloyd S. (1974). "A note on the Lemke-Howson algorithm." In: *Mathematical Programming Studies* 1, pp. 175–189.

Sigmund, Karl (2010). *The Calculus of Selfishness.* Princeton, NJ, USA: Princeton University Press.

Sigmund, Karl, Hannelore De Silva, Arne Traulsen, and Christoph Hauert (2010). "Social learning promotes institutions for governing the commons." In: *Nature* 466.7308, pp. 861–3.

Skyrms, Brian (2010). *Signaling: Evolution, Learning and Information.* New York, NY, USA: Oxford University Press.

Smuts, Barbara B. and Robert W. Smuts (1993). "Male Aggression and Sexual Coercion of Females in Nonhuman Primates and Other Mammals: Evidence and Theoretical Implications." In: *Advances in the Study of Behavior* 22, pp. 1–63.

Sobel, Joel (2009). *Signaling Games.* Ed. by Robert A. Meyers.

Spence, Michael (1973). "Job Market Signaling." In: *Quarterly Journal of Economics* 87.3, pp. 355–374.

Steels, Luc (1999). *The Talking Heads Experiment.* pre-editio. Brussels, Belgium: VUB AI Lab.

— (2001). "Language Games for Autonomous Robots." In: *IEEE Intelligent Systems* 16.5, pp. 16–22.

Sutton, Richard S. and Andrew G. Barto (1998). *Reinforcement Learning: An Introduction.* Cambridge, MA: MIT Press.

Számadó, Szabolcs (2011). "The cost of honesty and the fallacy of the handicap principle." In: *Animal Behaviour* 81.1, pp. 3–10.

Tambe, Milind (1997). "Towards Flexible Teamwork." In: *Journal of Artificial Intelligence Research* 7, pp. 83–124.

Taylor, Christine, Drew Fudenberg, Akira Sasaki, and Martin A. Nowak (2004). "Evolutionary game dynamics in finite populations." In: *Bulletin of mathematical biology* 66.6, pp. 1621–1644.

Tibbetts, Elizabeth A. and Amanda Izzo (2010). "Social Punishment of Dishonest Signalers Caused by Mismatch between Signal and Behavior." In: *Current Biology* 20, pp. 1637–1640.

Tibbetts, Elizabeth A., Alex Mettler, and Stephanie Levy (2010). "Mutual assessment via visual status signals in Polistes dominulus wasps." In: *Biology letters* 6.1, pp. 10–13.

Traulsen, Arne and Christoph Hauert (2009). "Stochastic evolutionary game dynamics." In: *Reviews of Nonlinear Dynamics and Complexity, vol. 2.* Ed. by H.-G. Schuster. Wiley-VCH.

Traulsen, Arne, Martin A. Nowak, and Jorge M. Pacheco (2006). "Stochastic dynamics of invasion and fixation." In: *Physical Review E* 74.1, p. 011909.

Von Frisch, Karl (1967). *The dance language and orientation of bees.* Cambridge, MA, USA: Harvard University Press.

Wagner, Elliott O. (2009). "Communication and Structured Correlation." In: *Erkenntnis* 71.3, pp. 377–393.

Watkins, Christopher John Cornish Hellaby (1989). "Learning from Delayed Rewards." PhD Thesis. Cambridge, UK: Cambridge University.

Wheeler, R. M. Jr. and Kumpati S Narendra (1986). "Decentralized learning in finite Markov chains." In: *IEEE Transactions on Automatic Control* 31.6, pp. 519–526.

Wilson, Charles (1977). "A model of insurance markets with incomplete information." In: *Journal of Economic Theory* 16.2, pp. 167–207.

Wu, Bin, Chaitanya S. Gokhale, Long Wang, and Arne Traulsen (2012). "How small are small mutation rates?" In: *Journal of mathematical biology* 64.5, pp. 803–27.

Xu, Qing, Tony Mak, Jeff Ko, and Raja Sengupta (2004). "Vehicle-to-vehicle safety messaging in DSRC." In: *Proceedings of the first ACM workshop on Vehicular ad hoc networks.* Philadelphia, PA, USA: ACM Press, pp. 19–28.

Yang, Xue, Jie Liu, Feng Zhao, and Nitin H. Vaidya (2004). "A vehicle-to-vehicle communication protocol for cooperative collision warning." In: *International Conference on Mobile and Ubiquitous Systems: Networking and Services.* Boston, MA, USA: IEEE, pp. 114–123.

Zahavi, Amotz (1975). "Mate selection - a selection for a handicap." In: *Journal of Theoretical Biology* 53, pp. 205–214.

— (1977). "The cost of honesty (further remarks on the handicap principle)." In: *Journal of Theoretical Biology* 67.3, pp. 603–5.

Zollman, Kevin J. S. (2005). "Talking to Neighbors : The Evolution of Regional Meaning." In: *Philosophy of Science* 72.January, pp. 69–85.

— (2013). "Finding alternatives to handicap theory." In: *Biological Theory* 8.2, pp. 127–132.

Zollman, Kevin J. S. and Rory Samuel Smead (2010). “Plasticity and language: an example of the Baldwin effect?” In: *Philosophical studies* 147.1, pp. 7–21.

Zollman, Kevin J. S., Carl T. Bergstrom, and Simon M. Huttegger (2013). “Between cheap and costly signals: the evolution of partially honest communication.” In: *Proceedings of the Royal Society B: Biological Sciences* 280.1750, p. 20121878.