

# Beliefs, intentions, power : essays on dynamic decision problems

Citation for published version (APA):

Méder, Z. Z. (2015). *Beliefs, intentions, power : essays on dynamic decision problems*. [Doctoral Thesis, Maastricht University]. Maastricht University. <https://doi.org/10.26481/dis.20150605zm>

## Document status and date:

Published: 01/01/2015

## DOI:

[10.26481/dis.20150605zm](https://doi.org/10.26481/dis.20150605zm)

## Document Version:

Publisher's PDF, also known as Version of record

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

# BELIEFS, INTENTIONS, POWER

ESSAYS ON DYNAMIC DECISION PROBLEMS

DISSERTATION

TO OBTAIN THE DEGREE OF DOCTOR AT MAASTRICHT UNIVERSITY,  
ON THE AUTHORITY OF THE RECTOR MAGNIFICUS, PROF. DR. L. L. G. SOETE,  
IN ACCORDANCE WITH THE DECISION OF THE BOARD OF DEANS,  
TO BE DEFENDED IN PUBLIC ON FRIDAY, JUNE 5, 2015, AT 12:00 HOURS

BY ZSOMBOR Z. MÉDER

*SUPERVISOR*

PROF. DR. P. J. J. HERINGS

*CO-SUPERVISORS*

DR. R. PEETERS

DR. J. FLESCHE

*ASSESSMENT COMMITTEE*

PROF. DR. H. J. M. PETERS, CHAIRMAN

PROF. DR. S. GÄCHTER, UNIVERSITY OF NOTTINGHAM

PROF. DR. K. I. M. ROHDE, ERASMUS UNIVERSITY ROTTERDAM

DR. E. TSAKAS





---

# Contents

<b>Acknowledgements</b>	<b>iii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Social Text . . . . .	1
1.2 Natural Text (no one is safe) . . . . .	4
1.3 Generated Text . . . . .	7
1.4 Economic Text . . . . .	11
1.5 This Text . . . . .	12
<b>2 Optimal choice for finite and infinite horizons</b>	<b>15</b>
2.1 Introduction and motivation . . . . .	16
2.2 The problem . . . . .	17
2.3 Comparing strategies . . . . .	18
2.4 Optimal strategies . . . . .	23
2.5 Pointwise limits . . . . .	26
2.6 Existence of repeatedly optimal strategies . . . . .	29
2.7 Discussion . . . . .	31
<b>3 Naiveté and sophistication in dynamic inconsistency</b>	<b>33</b>
3.1 Introduction . . . . .	34
3.2 Related literature . . . . .	35
3.3 Basic concepts . . . . .	37
3.3.1 Markov decision problem . . . . .	38
3.3.2 History . . . . .	38
3.3.3 Conceptual foundations . . . . .	39
3.3.4 Intentions, beliefs, strategy . . . . .	40
3.3.5 Truncation . . . . .	42

---

3.3.6	Frames . . . . .	44
3.3.7	Induced strategy . . . . .	49
3.3.8	Remarks . . . . .	49
3.4	Utility and discounting . . . . .	50
3.5	Naiveté . . . . .	52
3.6	Sophistication . . . . .	55
3.7	Hybrid decision makers . . . . .	60
3.8	Concluding remarks and future research . . . . .	69
3.A	Appendix – Comparison of naifs and sophisticates . . . . .	71
3.A.1	Single tasks – procrastination and impulsiveness . . . . .	72
3.A.2	Repeated tasks – underinvestment and binges . . . . .	77
3.A.3	Indulgence . . . . .	85
<b>4</b>	<b>Centralization of Power Sustains Cooperation</b>	<b>89</b>
4.1	Introduction . . . . .	90
4.2	Methods . . . . .	91
4.2.1	The power transfer game . . . . .	91
4.2.2	Experiment . . . . .	92
4.3	Results . . . . .	93
4.4	Discussion . . . . .	99
4.A	Appendix – Supplementary materials . . . . .	102
4.A.1	Game description . . . . .	102
4.A.2	Timeline and parameter setup . . . . .	105
4.A.3	Participants . . . . .	105
4.A.4	Experimental setup . . . . .	106
4.B	Appendix – Supplementary analysis . . . . .	121
4.B.1	Measurements . . . . .	121
4.B.2	Statistical models . . . . .	124
<b>5</b>	<b>Conclusion</b>	<b>143</b>
	<b>Bibliography</b>	<b>147</b>

# Acknowledgements

Science is a highly collaborative effort, where almost nothing hinges on the individual, and almost everything depends on her network. This work is no exception.

First and foremost, I would like to thank my supervisors, János Flesch and Ronald Peeters for their openness and patience. In fact, Ronald made a risky choice when he decided to hire me back in 2010. Many would have construed as a drawback what he instead perceived to be an advantage: my background as a student of philosophy. Then, János doubled down on the risk, and agreed to join us at work! I am still grateful for their decisions, and I hope I could give them at least a comparable amount of ideas to what they have given me. But besides ideas, they have given me much more: I still have to meet the PhD student who can be said to have worked in a more inspiring, yet less stressful milieu than what they have provided me with. Their joint supervision strategy is truly exemplary: Instead of using either carrots or sticks, they motivated me through the depth and honesty of their own scientific interest. They were not only patient, but actively helpful when I started engaging in side projects, unrelated to the main thread of my research on dynamic inconsistency. I think few academics can be said to be so free of envy, so altruistically helpful as János and Ronald are.

I wish to thank my promoter, Jean-Jacques Herings for a careful readthrough of all my papers, as well as for his numerous comments, suggestions, and criticisms. I believe the academic environment that is being created for game theorists in Maastricht has no match in the world. Within this environment, I have learned most from discussions with Christian Bach, Matthew Embrey, Andrés Perea, Arkadi Predtetchinski, Arno Riedl, and Elias Tsakas.

The Netherlands Organisation for Scientific Research (NWO) is gratefully acknowledged for its financial support.

Thanks to Jörg Gross for being such a brilliant collaborator and a tough

debate opponent just as much as a great friend. I can still recall the first discussion I shared with Jörg at a barbecue party about the encoding of utility in the brain. Later, when we were thinking about hierarchies and power, I still did not expect that I, then a staunch theorist, would be soon standing in the Behavioral and Experimental Economics Laboratory of Maastricht University, greeting participants. My learning-by-doing introduction to behavioral game theory is thus entirely Jörg's merit (or fault, depending on where one stands).

Mamoru Kaneko was extremely kind to accept me as a visiting researcher at Waseda University, Tokyo. He and his colleagues – Yuki Funaki and Róbert Veszteg – were superbly gracious hosts. I am thankful to Shuige Liu and Koji Yokote for their friendship, and for our Top Secret Konnyaku Mondo Reading Group! I am sure Shuige will soon be internationally renowned as the first game theorist who is also an accomplished novelist.

The city and the University of Maastricht generates a characteristic field of some sort, a magical bubble wrap that seems to protect those inside from harm and hardship. Everyone living here has a precise notion of this warm wrap, but we are still in want of a good verbal description. I am happy to have shared Maastricht's field of fuzziness with my many friends, of whom I shall highlight the names of Réka Felleg, Christine Gutekunst, Tibor Kapy, Abhimanyu Khan, Bob Meurs, and Dénes Pálvölgyi. The board games we played, the movies we watched, the beers we drank and the arguments we had connect all parts of this thesis through a hidden filament.

I am grateful for my family for their unwavering support. It has been a long way, but they have never questioned my want of study, never pressured me into changing the course of my life. The behavior of my parents thus flies in the face of all models of selfish-maximization. In fact, they have given me the most difficult, yet most valuable gift a parent can: freedom. Without this freedom, no part of this thesis would have been possible.

My wife, Katalin Tamási, has played an uncountable number of roles both during my years as a doctoral student, as well as during the preparation of these chapters. A patient analyzer of ideas; a constant source of reinforcement and optimism; a most critical copy editor; a tough *Netrunner* opponent; an occasional cook; and the list goes on. I really hope we can soon buy a proper board gaming table for the long term!

# Chapter 1

## Introduction

**Theorem 1.** (Euler’s average discreteness hypothesis). *For any  $b \leq \delta$ , there exists an isotropic tangent space of  $\text{Ker}(\phi)$  at  $\psi$  such that we can construct an infinite sequence in  $H \otimes \mathbb{Q}_p^n$  that diverges.*

*Proof.* Suppose  $\alpha$  is a relation  $f^{-1}$  of  $\nu$  and let  $v$  be an unbounded series in  $M_n(\mathbb{Z})$ . The result follows by dévissage.  $\square$

Theorem of the day at <http://davidsd.org/theorem/>, October 24, 2014

Derrida’s perceptive reply went to the heart of classical general relativity: “The Einsteinian constant is not a constant, is not a center. It is the very concept of variability – it is, finally, the concept of the game. In other words, it is not the concept of something – of a center starting from which an observer could master the field – but the very concept of the game.”

Derrida (1993), quoted by Sokal (1996a)

### 1.1 Social Text

The reader of a doctoral thesis in economics is in a precarious situation, arguably more so than its author. The goal of the current introduction is to show the veracity of this statement. However, if the reader wishes to skip to Chapter 2 without hesitation, I will gladly provide her with the main point of this introduction: ‘Economists, beware!’.

In 1996, a professor of physics named Alan Sokal published a paper in the academic journal *Social Text*. The work, entitled *Transgressing the Boundaries: Towards a Transformative Hermeneutics of Quantum Gravity*, was a hoax; an amalgam of syntactically correct, but intentionally meaningless claims, word-plays, and quotations from the masters of postmodernist social science and philosophy. Even a cursory look at its structure might make one incredulous that such a paper could have passed through any kind of review. The merely 6000-word essay boasts 109 footnotes along with a 218-item bibliography (!), mostly monographs by well-known postmodernist writers, mathematical and physical textbooks, as well as some popular science books.

On the day the article was published, the magazine *Lingua Franca* brought out another writing (Sokal, 1996b). In this, Sokal described his scam, and denounced not only the editors of *Social Text*, but also the intellectual milieu that made it possible. In the exchange of letters that ensued in *Lingua Franca* and many other journals over the next years a number of possible interpretations of the events surfaced.

One (anonymous) editor of *Social Text* suggested that although Sokal originally did have serious intentions with his paper, he changed his mind just before its publication (Sokal, 1996c). This of course seems utterly preposterous. Others have argued that while Sokal might have wanted to discredit sociology of science, in fact, he was inadvertently contributing to already extant debates within cultural politics (Robbins, 1996). In other words, through a reversal so typical to postmodernist discourse, he became part of the very debate he sought to discredit.

It has been emphasized that at the time, *Social Text* did not practice peer review; however, its editors could have asked for an expert physicist or mathematician to check the cogency of its contents. Moreover, the editors should at the very least have addressed the points relevant to social theory, but they chose not to do so. Sokal interpreted this as a manifestation of intellectual laziness, and lamented the general decline in standards for leftist intellectuals (Sokal, 1996b). Surprisingly, the editors replied by admitting to the selection bias, claiming that they were excited by the prospect of a respectable physicist addressing social science (Robbins and Ross, 1996). It is quite ironic that a high panel of a purportedly anti-authoritarian intellectual movement do not find it shameful to relinquish scientific standards just because the text bore the seal of an NYU physics professor. However that may be, the emphasis here is on the *technical* organization of the academia.

Intellectual laziness also has a more pronounced political aspect. The manifest main point of the original article was to derive a leftist agenda from ‘post-

---

modern science' (quantum mechanics, chaos theory, differential topology, etc.):

The content and methodology of postmodern science thus provide powerful intellectual support for the progressive political project, understood in the broadest sense: the transgressing of boundaries, the breaking down of barriers, the radical democratization of all aspects of social, economic, political and cultural life.

According to Sokal, this was one of the crucial components of his trap: He injected such leftist political ideas across the writing, because he expected the editors would find them favorable and publish his essay anyway, regardless of the validity of the supporting arguments. Of course, Sokal's point rests on a post facto unverifiable counterfactual: Had the essay been imbued with right-wing, or politically neutral ideas, it had not been accepted. Whether this political ingroup-outgroup bias motivated *Social Text*, or even whether it is a general characteristic of (postmodernist) social science is impossible to tell based on this observation alone. However, there is some empirical evidence that compared to the general population, the political views of academics are strongly skewed towards the Left, which could feed such a bias (Gross and Simons, 2007).

Perhaps surprisingly, Sokal's main contention with postmodernist thought concerns the existence of observer-independent reality (Sokal, 1996b). According to him, the emphasis on power relations and the 'laboratory life' (Aronowitz, 1997) of the scientist stems from an anti-realist philosophical position. This, in turn, leads to relativism and a disbelief in objective truth. If there is no objective truth, natural science is not an approximation of reality, but a mere phantasm created by those in power in academia. Thus, the entire edifice of science is rendered structurally and epistemologically indistinguishable from pseudo-science. For instance, social scientists were much too ready to accept a dumbed-down understanding – spread partially by popular science books – of the Copenhagen interpretation of quantum mechanics. According to this interpretation, the act of observation and measurement changes reality itself; *therefore*, the notion of objective reality has been dismantled by physics. Similar non sequiturs were made from the theory of general relativity, Gödel's first incompleteness theorem, etc. For these misconceptions, Sokal partially blames physicists and mathematicians for a lack of oversight in the communication of their results towards the general public. However, social scientists were also much too ready to rely on superficial information when they used the names of Bohr, Heisenberg, and Gödel to advance their philosophical and political agendas.

## 1.2 Natural Text (no one is safe)

It might be thought that ultimately, the Sokal affair has little relevance to other branches of science. After all, postmodernism is clearly out of fashion, even in philosophy departments and art galleries (see Figure 1.1). Moreover, relativism and anti-realism did not change the philosophical stance of most scientists. Few archaeologists or psychologists would claim that the object of their study is a purely social construction; and even fewer biologists or physicists would subscribe to such a tenet. Whether or not the Sokal affair has contributed to postmodernism’s decline, the whole story is just a historic artifact,<sup>1</sup> with little impact on what or how we research today. In short, the Sokal affair is a cautionary tale about postmodernism and its bad practices – about Them, rather than Us.

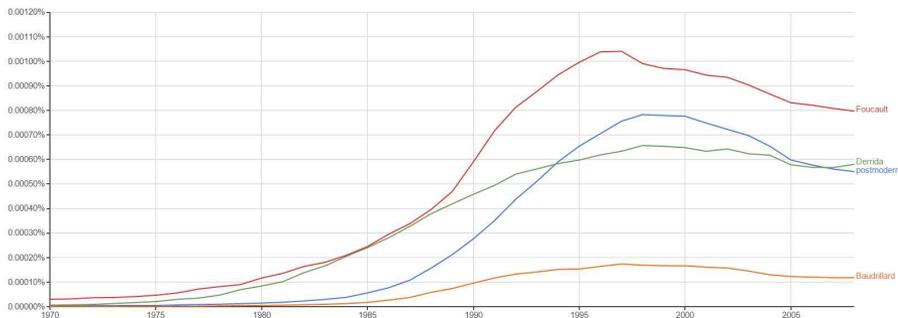


Figure 1.1: Frequencies of the word ‘postmodernism’, and the names of ‘Foucault’, ‘Derrida’ and ‘Baudrillard’ in the Google NGram corpus from 1970 to 2008.

The 20<sup>th</sup> century has a long-standing tradition of provoking the establishment through hoaxes. Various avantgardist movements – most notably, dadaists – have tried to challenge the art elite through such scams. The idea for one of the most well-known works of the century, Duchamp’s ‘Fountain’ – a readymade pissoir – came when Duchamp was thinking of ways to provoke the American cultural elite. Duchamp lamented that instead of being approached with irony and humor, art was taken far too seriously in the U.S. In a way, the idea backfired, because instead of relaxing the attitude to art, ‘Fountain’ unleashed a

<sup>1</sup> The concerned might be interested to know that the journal *Social Text* is alive and well today, and *does* practice peer review.

torrent of all-too academic discussion around the problem of the definition of art.

In another early incident, the American poets Witter Bynner and Arthur Davison Ficke, upholders of the classical form, tried to mock modernist poetry by writing intentionally bad verse. Though arguably they were quite successful with the latter: Their publisher thought the work serious, and eventually brought out *Spectra: A Book of Poetic Experiments*.

ANNE KNISH: Opus 40

I HAVE not written, reader,	I have written, reader,
That you may read...	For abstruse reasons.
They sit in rows in the bare school-room	Gold in the mine...
Reading.	Black water seeping into tunnels
Throwing rocks at windows is better,	A plank breaks, and the roof falls...
And oh the tortoise-shell cat with the can fled on!	Three men suffocated.
I would rather be a can-tier	The wife of one now works in a laundry;
Than a writer for readers.	The wife of another has married a fat man;
	I forget about the third.

Ficke (1916)

Though the artistic world might have been the birthing point of ‘Sokal hoaxes’, other domains have not been exempt from questioning their own systems of operation. As part of an experiment, Stanford researcher David Rosenhan sent five healthy participants to psychiatric wards across the United States. The participants reported (fake) hallucinations at admission, but started to behave normally afterwards. All of the participants were diagnosed as having psychiatric disorders, and were assigned treatment. However, after their symptoms ceased, their subsequent release was conditional on declaring themselves ill, and on promising to take anti-psychotic drugs in the future (Rosenhan, 1973). In the same way that Duchamp’s ‘Fountain’ emphasized the role of the gallery for art, and thus the distinction between art and non-art as essential, this study indicated that normal and abnormal behavior, psychiatric health and illness are highly context-dependent terms. Thus, the Rosenhan experiment gave a major lesson in self-doubt for professional psychiatrists.

In fact, it turned out that Sokal cannot even claim primacy for his own experiment within scientific publishing. Psychologists have been studying various

biases of academic journal editors since at least the eighties. Peters and Ceci (1982) resubmitted previously published articles to the very same psychology journals. They found that only a small share of editors detected the resubmission. Papers that proceeded to reviewers were mostly *rejected*. Mahoney (1977) and Epstein (1990) find evidence of confirmation bias on the side of researchers within behavioral psychology and social work.

However, the impact of these early experiments is hardly comparable to that of the Sokal hoax itself, which made its way out of academic debates into popular media. When Sokal and Bricmont published their book on the topic, violently attacking leading postmodernist intellectuals such as Lacan, Deleuze, and Baudrillard, the issue went viral. The book (Sokal and Bricmont, 1998) was included in the New York Times' 100 most notable books of the year, and went on to become a Boston Globe and a San Francisco Chronicle bestseller. The academic public – including, but not limited to social scientists – started to question the trustworthiness of its practices, in particular, the process of peer review. Where some saw danger, others noticed an opportunity, and before long, the ‘Sokal phenomenon’ started spreading like wildfire across all fields and branches of science.

Perhaps the weirdest Sokal-like event to date took place in the realm of physics. A French pair of twins, Igor and Grichka Bogdanov, were working towards a doctoral degree in theoretical physics at the University of Burgundy. They had trouble getting their thesis work accepted. After some struggles, Igor Bogdanov was told that if he could publish three peer-reviewed papers, he would finally be granted a Ph.D. And so, Igor and Grichka sent out their written work for publication. In 2001, the journal *Classical and Quantum Gravity* was the first to give way, accepting one of their papers (Bogdanov and Bogdanov, 2001). The twins soon became even bolder, publishing virtually identical papers in three different journals over the course of the next year; and as a result, Igor got his Ph.D. The papers started coming under scrutiny in late 2002, when some readers of the published articles felt that the journals became victims of a ‘reverse Alan Sokal hoax’ (Baez, 2002). For example, the string theorist Jacques Distler wrote: “The Bogdanov’s papers consist of buzzwords from various fields of mathematical physics, string theory and quantum gravity, strung together into syntactically correct, but semantically meaningless prose.” (Distler, 2004) The strange thing about the affair is that the Bogdanovs insisted – and, in fact, still do – that their papers were entirely serious. A 24-page CRNS committee report later determined that as the PhD thesis has no new content, and camouflages the lack of real arguments with flowery expressions, the work is void of any scientific value (Comité national du CNRS, 2003). It thus seems that

the natural science journals are in no better condition when it comes to distinguishing fake and real science than social scientists. If even the hallowed field of physics is fallible, surely every discipline is in danger!

A recent paper examined how the evaluations of scientific abstracts changes when a nonsensical mathematical formula is added to them (Eriksson, 2012). It found that especially the non-mathematically trained are prone to evaluate a text containing the mathematical gibberish more favorably. What we do not understand holds a magical attraction.

Sokal's technique has become the litmus test of the integrity of scientific fields and practices. For example, Maarten Boudry, a philosopher from Ghent University successfully submitted gibberish to two theology conferences (Coyne, 2012). For many, including Boudry himself, this points to the corruption of the field of theology. Sokal-style schemes have also been used on a larger scale, in the recent debate concerning the quality control of open access journals. Bohannon (2013) sent a fake article reporting on the cancer-cell killing property of a certain chemical to 304 open-access journals, getting 157 acceptances, 98 rejections, and 49 non-replies. While in Bohannon's eyes this indicates the unreliability of open-access journals, many have pointed out some fundamental flaws in Bohannon's experiment, including the fact that there was no control, i.e., he did not include subscription-based journals (Taylor, Wedel and Naish, 2013).

The proliferation of Sokal experiments indicates a fundamental change in the attitude to science. No longer can we, as readers of academic texts, assume that the author has serious intentions. One has to be on the lookout, even though (or, perhaps, especially because!) the field of economics has been spared of such scandals thus far.

There is one common element in the examples discussed above: These works were all, ultimately, direct products of the human mind, composed to deceive the relevant authorities. In the next subsection, we turn our attention to an even more intriguing type of hoax and research tool: works generated by computer programs.

### 1.3 Generated Text

The 17<sup>th</sup> century marked a fundamental change in the attitude towards language and representation. Previously, language was an irreducible component of the world, and the language perpetually analyzed itself as *part of the world* (Foucault, 2002). The defining genre of this worldview was commentary: commentary of the Bible, of texts of ancient philosophers, and often, commentaries

of commentaries. With the burgeoning of the natural sciences and increasing skepticism in the validity of classical texts, however, it became clear that analyzing nature requires a fundamentally new language. Such a language would allow the representation of natural objects and their relationships *as they are*, untainted by the inherent imperfections present in human communication. Most importantly, it would circumvent the variability of natural languages, inheriting instead the immutability of the laws of nature. A whole range of possible rectifications were recommended, like returning to the lost – and supposedly perfect – Adamic language, or the adoption of Chinese, where symbols of things were thought to perfectly match the things themselves. Scientists like John Wilkins developed artificial languages, complete with writing and phonological systems that tried to reflect the scientifically revealed order of things one-to-one (Méder, 2005).

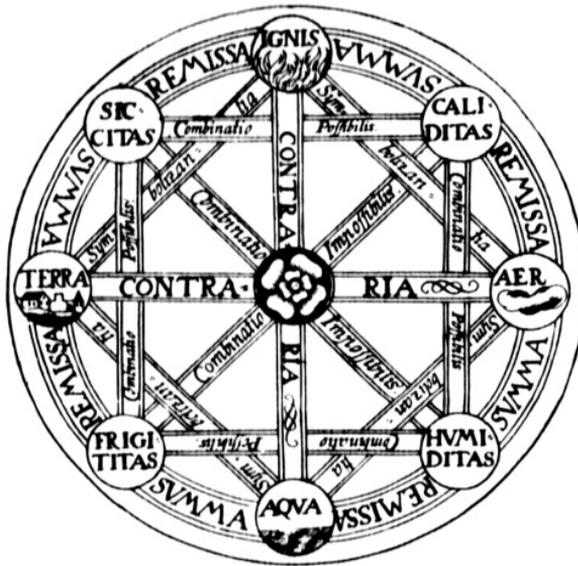


Figure 1.2: Leibniz's illustration of the symbolic *characteristica universalis*, to be used for reasoning.

Gottfried Wilhelm Leibniz recommended, though never fully worked out, a *characteristica universalis*, which was to serve as a universal language for all human thought (see Figure 1.2). This language would be processed by a *calculus ratiocinator* to produce deductive knowledge of the world through manipulating

---

words (symbols). This calculus could, in principle, be fully automatized, and thus be embodied by a mechanical calculating machine. Leibniz *constructed* such machines for algebraic calculation, which was obviously a small, though indispensable segment of this project. My goal here is of course not the exegesis of Leibniz's obscure and oft-debated writings on the subject. His revolutionary idea, one we are still struggling with today, was this: New knowledge can be generated not only by the workings of the human mind, but also in a purely *automatic* fashion, by a *machine*, through the manipulation of *symbols*.

This idea took a more intricate practical shape in the work of Charles Babbage, "the father of the computer". Babbage is perhaps most known for his ambitious project of a difference engine, a machine intended to calculate the values of polynomials. Due to financial difficulties, this machine materialized only long after his death. However, the project's ultimate goals were no less ambitious than Leibniz's, including the development of a universal symbolic language. Ada Lovelace, an avid correspondence partner of Babbage, credited as the first programmer in history, wrote:

The Analytical Engine does not occupy common ground with mere 'calculating machines.' It holds a position wholly its own. . . A new, a vast, and a powerful language is developed [...] in which to wield its truths so that these may become of more speedy and accurate practical application for the purposes of mankind than the means hitherto in our possession have rendered possible.

(Lovelace, quoted in Gleick, 2012)

Even from such a brief outlook, it is clear that the founding fathers and mothers of computing tried to automatize the search for scientific knowledge, the quest for truth. Did the advent of the digital age fulfill their dreams?

The cynical answer is that it did, though not quite in the expected way: Computers can produce text that *appears* like scientific output, but is, in fact, randomly generated by software. An example is the Postmodernism Generator,<sup>2</sup> back from the early days of the Internet (and incidentally, published online in 1996, the year of the Sokal affair). Whilst mainly an exercise in computational linguistics, the texts it creates based on a postmodernist corpus are, at first glance, surprisingly similar to the productions of many a deconstructivist. Nonetheless, its output is unlikely to pass any stage of peer review.

---

<sup>2</sup> <http://www.elsewhere.org/pomo/>.

There are more ambitious attempts to computer-generated scientific text, however. Perhaps the most successful program to date, SCIgen,<sup>3</sup> spawns research papers in computer science. SCIgen's first success in breaking academic barriers was *Rooter: A Methodology for the Typical Unification of Access Points and Redundancy*, which was invited to 2005 World Multiconference on Systemics, Cybernetics and Informatics. Other papers were accepted for publication in various journals. In line with the original intention of its programmers, SCIgen can be effectively put to use as an automated Sokal tester for computer science conferences and proceedings. A similar engine, Mathgen<sup>4</sup>, produces L<sup>A</sup>T<sub>E</sub>X source code for mathematics research papers.

Ultimately, of course, the attempt to *simulate* research through random text generation would indicate the *inability* of digital machines to produce genuine scientific output. This struggle is analogous to that of passing the infamous Turing test. Instead of creating genuine artificial intelligence, most competitors today engage in a purely statistical emulation of human textual output. But the role of computation in engendering new knowledge is not limited to mere imitation. *Automatic reasoning*, a branch of computer science, has achieved considerable success in automated proof checking, and, more importantly for our argument, automated theorem proving. Its most publicized result is the four color theorem, which states that every planar map divided into a finite number of contiguous regions can be colored with four colors in a way such that there will be no two regions having the same color, and sharing a common line boundary. The theorem was proved in 1976 through a task-specific computer program (Appel and Haken, 1980). Whether the output of theorem-proving software, output that has not (or, cannot) be checked by hand by a human should be considered a valid proof is still somewhat contentious (Mackenzie, 2004).

Still, the four color conjecture was formulated by humans, only the proof was left to the program. However, in the foreseeable future, improvements in automated theorem proving can lead to a point where every step of (formal) research is taken by software. Automatization could push not only manual workers, but even mathematicians out of jobs. There is already some indication that computers are taking over areas of intellectual labor that was previously deemed impossible, like law or architecture (Meltzer, 2014). Perhaps theoretical economists should be more wary of potential *genuine* automatically generated results in their field.

---

<sup>3</sup> <http://pdos.csail.mit.edu/scigen/>.

<sup>4</sup> <http://thatismathematics.com/mathgen/>.

---

## 1.4 Economic Text

Put in logical terms, the boundary between syntax and semantics has become increasingly obscure in science for two reasons. First, academic pranksters attempt to publish fake papers to break down the distinction between real and sham science, or to test the ‘quality assurance’ of science, that is, the peer review process. Second, the job of thinking might be taken over by artificial intelligence in the long term, a process that has begun with the proving and generating of mathematical theorems by sophisticated software.

What does all this imply for economics in particular? Specifically, what is the social role of a researcher in decision and game theory, such as the author of this academic thesis is attempting to become? Naturally, these questions cannot be answered within the confines of an introduction. In fact, it is unclear what a *satisfactory* answer would constitute at all.

Consider the original aim of game theory, clearly articulated by Thomas Schelling:

A new kind of inquiry that gave promise, fifteen years ago, of leading to such a theory of strategy is *game theory*. [ ... ] But in international strategy, the promise of game theory is so far unfulfilled. Game theory has been extremely helpful in the formulation of problems, and the clarification of concepts, but its greatest successes have been in other fields. It has, on the whole, been pitched at a level of abstraction where it has made little contact with the elements of a problem like deterrence.

Schelling (1960)

Fifty-five years later, the situation has hardly changed, if indeed at all. Game theory was spectacularly successful in some domains (e.g., auctions, matching), but still has little to say on the strategy of (non-zero-sum) conflicts.

As of yet, the fundamental story of game theory is thus a story of a promise unfulfilled. Should this observation bother us? Should the decline in the accessibility of theoretical results be a matter of concern? After all, it could be argued that technical complexity, a rigid, unyielding language, and a proliferation of buzzwords were the factors that made the Sokal hoax possible in the first place. I believe not. In my view, the community of researchers has – as it should have – complete authority over the direction of research, the accepted methodology, sophistication of the terminology, and the evaluation of individuals and their work. In short, I subscribe to the oft-cited methodological anarchism of Paul Feyerabend: “anything goes” (Feyerabend, 1993).

This methodological anarchism admits that there are no *a priori* ways to make a distinction between syntax and semantics, meaningless and meaningful, fake science and real science. Nor can we devise any failproof institutional mechanism that protects us from tricks and traps. The simple reason is that the scientific method cannot be formalized. In short, by looking at a text *formally*, we can evaluate whether the text obeys the laws of grammar, we can at most check whether mathematical derivations are correct; but we cannot evaluate it as a piece of science. Meaning is in the eyes of the beholder. This is why economists should not be afraid of tricksters flooding economics journals, or artificial intelligence taking over economists' jobs. As meaning is generated through the scientist's use of concepts – the scientific language game –, it is impossible for hoaxes to succeed for extended periods of time. It is equally impossible that automated systems of formal theorem-proving machines would substitute humans, as the connection of these formal theorems to the social world will always require a human interpreter. Thus, while nobody is safe on the short run – the social role of the economist (decision theorist, game theorist) is guaranteed for the long run.

## 1.5 This Text

At this point, a clarification seems warranted: The text presented in the following chapters (or, for that matter, in this introduction) was not produced by either a text- or a theorem-generating algorithm. It is not meant to be ironic. In fact, a deliberate attempt was made by the author to limit his unserious and half-serious thoughts by letting them expand up to this section, but no further. Thus, the role of this thesis is not to exhibit some inherent flaw in the design of graduate education of economics. The reader can thus take a relaxing breath here, and put his academic hat back on before diving in.

It would be natural to try to present the chapters of this thesis as part of one overarching project. In fact, the title 'Beliefs, intentions, power' seems to suggest that there is one such grand narrative. However, any such *ex post* construction would do injustice to the actual genealogy of these works. Instead, I must admit to the incidental nature of their connections. If one wishes to see such an overarching trend or tendency across them, then they can be said to move from the abstract to the concrete, from general considerations on optimality (Chapter 2), through decision makers facing their own fallible nature (Chapter 3), to the behavior of actual human beings in a social context (Chapter 4).

Chapter 2 investigates notions of limit optimality. It is partly inspired by the

---

works of Frank Ramsey and Weizsäcker (Ramsey, 1928; von Weizsäcker, 1965). However, while they mostly focus on intergenerational distributive decisions, i.e., optimal consumption paths over a succession of generations, we focus on a single individual decision maker facing a single problem in discrete time. The problem is modeled as a Markov decision problem, so that the set of states and the set of possible actions in each state are finite. Preferences over future payoff streams are represented by a utility function. The standard economic problem is identifying the decision maker's optimal strategy. In general, this identification requires the evaluation of infinite streams of payoff. If a decision maker is able to make these evaluations, we say that her decision horizon is *infinite*. However, not everyone is able or willing to execute the occasionally involved calculations required by the infinite horizon. Instead, it might be the case that payoffs after a certain point in time are disregarded. In this case, we say that the horizon of the decision maker is *finite*. We are concerned with strategies that are optimal not on a *particular* finite horizon, but on an infinite sequence of finite horizons, say, horizons of length 2, 4, 6, . . . The chapter examines the relationship between sets of strategies that are optimal on the infinite horizon, and those that are optimal on an infinite number of finite horizons. It turns out that the latter set is included in the former, but not vice-versa. Our main result thus provides a refinement notions for infinite-horizon optimality. There are at least two reasons why strategies that are optimal according to such refinements are interesting and/or should be preferred. First, if the game stops suddenly, due to an external reason not included in the model – an ‘act of God’, so to speak –, then the decision maker can be strictly better off. Second, it might be easier to find such a strategy, because one can start finding optimal strategies by backward induction on a finite horizon, and then expand this horizon step by step.

The subsequent chapter is an extensive analysis of dynamic inconsistency. Dynamic inconsistency describes situations in which preferences change in such a way that optimal strategies formed at various points in time prescribe different choices for the future. Such inconsistencies are empirically well documented (Frederick, Loewenstein and O'Donoghue, 2002). Previous research has established that there are two fundamentally different types of behavior in the face of preference change over time: naiveté and sophistication. Naifs do not consider their changing preferences, while sophisticates give a best response to the behavior of their future selves. Again, we model these types as a discrete-time Markov decision problem. The chapter makes an important distinction between the intentions and beliefs of agents about future behavior, and it provides definitions and optimality for naiveté and sophistication using this distinction. We

can compare naifs and sophisticates directly in a number of scenarios. Is sophistication always preferable? Does it bring advantages in situations that involve temptations? Can it help in avoiding procrastination? Finally, it is known that sexual arousal or alcohol use might induce people to behave more myopically and naively (Loewenstein, 1996). Thus, we might want to model some situations as involving changes in type. Our general framework enables us to define a new hybrid type of decision maker, who is naive in some states, and sophisticated in others, with transitions depending stochastically on the current state, and decisions taken therein.

Chapter 4 moves from empirically-inspired theory to an experimental context, and also from studying single decision makers to a multi-player environment. We address the fundamental question of how cooperation is maintained, given that it carries the danger of being exploited by free riding. In this, the chapter follows a long line of behavioral economists and evolutionary game theorists who studied this problem through the prisoner's dilemma and its generalization, the public goods game. In such settings, it has been shown that peer punishment can sustain cooperation (Fehr and Gächter, 2000, 2002). However, it can only function effectively if the impact of punishment is sufficiently large compared to its cost. Though many experiments assume efficient punishment options to be readily available, we were interested in its emergence from a 'state of nature', having equal and decentralized punishment. Moreover, the problem of second-order free riding is still not solved theoretically. Since punishment is assumed to be costly, in the subgame-perfect equilibrium of a finite public goods game with punishment, no one should engage in it. To see whether these limitations could be overcome, we designed a novel economic game, the 'power transfer game'. In the power transfer game, players are confronted with a social dilemma, and can engage in costly peer punishment. In addition, our setup enables players to give up and transfer their punishment power to others. The chapter presents the results of this behavioral experiment.

Overall, this thesis is an attempt to showing – instead of proving – that the games we theorists play are worth it.

## Chapter 2

# Optimal choice for finite and infinite horizons

This chapter lays down conceptual groundwork for optimal choice of a decision maker facing a finite-state Markov decision problem on an infinite horizon. We distinguish two notions of a strategy being favored on the limit of horizons, and examine the properties of the emerging binary relations. After delimiting two senses of optimality, we define a battery of optimal strategy sets – including the Ramsey-Weizäcker overtaking criterion – and analyze their relationships and existence properties. We also relate to the work on pointwise limits of strategies by Fudenberg and Levine (1983).

## 2.1 Introduction and motivation

Some decision problems are clearly limited in time: they only involve decisions and payoffs until a certain temporal point, having no relevant options and effects for the decision maker beyond that point. At one extreme, think of choosing the flavor of (a scoop of) ice cream: the decision is blatantly one-off, and any payoffs from choosing vanilla or chocolate are gained immediately. Other decisions, however, are potentially infinite, in the sense that utility-changing decisions might have to be made after arbitrary long times. Here, one can think of choosing the palette of flavors for an ice-cream company: Managers come and go, today's kids grow old, but until the company is up and running, the flavor palette will need to be reconsidered. Such decisions will have both immediate and long-term effects.

The horizon of a decision problem is the temporal distance between the time when the decision maker is confronted with the problem, and the furthest horizon that is payoff-relevant. This chapter starts with an infinite-horizon decision problem. However, even when the horizon is infinite, a decision-maker might still behave as if it were finite. There are various possible reasons for the decision maker acting in such a way. Maybe she (wrongly) believes that the decision problem is actually limited. It could also happen that, due to cognitive limitations, she is unable or unwilling to calculate with any utility that she receives too far in the future. In our example, a manager might care only about the short-run profits of the ice-cream company, and postpone R&D expenditures.

Short-sighted attitudes amount to cutting the decision tree so that it becomes finite. One alternative is to take into consideration the entire decision tree, including effects arbitrarily far into the future. Alternatively, if more cognitive resources, right beliefs, or deep commitments are available, the full decision problem can be tackled. Suppose now we let the horizon of the decision-maker grow ad infinitum, i.e., we cut off smaller and smaller parts of the tree. Intuition would suggest that decision strategies that are optimal on the full tree can always be approximated by strategies that are optimal for infinitely many such cuts. However, this is not always the case, and strategies that are optimal on the complete horizon can be suboptimal for *all* finite truncations of the horizon. A third and distinct optimality notion emerges by taking pointwise limits of strategies optimal on finite horizons.

In investigating the limit properties of decision strategies, the present work provides a systematic approach towards formulating and comparing optimality criteria such as the Ramsey-Weizsäcker overtaking criterion (von Weizsäcker,

1965) or the limit-equilibrium of Fudenberg and Levine (1983).

Our results provide grounds for optimality refinements. As we will see, it is possible that a strategy, despite being optimal on the infinite horizon, can be beaten by another strategy on *any* finite horizon. This means that up to any period, the decision maker would be strictly better off in expectation with the other strategy. In such cases, there might be good intuitive reasons to go with one strategy rather than the other, albeit they induce the same expected payoff on the complete horizon.

The next section outlines our model and notation. In Section 2.3, we define some relations between strategies, delimiting the strong and weak senses of taking limits of finite horizons. Section 2.4 distinguishes two senses of optimality, and analyzes the resulting strategy sets. The limit of strategies approach of Fudenberg and Levine (1983) is considered in Section 2.5. Then, we consider an approach to prove the nonemptiness of one of our optimality refinements, while the last section provides some concluding remarks.

## 2.2 The problem

Our decision maker faces a finite Markov decision problem.

**Definition 1.** *A finite Markov decision problem is given by:*

- the set of time periods  $\{1, 2, \dots\}$ ;
- a finite set of states  $\Omega$ , with  $\omega_1 \in \Omega$  as the initial state;
- a finite and nonempty set of pure actions  $A_\omega$  that the decision maker can choose from in state  $\omega$ ;
- a payoff function  $u_\omega : A_\omega \rightarrow \mathbb{R}$  that assigns a payoff to every action in state  $\omega$ ;
- transition probabilities  $m_\omega : A_\omega \rightarrow \Delta(\Omega)$ , with  $m_\omega(\omega' | a_\omega)$  denoting the probability to transit from state  $\omega$  to state  $\omega'$  when action  $a_\omega$  is chosen.

In every period, the decision-maker chooses an action from those available to her. We call a path of states and actions that the decision maker can go through a ‘history.’

**Definition 2.** A history  $h$  has the form  $h = (\omega_1, a_{\omega_1}, \dots, \omega_{t-1}, a_{\omega_{t-1}}, \omega_t)$ , with:

- $\omega_i \in \Omega$ , for  $i \in \{1, \dots, t\}$ ;

- $a_{\omega_i} \in A_{\omega_i}$ , for  $i \in \{1, \dots, t-1\}$ ;
- $m_{\omega_{i-1}}(\omega_i | a_{\omega_{i-1}}) > 0$ , for  $i \in \{1, \dots, t-1\}$ .

The length of  $h$  or current time at  $h$  is denoted by  $t = t(h)$ . In a similar vein, the function  $\omega(h) = \omega_{t(h)}$  indicates the current or end state at history  $h$ . We use  $H$  to refer to the set of all histories.

Obviously, the well-being of the decision-maker depends not only on her current choice, but also on her future actions. Thus, the decision maker needs a decision rule that tells her what to do at any history. Such a decision rule is called a strategy.

**Definition 3.** A strategy of the decision maker is a function  $s$  that assigns to every history  $h$  an action  $s(h) \in A_{\omega(h)}$ . The set of all strategies is denoted by  $S$ .

We suppose that discounting is exponential, so the expected utility induced by a strategy  $s$  on horizon  $T \in \mathbb{N}$  is:

$$U_T(s) = \sum_{h | t(h) \leq T} \delta^{t(h)-1} u_{\omega(h)}(s(h)) P_s(h),$$

where  $u_{\omega(h)}(s(h))$  is the payoff received from taking the action  $s(h)$  at history  $h$ , while  $P_s(h)$  is the probability that history  $h$  occurs when the decision maker chooses strategy  $s$ . The horizon  $T$  of this function is interpreted as the period beyond which the decision maker ignores payoffs. If the decision maker considers all future payoffs, however far they might be, we say that her horizon is infinite. The utility induced by  $s$  on the infinite horizon is  $U_\infty(s) = \lim_{T \rightarrow \infty} U_T(s)$ .

## 2.3 Comparing strategies

We start by defining relations to compare two strategies within  $S$ . First, suppose the decision maker has a finite horizon.

**Definition 4.** A strategy  $s$  is (strictly) favored over another strategy  $s'$  on a finite horizon  $T$  if it induces a (strictly) higher utility on that horizon. More precisely, we write  $s \succcurlyeq_T s'$  if  $U_T(s) \geq U_T(s')$ . Similarly, we write  $s \succ_T s'$  whenever  $U_T(s) > U_T(s')$ .

Next, it is possible that the decision maker considers the complete horizon of infinite length of the decision problem.

**Definition 5.** A strategy  $s$  is (strictly) favored over another strategy  $s'$  on the complete horizon if it induces a (strictly) higher utility on that horizon. More precisely, we write  $s \succ_{\text{CH}} s'$  if  $U_{\infty}(s) \geq U_{\infty}(s')$ . Similarly, we write  $s \succ_{\text{CH}} s'$  whenever  $U_{\infty}(s) > U_{\infty}(s')$ .

Another alternative is to conceive of the infinite horizon as the limit of finite horizons, as the parameter  $T$  that determines the length of the horizon goes to infinity. This option involves comparing strategies by their induced expected utility for large enough  $T$ . There are, however, at least two ways for looking at this limiting behavior, in a stronger and in a weaker sense.

**Definition 6.** A strategy  $s$  is (strictly) favored over another strategy  $s'$  on the limit of finite horizons in the strong sense (LHS) if it induces a (strictly) higher utility for all horizons  $T$  beyond a certain horizon  $T'$ . More precisely, we write  $s \succ_{\text{LHS}} s'$  if there is a  $T'$  such that  $s \succ_T s'$  for all  $T \geq T'$ . Similarly, we write  $s \succ_{\text{LHS}} s'$  whenever there is a  $T'$  such that  $s \succ_T s'$  for all  $T \geq T'$ .

**Definition 7.** A strategy  $s$  is (strictly) favored over another strategy  $s'$  on the limit of finite horizons in the weak sense (LHW) if, for any horizon  $T'$ , we can find a longer horizon  $T$  such that  $s$  induces a (strictly) higher utility on  $T$  than  $s'$ . More precisely, we write  $s \succ_{\text{LHW}} s'$  if for all  $T'$  there is a  $T \geq T'$  such that  $s \succ_T s'$ . Similarly, we write  $s \succ_{\text{LHW}} s'$  whenever for all  $T'$  there is a  $T \geq T'$  such that  $s \succ_T s'$ .

Note that the latter definition is equivalent to saying that there are infinitely many horizons on which the first strategy induces a (strictly) higher utility.

Following Ramsey (1928), von Weizsäcker (1965) investigates both  $\succ_{\text{LHS}}$  and  $\succ_{\text{LHW}}$ , although in a different setting than ours. The optimal growth literature adopts the stronger version as the Ramsey-Weizsäcker overtaking criterion (Gale, 1967; Brock, 1970; Brock and Haurie, 1976). Within a game-theoretic framework, Rubinstein (1979) considers an even stronger version of this relation.

**Theorem 2.** *The relations  $\succ_T$ ,  $\succ_{\text{CH}}$ ,  $\succ_{\text{LHS}}$ ,  $\succ_{\text{LHW}}$  and their strict versions satisfy the properties given in Table 2.1.*

*Proof.*

**The relations  $\succ_T$ ,  $\succ_{\text{CH}}$ ,  $\succ_T$  and  $\succ_{\text{CH}}$ .**

The properties of  $\succ_T$  and  $\succ_T$  simply inherit the properties of the non-strict and strict orders  $\geq$  and  $>$  on  $\mathbb{R}$ , since they are equivalent to single utility comparisons. Similarly,  $\succ_{\text{CH}}$  and  $\succ_{\text{CH}}$  inherit the properties of  $\geq$  and  $>$  on  $\mathbb{R}$ .

**The relation  $\succ_{\text{LHS}}$ .**

	$\succsim$			$\succ$		
	T/CH	LHS	LHW	T/CH	LHS	LHW
Total	+	-	+	-	-	-
Reflexive	+	+	+	-	-	-
Irreflexive	-	-	-	+	+	+
Symmetric	-	-	-	-	-	-
Asymmetric	-	-	-	+	+	-
Antisymmetric	-	-	-	+	+	-
Transitive	+	+	-	+	+	-

‘+’ indicates that the property necessarily holds.

‘-’ indicates that the property does not hold in general.

A relation  $R$  is:

*total* if  $xRy$  or  $yRx$  for all  $x, y$ ;

*reflexive* if  $xRx$  for all  $x$ ;

*irreflexive* if  $xRx$  for no  $x$ ;

*symmetric* if  $xRy$  implies  $yRx$ ;

*asymmetric* if  $xRy$  implies not  $yRx$ ;

*antisymmetric* if  $xRy$  and  $yRx$  imply  $x = y$ ;

*transitive* if  $xRy$  and  $yRz$  imply  $xRz$ .

Table 2.1: Properties of the relations.

To verify that  $\succsim_{\text{LHS}}$  is not total, consider the decision problem in Figure 2.1 with  $\odot$  as the initial state. The two options of the decision maker are going west (strategy  $s$ ) and going east ( $s'$ ). Suppose the discount factor is  $\delta = 0.5$ . We see that

$$U_{\text{T}}(s) = \begin{cases} 0 & \text{if } T = 2k + 1 \\ -(0.5)^T & \text{if } T = 2k \end{cases}$$

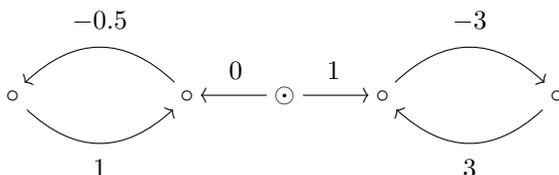
and

$$U_{\text{T}}(s') = \begin{cases} 0.5^{T-1} & \text{if } T = 2k + 1 \\ -(0.5)^{T-1} & \text{if } T = 2k. \end{cases}$$

This means that whenever  $T = 2k$ , then  $s \succ_{\text{T}} s'$ , but whenever  $T = 2k + 1$ , then  $s' \succ_{\text{T}} s$ . It follows that neither  $s \succsim_{\text{LHS}} s'$ , nor  $s' \succsim_{\text{LHS}} s$ , therefore  $\succsim_{\text{LHS}}$  is not total.

The relation  $\succsim_{\text{LHS}}$  is reflexive (and not irreflexive), as any strategy is weakly better than itself on all horizons.

We turn to symmetry:  $\succsim_{\text{LHS}}$  is not symmetric, asymmetric or antisymmetric. Lack of symmetry follows trivially from cases when one strategy generates strictly higher utility on all finite horizons. Due to its reflexivity, it is not asym-

Figure 2.1: The relation  $\succ_{\text{LHS}}$  is not total.

metric. To see that it is not antisymmetric either, take two different strategies  $s \neq s'$  that induce the same expected utility on all horizons. For these, we have  $s \succ_{\text{LHS}} s'$  and  $s' \succ_{\text{LHS}} s$  with  $s \neq s'$ . Thus  $\succ_{\text{LHS}}$  is not antisymmetric.

To show that  $\succ_{\text{LHS}}$  is transitive, suppose  $s \succ_{\text{LHS}} s'$  and  $s' \succ_{\text{LHS}} s''$ . This means that there is a  $T'$  such that  $s \succ_T s'$  for all  $T \geq T'$ , and that there is a  $T''$  such that  $s' \succ_T s''$  for all  $T \geq T''$ . Let  $T''' = \max\{T', T''\}$ . Consequently,  $s \succ_T s''$  whenever  $T \geq T'''$ . It follows that  $s \succ_{\text{LHS}} s''$ .

#### The relation $\succ_{\text{LHW}}$ .

In contrast with  $\succ_{\text{LHS}}$ , the relation  $\succ_{\text{LHW}}$  is total. Choosing any  $s$  and  $s'$ , for any  $T$ , either  $s \succ_T s'$  or  $s' \succ_T s$ . But the set of possible choices for  $T$  is infinite. Therefore,  $s \succ_T s'$  for infinitely many  $T$  or  $s' \succ_T s$  for infinitely many  $T$  (or possibly, both).

Next,  $\succ_{\text{LHW}}$  is reflexive for the same reason as  $\succ_{\text{LHS}}$ , and thus not irreflexive.

The symmetry properties of  $\succ_{\text{LHW}}$  are also the same as those of  $\succ_{\text{LHS}}$ , again for the same reasons.

Figure 2.2 shows a decision problem where  $\succ_{\text{LHW}}$  is not transitive. Suppose again that  $\odot$  is the starting state and the discount factor is  $\delta = 0.5$ . The decision maker has three options: going southwest (strategy  $s$ ), south ( $s'$ ) or southeast ( $s''$ ). The payoffs for these strategies on different horizons are summarized in Table 2.2. We see that whenever  $T = 3k$ , then  $s \succ_T s'$ , so  $s \succ_{\text{LHW}} s'$ . Moreover, whenever  $T = 3k + 2$ , then  $s' \succ_T s''$ , so  $s' \succ_{\text{LHW}} s''$ . But  $U_T(s) < 0 = U_T(s'')$  for all  $T$ , so there is no  $T$  such that  $s \succ_T s''$ . It follows that  $s$  is not favored over  $s''$  according to  $\succ_{\text{LHW}}$ , therefore  $\succ_{\text{LHW}}$  is not transitive.

#### The relation $\succ_{\text{LHS}}$ .

By comparing any strategy  $s$  to itself, or two distinct strategies that generate the same payoffs on all finite horizons, we can see that  $\succ_{\text{LHS}}$  is not total.

In contrast to its non-strict version,  $\succ_{\text{LHS}}$  is irreflexive, since no strategy can strictly beat itself on any horizon. Therefore, it is also not reflexive.

The case when one strategy yields higher payoff than the other for all hori-

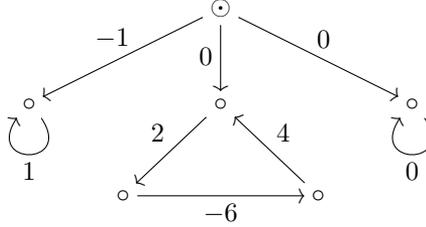


Figure 2.2: The relations  $\succ_{\text{LHW}}$  and  $\succ_{\text{LHS}}$  are not transitive.

	$U_T(s)$	$U_T(s')$	$U_T(s'')$
$T = 3k + 1$	$-0.5^{T-1}$	0	0
$T = 3k + 2$	$-0.5^{T-1}$	$0.5^{T-2}$	0
$T = 3k$	$-0.5^{T-1}$	$-0.5^{T-2}$	0

Table 2.2: Utility comparison for Figure 2.2.

zons shows that  $\succ_{\text{LHS}}$  is not symmetric. It is, however, both asymmetric and antisymmetric: If  $s \succ_{\text{LHS}} s'$ , then there exists  $T'$  such that  $s \succ_T s'$  for all  $T \geq T'$ . But then it is not possible that also a  $T''$  exists for which  $s' \succ_T s$  for all  $T \geq T''$ , since this would imply  $s \succ_T s'$  and  $s' \succ_T s$  for  $T \geq \max\{T', T''\}$ , which is impossible.

To show that  $\succ_{\text{LHS}}$  is transitive, we can use the argument for the transitivity of its non-strict counterpart, replacing non-strict inequalities with strict ones everywhere.

**The relation  $\succ_{\text{LHW}}$ .**

For similar reasons as  $\succ_{\text{LHS}}$ , the relation  $\succ_{\text{LHW}}$  is not total.

This relation is irreflexive (and not reflexive), and we can again use the same argument as for  $\succ_{\text{LHW}}$ .

The analogy also carries over for non-symmetry. However,  $\succ_{\text{LHW}}$  is not asymmetric or antisymmetric: the example in Figure 2.1 shows that it is possible that  $s \succ_T s'$  for infinitely many  $T$  and also  $s' \succ_T s$  for infinitely many  $T$ .

It is easy to see that the example on Figure 2.2 also shows the non-transitivity of  $\succ_{\text{LHW}}$ , since all of the strategies in Table 2.2 are strictly best on the respective horizons.  $\square$

## 2.4 Optimal strategies

The decision maker is looking for an optimal strategy. However, she might, for any reason, decide to restrict her choice to a nonempty subset  $G$  of the full strategy space  $S$ . For example the decision maker might restrict her attention to stationary strategies.

Given a relation  $R$  on  $S$ , a strategy in  $G$  can be said to be optimal in two senses. First, a strategy can be regarded as optimal if no further improvements can be made on it, i.e., if there is no other strategy in  $G$  that is favored to it. One can think of a former criterion for retaining the title of World Chess Champion, which required not being beaten in a title match – in case of a tie, the incumbent would remain the world champion. Formally, a strategy  $s^* \in G$  is optimal in  $G$  according to relation  $R$  in the ‘not-beaten’ (NB) sense, if there is no other strategy  $s \in G$  such that  $sRs^*$ .

Alternatively, a strategy in  $G$  can be said to be optimal if it is favored to all other strategies in  $G$ . Here, one can think of horse races, where in case of a tie bookmakers consider all the horses in the dead heat as ‘winners’. Formally, a strategy  $s^* \in G$  is optimal according to relation  $R$  in the ‘beat-all’ (BA) sense if  $s^*Rs$  for all other strategies  $s \in G$ .

The orders defined on finite and the complete horizons involve single utility comparison. Therefore, a strategy is not beaten by the strict order if and only if it is favored to every other strategy according to the weak order. Similarly, a strategy is not beaten by the weak order if and only if it is favored to every other strategy according to the strict order. Therefore, we have the following definitions for optimal strategies.

**Definition 8.** The set of strategies in  $G$  that are optimal on horizon  $T$  is given by:

$$G_T = \{s \in G_s \mid \nexists s' \in G_s, s' \succ_T s\} = \{s \in G \mid \forall s' \in G_s, s \succcurlyeq_T s'\},$$

with  $G_s = G \setminus \{s\}$  here and henceforth.

The set of strategies in  $G$  that are uniquely optimal on horizon  $T$  is given by:

$$\overline{G}_T = \{s \in G \mid \nexists s' \in G_s, s' \succcurlyeq_T s\} = \{s \in G \mid \forall s' \in G_s, s \succ_T s'\}.$$

**Definition 9.** The set of strategies in  $G$  that are optimal on the complete horizon is given by:

$$G_{CH} = \{s \in G \mid \nexists s' \in G_s, s' \succ_{CH} s\} = \{s \in G \mid \forall s' \in G_s, s \succ_{CH} s'\}.$$

The set of strategies in  $G$  that are uniquely optimal on the complete horizon is given by:

$$\overline{G}_{\text{CH}} = \{s \in G \mid \nexists s' \in G_s, s' \succ_{\text{CH}} s\} = \{s \in G \mid \forall s' \in G_s, s \succ_{\text{CH}} s'\}.$$

As its name suggests, a uniquely optimal strategy set contains at most one element.

The simple duality of the two senses of optimality fails in the case of the limit of finite horizon relations since these are not total ( $\succ_{\text{LHS}}$ ), or not transitive ( $\succ_{\text{LHW}}$ ). Thus, we have a total of eight possible definitions for the concept of ‘optimal strategy on the limit of finite horizons’:

**Definition 10.** The sets of strategies  $s$  in  $G$  that are ‘not-beaten-optimal’ or ‘beat-all-optimal’ on the limit of finite horizons in the strong or weak sense, according to the respective strict or non-strict relations, are given by the conditions in Table 2.3:

	LHS	LHW
NB, $\succ$	$\nexists s' \in G_s, s' \succ_{\text{LHS}} s$	$\nexists s' \in G_s, s' \succ_{\text{LHW}} s$
NB, $\succ$	$\nexists s' \in G_s, s' \succ_{\text{LHS}} s$	$\nexists s' \in G_s, s' \succ_{\text{LHW}} s$
BA, $\succ$	$\forall s' \in G_s, s \succ_{\text{LHS}} s'$	$\forall s' \in G_s, s \succ_{\text{LHW}} s'$
BA, $\succ$	$\forall s' \in G_s, s \succ_{\text{LHS}} s'$	$\forall s' \in G_s, s \succ_{\text{LHW}} s'$

Table 2.3: Optimality conditions.

We denote  $\{s \in G \mid \nexists s' \in G_s, s' \succ_{\text{LHS}} s\}$  by  $G_{\text{LHS}}^{\text{NB}\succ}$ . The other sets are denoted in a consistent manner.

**Theorem 3.** For any Markov decision problem and  $G \subseteq S$ , the different notions of optimality induce the following relations on the resulting sets of optimal strategies:

$$\overline{G}_{\text{CH}} \subseteq G_{\text{LHW}}^{\text{NB}\succ} = G_{\text{LHS}}^{\text{BA}\succ} \subseteq G_{\text{LHS}}^{\text{NB}\succ} = G_{\text{LHW}}^{\text{BA}\succ} \subseteq G_{\text{CH}}.$$

*Proof.*

$$\overline{G}_{\text{CH}} \subseteq G_{\text{LHS}}^{\text{BA}\succ}.$$

Take  $s \in \overline{G}_{\text{CH}}$ . Then,  $U_{\infty}(s) > U_{\infty}(s')$  for all  $s' \in G_s$ , hence there exists a

$T'(s')$  so that for all subsequent horizons  $T \geq T'(s')$ , we get  $U_T(s) > U_T(s')$ . Thus,  $s \in G_{\text{LHS}}^{\text{BA}\succ}$ .

$$\mathbf{G}_{\text{LHW}}^{\text{NB}\succ} = \mathbf{G}_{\text{LHS}}^{\text{BA}\succ}, \mathbf{G}_{\text{LHW}}^{\text{NB}\succ} = \mathbf{G}_{\text{LHS}}^{\text{BA}\succ}, \mathbf{G}_{\text{LHS}}^{\text{NB}\succ} = \mathbf{G}_{\text{LHW}}^{\text{BA}\succ}, \mathbf{G}_{\text{LHS}}^{\text{NB}\succ} = \mathbf{G}_{\text{LHW}}^{\text{BA}\succ}.$$

Strategy  $s$  is in  $G_{\text{LHW}}^{\text{NB}\succ}$  if and only if there is no  $s' \in G_s$  such that  $s' \succ_T s$  for infinitely many  $T$ . This is true if and only if  $U_T(s') \geq U_T(s)$  for only finitely many  $T$ , which is again equivalent to there being a  $T'(s')$  such that  $U_T(s) > U_T(s')$  for all  $T \geq T'(s)$ . The latter condition is the definition of  $G_{\text{LHS}}^{\text{BA}\succ}$ , so indeed  $G_{\text{LHW}}^{\text{NB}\succ} = G_{\text{LHS}}^{\text{BA}\succ}$ . Similarly, the definitions of all the other pairs of sets are logically equivalent.

$$\mathbf{G}_{\text{LHS}}^{\text{BA}\succ} \subseteq \mathbf{G}_{\text{LHS}}^{\text{BA}\succ}, \mathbf{G}_{\text{LHW}}^{\text{BA}\succ} \subseteq \mathbf{G}_{\text{LHW}}^{\text{BA}\succ}.$$

These cases are trivial, since  $s \succ_T s'$  implies  $s \succ_T s'$ .

$$\mathbf{G}_{\text{LHS}}^{\text{BA}\succ} \subseteq \mathbf{G}_{\text{LHW}}^{\text{BA}\succ}.$$

Choose any  $s \in G_{\text{LHS}}^{\text{BA}\succ}$ . Thus, for all  $s' \in G$ , there is a  $T'$  such that  $U_T(s) > U_T(s')$  for all  $T \geq T'$ . Obviously then, for all  $s'$  there are infinitely many  $T'$  for which  $U_T(s) \geq U_T(s')$ , so we have  $s \in G_{\text{LHW}}^{\text{BA}\succ}$ .

$$\mathbf{G}_{\text{LHS}}^{\text{BA}\succ} \subseteq \mathbf{G}_{\text{LHW}}^{\text{BA}\succ}.$$

This case is analogous to the previous one, by replacing  $\succ$  with  $\succ$ .

$$\mathbf{G}_{\text{LHW}}^{\text{BA}\succ} \subseteq \mathbf{G}_{\text{CH}}.$$

Fix  $s \in G_{\text{LHW}}^{\text{BA}\succ}$  and  $s' \in G$ . Since  $s \succ_{\text{LHW}} s'$  for infinitely many  $T$ ,  $U_\infty(s) = \lim_{T \rightarrow \infty} U_T(s) \geq \lim_{T \rightarrow \infty} U_T(s') = U_\infty(s')$ . Therefore,  $s \in G_{\text{CH}}$ .  $\square$

We know that whenever  $G \subseteq S$  is nonempty and closed for pointwise limits, strategies optimal on the complete horizon exist, i.e.,  $G_{\text{CH}}$  is nonempty (see Theorem 3).

The set  $G_{\text{LHS}}^{\text{NB}\succ} = G_{\text{LHW}}^{\text{BA}\succ}$  contains those strategies that are non-strictly favored to any other strategy on infinitely many horizons. For this reason, we call them ‘repeatedly’ optimal strategies, and denote the resulting set of strategies by  $G_{\text{R}}$ . We conjecture that whenever  $G$  is nonempty and closed for pointwise limits, the set of repeatedly optimal strategies is also nonempty. In Section 2.6 we deal with an attempt to prove the nonemptiness of  $G_{\text{R}}$ .

Members of the set  $G_{\text{LHW}}^{\text{NB}\succ} = G_{\text{LHS}}^{\text{BA}\succ}$  are those strategies  $s$  for which, for every other strategy  $s'$ , there is a certain horizon – depending on  $s'$  – after which  $s$  is non-strictly favored to  $s'$  for all further horizons. This is exactly the set of strategies that von Weizsäcker (1965) calls optimal. Therefore, we call these ‘overtakingly’ optimal strategies, and denote the resulting set by  $G_{\text{OT}}$ . Such

strategies are not guaranteed to exist. Indeed, for the decision problems in Figures 2.1 and 2.2, there are only two, respectively three possible strategies, those defined in Theorem 2. Since  $s, s' \notin S_U$  for the first problem, and  $s, s', s'' \notin S_U$  for the second problem, there are no overtakingly optimal strategies in these decision problems for  $G = S$ . The seminal paper of von Weizsäcker (1965) remarks in a different setting that it might be possible for a strategy to be repeatedly, but not overtakingly optimal; however, he does not provide an example.

It can be easily seen that  $G_{LHS}^{NB\>} = G_{LHW}^{BA\>}$  can be empty. Think of a decision problem where there are only two different strategies and they induce the same payoffs over all horizons. Then neither of them can be called optimal according to this optimality criterion. Due to such instances, this particular notion is less appealing to us.

One could call the set of strategies in  $G_{LHW}^{NB\>} = G_{LHS}^{BA\>}$  ‘strictly overtakingly optimal’. Thus, we can denote them by  $G_{OTS}$ . From our previous results, it is obvious that a decision problem might have no strictly overtakingly optimal strategies. Moreover, if  $G_{OTS}$  is nonempty, then it contains exactly one element.

Figure 2.3 illustrates the set inclusion relationships between the various optimality notions defined in this section.

## 2.5 Pointwise limits

We now turn to examine the relationship between the pointwise limits of strategies optimal on finite horizons examined by Fudenberg and Levine (1983) and our optimality notions.

**Definition 11.** Suppose  $\mathfrak{s} = (s_i)_{i \in \mathbb{N}}$  is an infinite sequence of strategies. We say that strategy  $s \in S$  is a pointwise limit of  $\mathfrak{s}$  if there is an index set  $I \subseteq \mathbb{N}$  of infinite cardinality so that for every history  $h$  there is an index  $m_h$  such that  $s(h) = s_i(h)$  for every  $i \in I$  with  $i \geq m_h$ .

It is easy to provide an iterative method for finding pointwise limits of a sequence of strategies. First, order all histories by length. Then, go through the histories step by step: At each history, choose any action that is taken by infinitely many strategies, and eliminate the strategies that do not take this action at that history. After each step of choice and elimination, the number of remaining strategies will be infinite. Thus the algorithm will never halt, and at least one pointwise limit will exist for any infinite sequence of strategies.

For a set of strategies  $G \subseteq S$ , denote by  $G_{FL}$  the set of all strategies  $s$  for which there exists a sequence  $\mathfrak{s} = (s_T)$  with  $s_T \in G_T$  for each  $T \in \{1, 2, \dots\}$  such

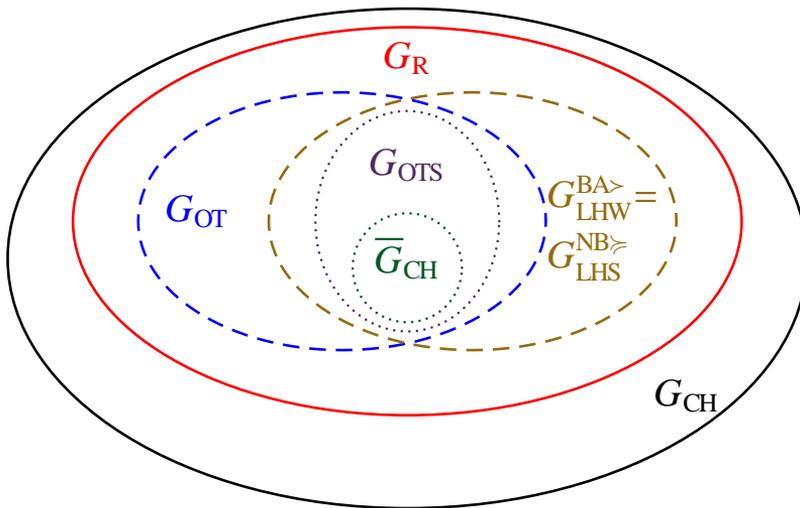


Figure 2.3: The different optimality notions.

A solid boundary indicates that the set is (conjectured to be) nonempty. A dashed or dotted boundary indicates that the set might be empty. A dotted boundary also indicates that the set contains at most one element.

that  $s$  is a pointwise limit of  $\mathfrak{s}$ . The following theorem follows from the results of Fudenberg and Levine (1983). Since they use a game-theoretic approach, and their formalism is somewhat different, we attach a short proof.

**Theorem 4.** *Suppose  $G \subseteq S$  is nonempty and closed for pointwise limits. Then,  $G_{\text{FL}}$  is a nonempty subset of  $G_{\text{CH}}$ .*

*Proof.* Since  $G$  is nonempty and closed for pointwise limits, the nonemptiness of  $G_{\text{FL}}$  follows from the discussion above on the iterative method. We will now show that  $G_{\text{FL}} \subseteq G_{\text{CH}}$ . Take an arbitrary  $s \in G_{\text{FL}}$ . Then, there exists a sequence  $(s_T)_{T \in \mathbb{N}}$ , with  $s_T \in G_T$  for each  $T \in \mathbb{N}$ , and an infinite index set  $I \subseteq \mathbb{N}$  such that  $s$  is the pointwise limit of the subsequence  $(s_T)_{T \in I}$ . Let  $s' \in G$ . Because  $s_T \in G_T$ , we have  $U_T(s_T) \geq U_T(s')$  for every  $T \in \mathbb{N}$ . Let  $R$  be the maximum of the absolute values of all payoffs, and for every  $T \in \mathbb{N}$  let

$$w_T = \sum_{t=T+1}^{\infty} \delta^{t-1} R = \frac{\delta^T}{1-\delta} R.$$

Then,  $U_T(s_T) \leq U_{T'}(s_T) + w_{T'}$  for every  $T, T' \in \mathbb{N}$  with  $T' \leq T$ , and  $U_T(s') \geq U_{\infty}(s') - w_T$  for every  $T \in \mathbb{N}$ . Consequently, for every  $T, T' \in \mathbb{N}$  with  $T' \leq T$  we have

$$U_{T'}(s_T) \geq U_T(s_T) - w_{T'} \geq U_T(s') - w_{T'} \geq U_{\infty}(s') - w_T - w_{T'}.$$

For a fixed  $T'$ , taking the limit when  $T$  tends to infinity within the set  $I$  yields

$$U_{T'}(s) \geq U_{\infty}(s') - w_{T'}.$$

Next, taking the limit when  $T' \rightarrow \infty$  yields

$$U_{\infty}(s) \geq U_{\infty}(s').$$

Because  $s'$  was arbitrary in  $G$ , we have shown that  $s \in G_{\text{CH}}$ . □

**Remark.** *A strategy generated through pointwise limits is not necessarily repeatedly optimal, i.e.,  $G_{\text{FL}}$  is not necessarily a subset of  $G_{\text{R}}$ , even if  $G$  is closed for pointwise limits.*

To see this, consider the decision problem in Figure 2.4. Suppose that  $\odot$  is the starting state and the discount factor is  $\delta = 0.5$ . Let  $g_t$  be the following strategy for  $t > 1$ : Go southeast, stay there until  $t$ , go east in that period, then stay there.

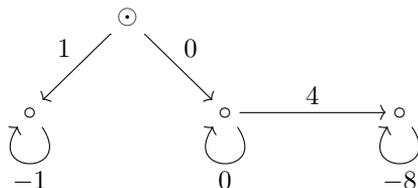


Figure 2.4: Pointwise limit of strategies optimal on finite horizons is dominated on all horizons.

First we show that  $g_t$  is the unique optimal strategy on horizon  $t$ . On this horizon, it induces the payoff sequence  $(0, \dots, 0, 4)$  and hence a utility of  $4 \cdot 0.5^{t-1}$ . There are only three types of alternative strategies to it: taking less time or more time to get the payoff of 4, or going southwest. Since the payoff of 4 is taken at the last possible moment, any strategy that takes more time will induce a utility of 0 on horizon  $t$ . On the other hand, any strategy that takes less time will end up with a utility smaller or equal to 0, because of the repeated payoff of  $-8$  on every round after taking the payoff of 4. Last, going southwest induces a utility of  $0.5^{t-1}$ . Thus  $g_t$  is uniquely optimal on horizon  $t$ .

Since  $g_t$  is uniquely optimal on horizon  $t$ , it can easily be verified that  $G_{\text{FL}}$  is a singleton set, only consisting of the strategy  $s$  of going southeast, then staying at that point forever. This strategy earns 0 on all horizons. However, we have seen that going southwest earns  $0.5^{t-1}$  on all finite horizons, therefore  $s' \succ_{\text{LHS}} s$ , and  $s$  is not repeatedly optimal.

## 2.6 Existence of repeatedly optimal strategies

Take  $G \subseteq S$  to be nonempty and closed for pointwise limits. A natural attempt to prove the existence of a repeatedly optimal strategy within  $G$  would be by using Zorn's lemma, which is equivalent to the axiom of choice. For  $s, s' \in G$ , we write  $sRs'$  if and only if either  $s = s'$ , or  $s \succ_{\text{LHS}} s'$ . It follows from Theorem 2 that relation  $R$  is a partial order on  $G$ , i.e.,  $R$  is reflexive, antisymmetric and transitive. Take a nonempty set  $G' \subseteq G$ . We call  $G'$  a chain in  $G$  if  $R$  is a total order on  $G'$ . A strategy  $s \in G$  is called an upper bound for a chain  $G'$  if  $sRs'$  for every  $s' \in G'$ . By Zorn's lemma, if every chain  $G'$  in  $G$  has an upper bound, then there exists a strategy  $s \in G$  for which there is no strategy  $s' \in G$  such that  $s' \succ_{\text{LHS}} s$ . This would mean that  $s \in G_{\text{LHS}}^{\text{NB}\succ}$ , i.e., that  $s$  is a repeatedly optimal strategy of  $G$ . So we would only have to prove that every chain  $G'$  in  $G$

has an upper bound. The following example demonstrates that this is not easy. In this example, we first construct a chain of countably infinite cardinality, and subsequently show that the unique pointwise limit of this chain is not an upper bound.

For the decision problem represented in Figure 2.5, the decision maker has the options of going east ( $E$ ) or staying in place ( $S$ ) in the initial state  $\odot$ . By choosing to go east, she first gains a payoff of 2, and she has a 50/50 chance of either returning to the initial state or moving to the absorbing state. The behavior-relevant segment of a strategy is represented by a sequence of  $E$ -s and  $S$ -s:  $(a_T)_{T \in \mathbb{N}}$ , with  $a_T \in \{E, S\}$  as the action for the decision maker if she is in state  $\odot$  in period  $T$ .

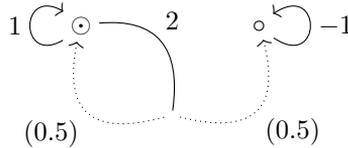


Figure 2.5: Pointwise limit of a sequence of strategies is dominated by each strategy in the sequence on all horizons.

Suppose the discount factor is  $\delta = 0.5$ . It is easy to see that any strategy is optimal on the complete horizon, earning an expected utility of 2. Moreover, it can be shown that for any finite horizon  $T$ , the number of  $E$ -s in the strategy up to and including  $T$  determine the expected utility up to  $T$ , their order being irrelevant. In general, if a strategy  $s$  contains  $k$  number of  $E$ -s up to  $T$ , then its expected utility up to  $T$  is

$$U_T(s) = 2 + \frac{1}{2^{T-1}} \left( 1 - \frac{1}{2^{k-1}} \right),$$

which is increasing in  $k$ . Take the following sequence of strategies:  $\mathbf{s} = (s_T)_{T \in \mathbb{N}}$ , where

$$s_T = \underbrace{SS \dots S}_{T-1} \underbrace{EE \dots E}_T SSSS \dots$$

Fix any horizon  $\bar{T}$ . We then have

$$s_1 \prec_T s_2 \prec_T \dots \prec_T s_{\bar{T}}$$

for all  $T$  greater or equal to  $2\bar{T} - 1$ . Therefore,

$$s_1 \prec_{\text{LHS}} s_2 \prec_{\text{LHS}} \dots \prec_{\text{LHS}} s_{\bar{T}}.$$

---

Obviously,  $\mathfrak{s}$  has a single pointwise limit, namely  $s = SSS\dots$  that prescribes staying in  $\odot$  forever. However,  $s$  is not favored to *any* strategy in the sequence on *any* horizon, since  $s \prec_T s_1$  for all  $T$ . Thus, the limit strategy, our most natural choice for an upper bound of this chain lacks the desired property, repeated optimality.

## 2.7 Discussion

First, it should be noted that our examples concern non-generic decision problems. We do not regard this as a serious drawback, as many interesting decision problems are non-generic. In the generic case, all the sets of optimal strategies defined in the previous sections collapse into a single one.

The existence of repeatedly optimal strategies remains an intriguing open question, although we conjecture that for subsets of the strategy space closed for pointwise limits, repeatedly optimal strategies do exist. If the existence of repeatedly optimal strategies can be shown, we have good reasons to use them as a refinement of optimal strategies on the complete horizon. Namely, the choice of a repeatedly optimal strategy will guarantee that for infinitely many periods, the decision maker has reason not to feel any regret over her strategy choice. If an overtakingly optimal strategy can be found, there is even more reason to rejoice: Compared with every other strategy, any regret for choosing the overtakingly optimal strategy will fully dissipate after just finitely many periods.

The assumption of exponential discounting can be relaxed for most of our results. We can use a more general, history-dependent discount function, as long as it is continuous at infinity in the sense of Fudenberg and Levine (1983).

While our model is presented in a decision-theoretic framework, the game-theoretic extension of the concepts introduced should be straightforward. Results concerning the relationships of various strategy sets will carry over, but the existence properties of various strategy sets will have to be readdressed for games.



## Chapter 3

# Naiveté and sophistication in dynamic inconsistency

This chapter introduces a general framework for dealing with dynamic inconsistency in the context of Markov decision problems. It decouples and examines concepts that are often entwined in the literature: It distinguishes between the decision maker and her various temporal agents, and between the beliefs and intentions of the agents. Classical examples of naiveté and sophistication are modeled and contrasted based on this new language. We show that naive and sophisticated decision makers can form optimal strategies at each possible history, and provide welfare comparisons for a class of decision problems including procrastination, impulsiveness, underinvestment, binges and indulgence. The creation of a unified formalism to deal with dynamic inconsistency allows for the introduction of a hybrid decision maker, who is naive sometimes, sophisticated at others. Such a hybrid decision maker can be used to model situations where type determination is endogenous. Interestingly, the analysis of hybrid types indicates that self-deception can be optimal.

### 3.1 Introduction

Imagine that you are sitting with friends, drinking beer. You have just finished your second pint, and your friends want to order another round. You think to yourself: ‘well, I could deal with one or two more, but then I really should go home.’ However, from previous experience, you are also acutely aware that after your third beer your mindset is likely to change: you will start fooling yourself, repeating over and over in the course of the evening: ‘just one more beer, and then I am *really* going home’. This would lead to an undesirable outcome, getting drunk and having a hangover the next morning. So you wisely leave your friends after just your second beer. What is happening here? The framework that we propose here makes it possible to model this and similar scenarios.<sup>5</sup>

Traditionally, there are three main ways to portray decision makers under time inconsistency. The first one regards decision makers as naifs (Akerlof, 1991; O’Donoghue and Rabin, 1999b), the second attributes sophistication to them (Laibson, 1997; Fischer, 1999; Harris and Laibson, 2001), while the third argues that even resolute behavior is possible (McClennen, 1990). A common assumption of these models in the classic papers on dynamic inconsistency is regarding the decision maker as falling entirely into one of the above three categories, treating her type as exogeneously given. More recently, mainly building on the work of O’Donoghue and Rabin (2001), hybrid decision makers have been considered, in models of so-called ‘partial naiveté’. However, these models still treat this type as exogeneous to the decision problem.

One way to interpret our example above is that you are sophisticated after finishing the first two beers, but as you drink more, you expect to become naive later on. The story indicates that in certain situations, a decision maker could cause her type to change; moreover, she might even be able to reason about such changes. Any perspective that assumes a fixed type is unable to capture such situations. In order to fix this shortcoming, we attempt a general interpretation of naiveté and sophistication for dynamic inconsistency. The language we develop allows the introduction of hybrid-type decision makers such as the one in our example.

Following a review of the relevant literature, we start building a formalism that allows for precise definitions of the two most commonly discussed types of decision makers, naifs and sophisticates. We work in discrete time, assuming that the situation of the decision maker can be captured as a Markov deci-

---

<sup>5</sup> The model of this situation is presented in Section 3.7.

---

sion problem. Adopting the terminology of multiple self models, we distinguish between the agent level and the level of the decision maker. Starting on the agent level, we introduce the notion of a strategy, which contains information about the intentions and beliefs of the individual agent at one particular history. Next, the properties of strategies (coherence, stationarity, consistency) are discussed. Moving to the level of the decision maker (i.e., the collection of all agents), we define the concept of a *frame*, and consider its properties. We link the properties of strategies to the properties of frames. Frames are a novel concept that provide all the relevant information about a decision maker facing dynamic inconsistency.

After clarifying our assumptions on utility functions, we define and introduce the two types of decision making, naiveté and sophistication. We provide existence results for optimal frames of both types, and discuss various properties of such frames. In the main text body, we then introduce decision makers with a hybrid type, also providing an existence theorem, and discuss two examples of hybrid decision making in detail. In the concluding section, we point towards further extensions of the model. The Appendix gives a minimalist summary of classical decision problems of the literature on dynamic inconsistency, and compares naively and sophisticatedly optimal frames for these problems. It also serves as an illustration for the applicability of our approach.

The contributions of this chapter are thus threefold. First, it provides new concepts and distinctions for problems of dynamic inconsistency. In particular, the representation of the decision maker through a frame, the distinction between a stationary strategy and frame, and the theorems relating consistency and stationarity should prove useful (Section 3.3.6). Second, it provides definitions for naiveté and sophistication, and shows the existence of naively and sophisticatedly optimal frames. Third, it introduces hybrid naive-sophisticated types, expanding the scope of dynamic inconsistency models. As a corollary, the analysis of hybrid types shows that self-deception can be optimal.

## 3.2 Related literature

Modelling approaches to dynamic inconsistency come in two varieties (Asheim, 2007). Dual-self planner-doer models bear a close analogy to principal-agent models (Thaler and Shefrin, 1981). A (single) *planner*, endowed with dynamically consistent preferences formulates plans; a present-biased *doer* can execute them or deviate from them. The conceptual background of planner-doer models is multifold: They sometimes rely on the hot-cold empathy gap (Loewenstein,

2005), on recent findings of neuroscience, or even the Freudian distinction between the *id* and the *ego*. Fudenberg and Levine (2006) argues in favor of a dual-self model as being analytically simpler, more in line with findings in neuroscience, and nevertheless being able to explain a large number of empirical phenomena. An important advantage of this approach is that welfare comparisons are relatively straightforward: The preferences of the planner are generally adopted to be normatively relevant. Recent models allow for the planner to learn about the doer's type through costly experimentation (Ali, 2011), or can include self-control (Gul and Pesendorfer, 2001; Benabou and Pycia, 2002; Fudenberg and Levine, 2012).

Despite all their advantages, dual-self models make the strong assumption that individuals have long-run, time-consistent preferences, and that these preferences can be identified. In this chapter we avoid this assumption, and instead adopt the multiple-self approach.<sup>6</sup> Our primary focus is on naive and sophisticated decision makers, and hybrid types. Multiple-self models have analyzed naive and sophisticated decision makers in a variety of settings. For instance, Laibson (1994); Fischer (1999); Laibson (1997); Angeletos et al. (2001) work with sophisticates, while Akerlof (1991) and O'Donoghue and Rabin (1999b) assume naiveté. In a general equilibrium setting, Herings and Rohde (2006, 2008) deal with both naifs and sophisticates. For us, a crucial precursor paper is O'Donoghue and Rabin (1999a), which compares naiveté and sophistication for the case of quasi-hyperbolic discounting, with so-called immediate costs and rewards.

The first model of hybrid types can be found in O'Donoghue and Rabin (2001). They use quasi-hyperbolic ( $\beta - \delta$ ) discounting to define partial naiveté. While naifs think their  $\beta$  is 1, the actual  $\beta$  is fully known to sophisticates, while partially naive agents think they have a  $\beta$  that is larger than their actual one. Partially naive agents thus entertain false beliefs about the future (just like naifs). This approach has proven its fruitfulness especially in the contract design literature (DellaVigna and Malmendier, 2004; Eliaz and Spiegel, 2006; Gilpatric, 2008); while DellaVigna and Malmendier (2006) focus on a monopolistic firm facing a mixed population of consumers. Heidhues and Kőszegi (2009) generalize the distributional assumptions on beliefs, but still in the context of quasi-hyperbolic discounting.

An important aspect of the O'Donoghue and Rabin (2001) approach is that – compared to sophistication – *any* degree of partial naiveté can generate arbi-

---

<sup>6</sup> Bach and Heilmann (2011) link multiple-self models to the philosophical literature on personal identity.

---

trarily large losses in efficiency for a decision maker. In this sense, the limit of partial naiveté (as the perceived present-biasedness approaches the real parameter) is not sophistication.

There were a few other attempts to treat hybrid decision makers. In Asheim (2007), agents have a perceived preference persistence, i.e., a probability (between 0 and 1) with which they think their preferences will be identical in the next period. However, this belief is always incorrect, as their preferences will change with probability 1. In Jehiel and Lilico (2010), agents endowed with exponential discounting have access to information about a number of future periods ('foresight'). They find that improving the length of foresight always improves welfare.

This chapter expands on the existing literature by providing the foundations of a hybrid model that is independent of the quasi-hyperbolic assumption of O'Donoghue and Rabin (2001), and where type determination is an endogenous part of the decision problem, as in our motivating example.

In the Appendix, we provide stylized models of a few standard problems often discussed in the context of dynamic inconsistency: procrastination (Akerlof, 1991; Fischer, 1999; Ariely and Wertenbroch, 2002; Steel, 2007), impulsiveness (Ainslie, 1974, 1975), underinvestment (Laibson et al., 1998; Angeletos et al., 2001; DellaVigna and Malmendier, 2004) and bingeing behavior/addiction (Gruber and Kőszegi, 2001; Ainslie and Monterosso, 2003). Indulgence is a particular type of timing problem, where the issue is when to consume a certain good. We adopt the latter model directly from O'Donoghue and Rabin (1999a). For all of these problems, we compare the welfare of naive and sophisticated decision makers. The Appendix also serves as a means of illustrating the potential use of our framework.

### 3.3 Basic concepts

In this section, we introduce our framework and notations. Standard definitions are provided for the notions of 'decision problem' and 'history'. However, before introducing strategies, we argue in Section 3.3.3 for a new way of defining strategies that includes both beliefs and intentions. We then proceed by defining strategies and frames, corresponding to two different levels of analysis. Towards the end of the section, we present some results on the relationship between consistency and stationarity.

### 3.3.1 Markov decision problem

We start with a decision maker facing a finite Markov decision problem on an infinite horizon.<sup>7</sup>

**Definition 12.** A *finite Markov decision problem* is given by:

- the set of time periods  $T = \{0, 1, 2, \dots\}$ ;
- a finite set of states  $\Omega$ , with  $\bar{\omega} \in \Omega$  as the initial state;
- a finite and nonempty set of pure actions  $A_\omega$  that the decision maker can choose from in state  $\omega$ ;
- a payoff function  $u_\omega : A_\omega \rightarrow \mathbb{R}$  that assigns a payoff to every action in state  $\omega$ ;
- transition probabilities  $m_\omega : A_\omega \rightarrow \Delta(\Omega)$ , with  $m_\omega(\omega' | a_\omega)$  denoting the probability to transit from state  $\omega$  to state  $\omega'$  when action  $a_\omega$  is chosen.

This definition excludes the possibility of randomization over actions, although in some cases such an extension is needed to guarantee the existence of a stationary optimal strategy (see Section 3.A.2). Nevertheless, in order to keep the presentation simpler, for the moment we only consider pure actions.

### 3.3.2 History

To capture all the informational aspects on which an action choice can be conditioned, we introduce the notion of a history:

**Definition 13.** A *history*  $h$  has the form  $h = (\omega_0, a_{\omega_0}, \dots, \omega_{t-1}, a_{\omega_{t-1}}, \omega_t)$ , with:

- $\omega_i \in \Omega$ , for  $i \in \{0, 1, \dots, t\}$ , and  $\omega_0 = \bar{\omega}$ ;
- $a_{\omega_i} \in A_{\omega_i}$ , for  $i \in \{0, 1, \dots, t-1\}$ ;
- $m_{\omega_i}(\omega_{i+1} | a_{\omega_i}) > 0$ , for  $i \in \{0, 1, \dots, t-1\}$ .

The length of  $h$  or current time at  $h$  is denoted by  $t = t(h)$ , and the function  $\omega(h) = \omega_{t(h)}$  indicates the current or end state at history  $h$ . We use  $H$  to refer to the set of all histories.

---

<sup>7</sup> The latter is not a restrictive requirement, since it is easy to rewrite a decision problem on a finite horizon to one on an infinite horizon.

If history  $h'$  begins with  $h$ , we say that  $h'$  *succeeds*  $h$ , or equivalently, that  $h$  *precedes*  $h'$ ; and denote this with  $h \preceq h'$ , or equivalently, with  $h' \succcurlyeq h$ .<sup>8</sup> The subset of  $H$  that consists of all histories that succeed  $h$  is denoted by  $H^{\succcurlyeq h}$ :

$$H^{\succcurlyeq h} = \{h' \in H \mid h' \succcurlyeq h\}.$$

We refer to  $h_0 = (\omega_0)$  as the ‘root history’. To shorten notation, when specifying a history, we sometimes omit the commas separating states and actions, and also the enclosing parentheses. Thus, history  $h = (\omega_0, a_{\omega_0}, \omega_1)$  will be occasionally written as  $h = \omega_0 a_{\omega_0} \omega_1$ .

### 3.3.3 Conceptual foundations

The fundamental entities in our model are *agents*<sup>9</sup>: They are the ones with the ultimate power of choosing an action and executing it. We assume a one-to-one correspondence between histories and agents. Our agents form beliefs and intentions about the future, and have the ability to choose an action in the present.<sup>10</sup> To refer to the collection of all agents, we use the notion of *decision maker*.

We keep the general assumption that past actions have no effect on the well-being generated by current and future actions of the agents, i.e., ‘bygones are bygones.’ Our agents have full control over their current actions.<sup>11</sup>

Using the distinction between experienced utility and decision utility (Kahneman, Wakker and Sarin, 1997), we can delimit two senses of ‘expected utility from taking action  $a$ .’ Let us disregard the immediate payoff for taking the action, and consider only future payoffs. In one sense, the phrase could mean ‘experienced utility from expectation’, i.e., utility that *is actually experienced* by the agent due to expecting a certain stream of future payoffs. Think of a student that decides to study for an upcoming exam instead of watching her favorite TV show. She might, in fact, *already* enjoy the benefits of the decision to study (she is already less anxious around the exam, maybe she relishes the idea that she is doing ‘the right thing’, etc.). The other sense of ‘expected utility’ could be rendered as ‘the expected present value of various streams of payoffs’

<sup>8</sup> Obviously,  $h' \preceq h$  &  $h \preceq h' \Leftrightarrow h = h'$ .

<sup>9</sup> Others in the multiple-selves approach refer to them as ‘selves’.

<sup>10</sup> We follow Cowen (1991), who identifies a self ‘with a set of preferences linked to certain cognitive and volitional capacities.’

<sup>11</sup> To appreciate that this choice is not so obvious, see Jehiel and Lilico (2010). Similarly, Elster’s interpretation of the Ulysses story is an example of a model where control over current action is essentially eliminated (Elster, 1979).

that the agent's current decision can lead to. In this sense, the agent does not experience any actual change in utility by choosing one or other course of action; she is merely able to calculate with these future payoffs. This distinction between the two senses of expected utility will be used in Section 3.4, and in our conceptual interpretations of naiveté and sophistication in subsequent sections.

### 3.3.4 Intentions, beliefs, strategy

We aim to give a full description for the two most prevalent decision maker types (naifs and sophisticates) and the hybrid types that we introduce later. For this purpose, we define a strategy as having three components: the current action, the intended future actions, and the belief about what future agents will in fact do. There is no special reason for assuming that the latter two coincide for future actions, although with our definitions, they coincide for naifs and sophisticates, but not for hybrid decision makers. To simplify notation, we reduce this triadic framework to just intentions and beliefs, and assume that for the current action, these two have to coincide: No agent can be wrong about which action she takes, and each agent takes the action that she intends to *at that moment*.

The basic building blocks of our model are all functions from the set of histories that succeed the agent to the set of available actions at those histories.

**Definition 14.** The *intentions* of an agent at history  $\bar{h}$  assign an intended action to each history that succeeds the present:

$$i^{\bar{h}} : h \in H^{\succ \bar{h}} \mapsto A_{\omega(h)}.$$

**Definition 15.** The *beliefs* of an agent at history  $\bar{h}$  assign an action to each history that succeeds the present:

$$b^{\bar{h}} : h \in H^{\succ \bar{h}} \mapsto A_{\omega(h)}.$$

Note that intentions and beliefs are defined at all succeeding histories, even at those that the agent does not intend to reach or believes will not be reached.

We now proceed to define strategies as the pair of intentions and beliefs for an agent.

**Definition 16.** A *strategy* of an agent at history  $\bar{h}$  is a pair of intentions and beliefs for that agent, with the added property that the belief and intention for the current action coincide:

$$s^{\bar{h}} = (i^{\bar{h}}, b^{\bar{h}}), \text{ with } i^{\bar{h}}(\bar{h}) = b^{\bar{h}}(\bar{h}).$$

The set of all strategies for this agent is denoted by  $S^{\bar{h}}$ .

For an agent at  $\bar{h}$ ,  $i^{\bar{h}}(h)$  refers to the intention, while  $b^{\bar{h}}(h)$  refers to the belief component of the strategy  $s^{\bar{h}}$  about the future agent at  $h$ . For example,  $b^{\bar{h}}(h) = a$  should be read as: ‘The agent at  $\bar{h}$  who holds strategy  $s^{\bar{h}}$  believes the agent at  $h$  will choose action  $a$ .’

We emphasize that our definition of a strategy in terms of an intention-belief pair is not standard. In fact, the most common practice is to conflate the two concepts. On the other hand, in epistemic game theory, beliefs are strictly separated from actions. It might be convenient for the reader to think of the term ‘strategy’ simply as an abbreviation for ‘intention-belief pair.’

We can now define two properties of strategies, stationarity and coherence, as well as a relation over the set of strategies, consistency.

**Definition 17.** The intentions (or beliefs) of an agent at  $\bar{h}$  are *stationary* whenever the intended (believed) actions depend only on the end-state. Thus,  $i^{\bar{h}}$  or  $b^{\bar{h}}$  is called stationary if, for all  $h, h' \in H^{\succ \bar{h}}$  with  $\omega(h) = \omega(h')$ , we have  $i^{\bar{h}}(h) = i^{\bar{h}}(h')$  or respectively,  $b^{\bar{h}}(h) = b^{\bar{h}}(h')$ . A strategy  $s^{\bar{h}}$  is stationary if both its constituent intentions  $i^{\bar{h}}$  and beliefs  $b^{\bar{h}}$  are stationary.

For example, if each day of the week can be modeled as a single state, the strategy of an agent who intends and believes eating in a restaurant every second Saturday, but staying home on every other one is not stationary.

**Definition 18.** An agent at  $\bar{h}$  is said to hold a *coherent* strategy, if her intention and beliefs about future actions coincide for all future histories. Formally, a strategy  $s^{\bar{h}} = (i^{\bar{h}}, b^{\bar{h}})$  of an agent at  $\bar{h}$  is coherent if  $i^{\bar{h}}(h) = b^{\bar{h}}(h)$  for all  $h \in H^{\succ \bar{h}}$ .

For example, a strategy of an agent who intends to stop drinking, but believes she will be unable to do so is not coherent.

**Definition 19.** The strategies of two agents at  $h$  and  $h'$  are said to be *consistent* if they assign the same intentions and beliefs to each history that succeeds both agents, i.e.,  $s^h$  and  $s^{h'}$  are consistent, if  $s^h(h'') = s^{h'}(h'')$  for all  $h'' \in H^{\succ h} \cap H^{\succ h'}$ .<sup>12</sup>

For example, a strategy formulated yesterday which intended eating apples today and a strategy formulated today which intends eating cookies instead are not consistent.

<sup>12</sup> So, if two strategies are defined at histories that neither succeed nor precede each other, then they are consistent, as there are no histories that succeed both.

While coherence concerns the relationship between the intentions and beliefs of the same strategy, i.e., belonging to one agent, consistency compares strategies of two distinct agents. In other words, coherence is an intrinsic property, whereas consistency is an extrinsic (relational) property of a strategy.

A natural question is whether the consistency of strategies is transitive, i.e., whether the consistency of  $s^h$  and  $s^{h'}$  and the consistency of  $s^{h'}$  and  $s^{h''}$  imply that  $s^h$  and  $s^{h''}$  are also consistent. If  $h \preceq h' \preceq h''$ , then this is indeed the case. However, without this constraint, consistency is not transitive in general – it is not even transitive within the set of stationary strategies. To see this, take the decision problem in Figure 3.1.<sup>13</sup> We construct three stationary strategies  $s^{h_\rho}$ ,  $s^{h_\sigma}$ , and  $s^{h_\tau}$  such that  $s^{h_\rho}$ ,  $s^{h_\sigma}$  are consistent, as well as  $s^{h_\sigma}$ ,  $s^{h_\tau}$ , but  $s^{h_\rho}$ ,  $s^{h_\tau}$  are not consistent. Fix  $h_\rho = (\rho)$ ,  $h_\sigma = (\rho, A, \sigma)$  and  $h_\tau = (\rho, B, \tau)$ . Also, let  $s^{h_\rho}(h_\rho) = (A, A)$ ,  $s^{h_\rho}(h) = (C, C)$  if  $\omega(h) = \sigma$ , and  $s^{h_\rho}(h) = (E, E)$  if  $\omega(h) = \tau$ . Intuitively,  $s^{h_\rho}$  means: ‘I choose  $A$ , believe and intend  $C$  in state  $\rho$ , and believe and intend  $E$  in state  $\tau$ .’ Define two other strategies through  $s^{h_\sigma}(h) = (C, C)$  for all  $h \succcurlyeq h_\sigma$  (‘do  $C$  in  $\sigma$ ’), and  $s^{h_\tau}(h) = (F, F)$  for all  $h \succcurlyeq h_\tau$  (‘do  $F$  in  $\tau$ ’). All of these strategies are stationary. Clearly,  $s^{h_\rho}$  and  $s^{h_\sigma}$  are consistent, since they both require the decision maker to choose  $C$  in state  $\sigma$ , and after history  $h_\sigma$  no state other than  $\sigma$  is reachable. Next,  $s^{h_\sigma}$  and  $s^{h_\tau}$  are consistent, since histories  $h_\sigma$  and  $h_\tau$  neither succeed, nor precede each other. But  $s^{h_\rho}$  and  $s^{h_\tau}$  are not consistent, as they assign different actions to the state  $\tau$ . This shows that consistency of strategies is not transitive on the set of stationary strategies.

### 3.3.5 Truncation

The following definitions of ‘truncation’, albeit technical in nature, are necessary for the definition of a stationary frame. Truncation formalizes the idea of ‘bygones are bygones,’ and chips away an initial segment of the history.

**Definition 20.** Take any history  $h = (\omega_0, a_{\omega_0}, \dots, \omega_t(h))$ . The *truncation* op-

<sup>13</sup> We use figures like this one to represent decision problems. States are denoted by Greek characters; in this decision problem, we have states  $\rho, \sigma, \tau$ , with  $\boxed{\rho}$  indicating the initial state. For actions, we use Roman capitals – in this case, we have action  $A, B$  (available in state  $\rho$ ),  $C$  (the only action available in  $\sigma$ ), and  $E, F$  (in  $\tau$ ). Whenever – unlike in this example – payoffs are important, the payoff associated with an action is written in the form  $A|3$ , meaning that choosing action  $A$  generates a payoff of 3. Our examples involve only deterministic transition probabilities, and the transitions associated with each state are represented by arrows. For example, for the decision problem in Figure 3.1 choosing action  $E$  in state  $\tau$  implies that the next state will be  $\tau$  with probability 1.

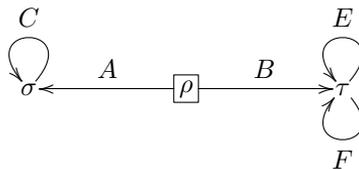


Figure 3.1: Stationary strategies with intransitive consistency.

erator  $\neg_k$ , defined for any  $k \leq t(h)$ , removes the first  $k$  pairs of this sequence, so  $\neg_k h = (\omega_k, a_{\omega_k}, \dots, \omega_{t(h)})$ . For a set of histories  $H' \subseteq H$ , we refer to the set of  $k$ -truncated histories by  $\neg_k H'$ .

A truncated strategy applies this idea to strategies: It is defined for an agent who ‘forgot’ (or disregards) all of her past; future histories obviously do not include descriptions of forgotten segments of the past anymore. Thus, the truncated strategy of an agent at  $\bar{h}$  will be defined on the set  $\neg_{t(\bar{h})} H^{\succ \bar{h}}$ .

**Definition 21.** For any strategy  $s^{\bar{h}}$ , the *truncated* strategy  $\neg s^{\bar{h}} : h \in \neg_{t(\bar{h})} H^{\succ \bar{h}} \mapsto A_{\omega(h)}$  denotes the function for which  $\neg s^{\bar{h}}(\neg_{t(\bar{h})} h) = s^{\bar{h}}(h)$ , for all  $h \in H^{\succ \bar{h}}$ .

We note that while truncated histories are defined for truncations of arbitrary length ( $\neg_k$ ), for strategies we only need truncations of length  $t(\bar{h})$  for a strategy  $s^{\bar{h}}$  of an agent at history  $\bar{h}$ .

To see this definition at work, think of the decision to stop smoking on the first day of the next month. Take an agent who resolves on July 24th: ‘I will stop smoking from August 1st’ and then fails. On August 24th, she makes another decision: ‘I will stop smoking from September 1st’. If we interpret each statement as a strategy, it is easy to see that they are not consistent: For instance, they prescribe different smoking behavior for August 28th – the first strategy forbids it, while the second allows it. However, there is an intuitive sense in which they are very similar. Indeed, they map them into the same resolve that uses indexicals instead of precise dates: ‘I can smoke for one more week, and then I will stop’.<sup>14</sup> Truncating the present history highlights this

<sup>14</sup> Actually, the difference between specifying a future consumption period by a *calendar date* or through its *temporal distance* from the present has already been noticed by Strotz (1956). This difference is experimentally explored by Read et al. (2005), which finds that subjects only exhibit hyperbolic discounting when future periods are identified via their temporal distance.

similarity by getting rid of the past. In our example, the original strategies are not identical or consistent; but their truncated versions are identical.

### 3.3.6 Frames

We now move from the agent level to the level of the decision maker. Since there is no *a priori* reason for the agents to have consistent strategies, different agents can form different intentions and entertain different beliefs about any certain future agent. To have an ‘external’ overview of all agents, we introduce the concept of a frame. In our terminology, a frame is an auxiliary tool for representing the strategies of all possible agents, and not something that is intentionally put together by the decision maker. Whereas each agent chooses a strategy, the decision maker does not choose a frame. Instead, a frame contains a full description of the intentions and beliefs in all contingencies, i.e., at all histories.

**Definition 22.** A *frame* assigns a strategy to each agent. That is, a frame is a function  $f : h \in H \mapsto S^h$ .

Figure 3.2 shows an extremely simple decision problem, for which an example of a frame is represented in Table 3.1. Each entry is a pair of  $A$ ’s and  $B$ ’s, an intended action and a belief about an action. Each row corresponds to a strategy for an agent at  $\bar{h}$ , defining an intention and a belief for each history that succeeds  $\bar{h}$ . For example, the entry  $AB$  for row  $\bar{h} = \rho A\rho$  and column  $h = \rho A\rho A\rho$  should be interpreted as: the agent at  $\rho A\rho$  intends to choose action  $A$  at history  $\rho A\rho A\rho$ , while believing the agent at  $\rho A\rho A\rho$  will, in fact, choose action  $B$ . The whole frame thus specifies the intentions and beliefs of all agents over all other (present and future) agents. Our definition of a strategy ensures that on the diagonal of the table, the intentions and the beliefs match.



Figure 3.2: A basic decision problem.

We now proceed to introduce three properties of frames. Our definition of stationarity makes use of the truncation operator defined above.

		$h \in H^{\succ \bar{h}}$					
		$\rho$	$\rho A \rho$	$\rho B \rho$	$\rho A \rho A \rho$	$\rho A \rho B \rho$	...
$\bar{h}$	$\rho$	$BB$	$AB$	$AB$	$AA$	$BB$	...
	$\rho A \rho$	–	$BB$	–	$AB$	$BA$	...
	$\rho B \rho$	–	–	$AA$	–	–	...
	$\rho A \rho A \rho$	–	–	–	$BB$	–	...
	$\rho A \rho B \rho$	–	–	–	–	$AA$	...
	...	–	–	–	–	–	...

Table 3.1: An example of a frame for the decision problem in Figure 3.2.

**Definition 23.** A frame  $f$  is said to be *stationary*, if only the end-state matters when assigning strategies to histories, i.e., for any histories  $h$  and  $h'$ , if  $\omega(h) = \omega(h')$ , then  $\neg f(h) = \neg f(h')$ .

Stationarity of a frame is different from the stationarity of the strategies involved. For the decision problem in Figure 3.2, Table 3.2 offers a non-stationary frame of stationary strategies. To check for this, what needs to be verified is that each row represents a stationary strategy. As there is only one state for this decision problem, this means that in each row, we should see the same intention-belief pair, which is indeed the case. Thus, Table 3.2 shows a frame of stationary strategies. The frame itself however is not stationary: By truncating the strategies in the first and second row, we get a different strategy.

On the other hand, Table 3.3 shows a stationary frame of non-stationary strategies. It is easy to see that it is a frame of non-stationary strategies, as each row represents one strategy, which are not stationary – for instance,  $s^\rho(\rho) = (B, B) \neq (A, A) = s^\rho(\rho A \rho)$ . The frame itself is stationary, which can be checked by comparing the rows. As there is only one state, we need to compare the truncations of all strategies. It can be read from Table 3.3 that by ‘forgetting the past’, we get the same strategy in each row, namely: The agent picks action  $B$  right away, intends to choose action  $A$ , and believes she will always do so in the future.

For a concrete example of a stationary frame of non-stationary strategies, think of the decision maker who, waking up every day, decides to take *just one more* shot of heroin, and intends (and believes) to quit the next day.

Next, we define a consistent frame. The intuitive idea is that a frame is consistent if no deviation can be expected from previous intentions and beliefs.

**Definition 24.** A frame  $f$  is said to be *consistent* if the strategies  $f(h)$  and

		$h \in H^{\succ \bar{h}}$					
		$\rho$	$\rho A \rho$	$\rho B \rho$	$\rho A \rho A \rho$	$\rho A \rho B \rho$	$\dots$
$\bar{h}$	$\rho$	$AA$	$AA$	$AA$	$AA$	$AA$	$AA$
	$\rho A \rho$	$-$	$BB$	$-$	$BB$	$BB$	$BB$
	$\rho B \rho$	$-$	$-$	$AA$	$-$	$-$	$AA$
	$\rho A \rho A \rho$	$-$	$-$	$-$	$AA$	$-$	$AA$
	$\rho A \rho B \rho$	$-$	$-$	$-$	$-$	$BB$	$BB$
	$\dots$	$-$	$-$	$-$	$-$	$-$	$\dots$

Table 3.2: Non-stationary frame of stationary strategies.

		$h \in H^{\succ \bar{h}}$					
		$\rho$	$\rho A \rho$	$\rho B \rho$	$\rho A \rho A \rho$	$\rho A \rho B \rho$	$\dots$
$\bar{h}$	$\rho$	$BB$	$AA$	$AA$	$AA$	$AA$	$AA$
	$\rho A \rho$	$-$	$BB$	$-$	$AA$	$AA$	$AA$
	$\rho B \rho$	$-$	$-$	$BB$	$-$	$-$	$AA$
	$\rho A \rho A \rho$	$-$	$-$	$-$	$BB$	$-$	$AA$
	$\rho A \rho B \rho$	$-$	$-$	$-$	$-$	$BB$	$AA$
	$\dots$	$-$	$-$	$-$	$-$	$-$	$\dots$

Table 3.3: Stationary frame of non-stationary strategies.

$f(h')$  assigned to any two histories  $h$  and  $h'$  are consistent.

Consistency is a very strong notion: A consistent decision maker – at whichever history she is contemplating the present and the future – would never change her mind about any intention-belief pair. An example would be a heroin user who goes cold turkey immediately and definitely, never ever restarting her substance use. If she relapses, she will not be consistent anymore: Her choice to quit for good implies that at the time of quitting her intentions and beliefs for the future history at which she relapses do not match her intentions and beliefs at the point of relapse for the (then-)current history.

According to this definition, if  $f$  is a consistent frame, then we get  $f(h)(h'') = f(h')(h'')$  whenever  $h'' \in H^{\succ h} \cap H^{\succ h'}$ , and either  $h' \in H^{\succ h}$ , or  $h \in H^{\succ h'}$ . Note that, since for a consistent frame,  $f(h)(h) = f(h')(h)$  for all  $h'$  and  $h \in H^{\succ h'}$ ; and our requirement for strategies that  $i^h(h) = b^h(h)$ , a consistent frame is necessarily made up by coherent strategies. This implies that a choice of an action for all histories uniquely determines a consistent frame. Similarly, a choice of an action for all states uniquely determines a consistent frame of stationary strategies.

**Theorem 5.** *A consistent, stationary frame consists of stationary strategies.*

*Proof.* Take any histories  $h, h'$  and  $h''$  for which  $\omega(h') = \omega(h'')$ . We have to show that  $f(h)(h') = f(h)(h'')$ . For this, see that:

$$\begin{aligned} f(h)(h') &= f(h')(h') = \neg f(h')(\neg_{t(h')}h') = \\ &= \neg f(h'')(\neg_{t(h')}h'') = f(h'')(h'') = f(h)(h''). \end{aligned}$$

For the respective equations, we use, in order, consistency, definition of truncation, stationarity of the frame, definition of truncation, and consistency again.  $\square$

**Theorem 6.** *A consistent frame of stationary strategies is a stationary frame.*

*Proof.* Take histories  $h, h', h'', h'''$  with  $\omega(h) = \omega(h')$ . We have to show that

$$(\neg f(h))(\neg_{t(h)}h'') = (\neg f(h'))(\neg_{t(h')}h''')$$

if  $\neg_{t(h)}h'' = \neg_{t(h')}h'''$ . Note that such  $h''$  and  $h'''$  exist, because the end-state in  $h$  and  $h'$  is identical. Using the fact that  $\omega(h'') = \omega(\neg_{t(h)}h'') = \omega(\neg_{t(h')}h''') = \omega(h''')$ , we get:

$$\begin{aligned} (\neg f(h))(\neg_{t(h)}h'') &= f(h)(h'') = f(h_0)(h'') = f(h_0)(h''') = \\ &= f(h')(h''') = (\neg f(h'))(\neg_{t(h')}h'''). \end{aligned}$$

We use, in turn, the definitions of truncation, consistency, stationarity of the strategies, consistency, and finally, truncation again. (We make use of the root strategy  $h_0$  as a history which is surely succeeded by both  $h''$  and  $h'''$ .)  $\square$

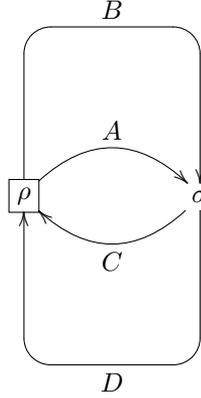


Figure 3.3: Stationary frame of stationary strategies is not necessarily consistent.

Based on the two theorems above, one might expect that a stationary frame of stationary strategies would be consistent. However, this is not necessarily so, as can be seen from the following example. Consider the Markov decision problem in Figure 3.3. For all  $h \succ \bar{h}$ , let:

$$s_1^{\bar{h}}(h) = \begin{cases} (A, A) & \text{if } \omega(h) = \rho, \\ (C, C) & \text{if } \omega(h) = \sigma, \end{cases} \quad \text{and} \quad s_2^{\bar{h}}(h) = \begin{cases} (B, B) & \text{if } \omega(h) = \rho, \\ (D, D) & \text{if } \omega(h) = \sigma. \end{cases}$$

Now, let us define a frame  $f$ , so that:

$$f(h) = \begin{cases} s_1^{\bar{h}} & \text{if } \omega(\bar{h}) = \rho; \\ s_2^{\bar{h}} & \text{if } \omega(\bar{h}) = \sigma. \end{cases}$$

This is obviously a frame of stationary strategies. It also is a stationary frame, since only the end-state matters in assigning a strategy to a history, according to the definition. However, it is not a consistent frame, since:

$$f(\rho)(\rho A \sigma) = s_1^\rho(\rho A \sigma) = (A, A) \neq (B, B) = s_2^{\rho A \sigma}(\rho A \sigma) = f(\rho A \sigma)(\rho A \sigma).$$

### 3.3.7 Induced strategy

We assume that, since each agent has control over her current action (and only that), the actual actions executed by each agent  $h$  can be obtained from frame  $f$  by looking at  $f(h)(h)$ , in other words, from the diagonal of the frame.

**Definition 25.** The *induced strategy* of a frame  $f$  specifies the actual actions chosen by each agent:

$$\Lambda(f) : h \in H \mapsto A_{\omega(h)}, \text{ given by } \Lambda(f)(h) = f(h)(h).$$

The induced strategy of the frame represented in Table 3.1 is  $\Lambda(f)(\rho) = (BB)$ ,  $\Lambda(f)(\rho A \rho) = (BB)$ ,  $\Lambda(f)(\rho B \rho) = (AA)$  etc.

It is handy to define for some frame  $f$ , and an agent at  $\bar{h}$ , the induced strategy for the (present and) future:

$$\begin{aligned} \Lambda^{\succ \bar{h}}(f) : h \in H^{\succ \bar{h}} &\mapsto A_{\omega(h)}, \text{ given by } \Lambda^{\succ \bar{h}}(f)(h) = f(h)(h); \\ \Lambda^{\succ \bar{h}}(f) : h \in H^{\succ \bar{h}} \setminus \{\bar{h}\} &\mapsto A_{\omega(h)}, \text{ given by } \Lambda^{\succ \bar{h}}(f)(h) = f(h)(h). \end{aligned}$$

### 3.3.8 Remarks

The considerations of this section highlight the interdependencies between various concepts. Coherence reflects a match between beliefs and intentions, desires and reality, and is a necessary condition for consistency. Consistency of a frame ensures that agents are not ‘let down’ by future agents, and thus that expectations match outcomes: A consistent decision maker should not experience *ex post* regret. In various contexts, stationarity of strategies and stationarity of a frame can also be desirable properties; primarily, stationary (Markovian) strategies embody the simplest form of behavior consistent with rationality (Maskin and Tirole, 2001). Finally, a stationary frame describes a decision maker who is stable over time.

The language developed here can be helpful for discussing problems of dynamic inconsistency formulated in the multiple-agent framework, and is independent of the specific assumptions on utility functions of the next section. In particular, we believe that our distinction between beliefs/intentions, agents/decision makers, strategies/frames, as well as our method of representing frames through tables should prove useful. To see the latter in use, we refer the reader to the Appendix. Moreover, the distinction between stationary strategies and stationary frames is, as far as we are aware, entirely new in the literature.

### 3.4 Utility and discounting

The term *payoff*, introduced in Definition 12, refers to the immediate gains or losses resulting from an action. Formally, a payoff gained in period  $t$  is denoted by  $u_t$ , and a stream of payoffs starting at period  $t$  by  $u_{t \rightarrow} = (u_t, u_{t+1}, \dots)$ . We say that a stream of payoffs  $u_{t \rightarrow}$  starting at period  $t$  coincides with a stream of payoffs  $u'_{t' \rightarrow}$  starting at period  $t'$  if  $u_t = u'_{t'}$  and  $u_{t+1} = u'_{t'+1}$ , and so on.

Payoffs are fully determined by the decision problem, the state and the action taken. However, time preference implies that identical payoffs might be regarded differently by various agents. Throughout the chapter, we make two assumptions on the utility functions  $U^h(u)$  that integrate a stream of future payoffs into a single number. The first assumption states that  $U^h$  is continuous at infinity for every  $h$ , and is adapted from Fudenberg and Levine (1983).

**Assumption 1.**  $U^h$  is continuous at infinity for every  $h$ , i.e., for any  $\epsilon$ , there is a horizon  $T(h)$  – possibly depending on  $h$  – such that the total variation of utility after  $t(h) + T(h)$  is less than  $\epsilon$ .

Another crucial assumption that we adopt is that agents are identical in the way they evaluate streams of payoffs, i.e., we assume stationary preferences (Peleg and Yaari, 1973).

**Assumption 2.** For any two agents at histories  $h$  and  $h'$ , and coinciding streams of payoffs  $u_{t(h) \rightarrow}, u'_{t(h') \rightarrow}$ , the utilities of the two agents are equal, i.e.,  $U^h(u_{t(h) \rightarrow}) = U^{h'}(u'_{t(h') \rightarrow})$ .

If the utility functions satisfy First and Second Order Separability<sup>15</sup> (Lapied and Renault, 2012), then the discount factor for a future payoff  $u_t$  can only depend on the time distance  $t(h) - t$ . In our examples in Section 3.7 and the Appendix, we use a discounted utility function of a particular form, namely, quasi-hyperbolic discounting:

$$U^h(u_{t(h) \rightarrow}) = u_{t(h)} + \beta \sum_{t=t(h)+1}^{\infty} \delta^{t-t(h)} u_t,$$

with  $0 \leq \beta \leq 1$  and  $0 \leq \delta < 1$ .

---

<sup>15</sup> First Order Separability means that preferences over a set of outcomes are such that there is no interaction between the outcomes of various periods. Second Order Separability allows the isolation of effects of temporal distance.

In Section 3.3.3, we distinguished between two senses of the term ‘expected utility’: utility *actually experienced* from expecting a future payoff stream, and ‘expected utility’ as simply *a means of calculating* with various future courses of action. This distinction is formally nailed down and further refined by the following definitions.<sup>16</sup>

**Definition 26.** The *expected utility based on intentions* of playing strategy  $s^h$  for an agent at  $h$  is:

$$U_i^h(s^h) = \mathbb{E}[i^h](U^h).$$

The next definition focuses on how much utility an agent can reasonably expect *ex ante*, whenever making utility calculations:

**Definition 27.** The *expected utility based on beliefs* of playing strategy  $s^h$  for an agent at  $h$  is:

$$U_b^h(s^h) = \mathbb{E}[b^h](U^h).$$

Our final definition disregards mere expectations, and captures the utility gained by an agent by considering which actions future agents will have *actually* implemented under various eventualities. Using the notion of an induced strategy, we can define *induced utility*:

**Definition 28.** Given a frame  $f$ , the (ex post) *induced utility* of the root agent at  $h_0$  is:

$$U_r(f) = \mathbb{E}[\Lambda^{\succsim_{h_0}}(f)](U^{h_0}).$$

We will use the notion of induced utility for welfare comparisons between various frames, especially in the Appendix, when contrasting naive and sophisticated decision makers in particular decision problems. Implicitly, this means picking the perspective of the root agent for welfare comparisons of frames. We do so in order to avoid any normative assumptions on the agent’s long-run preferences.<sup>17</sup>

Notice that traditionally, the above three meanings of the term ‘expected utility’ coincide. The reason is that where dynamic inconsistency does not pose a problem, intentions and beliefs on future actions coincide; moreover, the decision maker always executes the intentions of past agents.

<sup>16</sup> Sáez-Martí and Weibull (2005) connect discounting with altruism towards future selves. They note that “[c]urrent welfare or ‘total utility’, so defined, does not stem only from current instantaneous utility but also from (the anticipation of) the stream of future instantaneous utilities”. Our distinction highlights exactly this: Is it the actual future stream, or the anticipation of that stream that is really important?

<sup>17</sup> One alternative would be using a Pareto criterion, as for instance, in Herings and Rohde (2006).

### 3.5 Naiveté

At first sight, it is not even clear whether naiveté is a property of the decision maker or that of an agent. In this and the following section, we define naiveté and sophistication primarily for agents, assuming that the decision maker is always naive (or sophisticated). We return to the introductory example presented in Section 3.1 after analyzing and contrasting these base cases.

Naiveté has been characterized in several ways in the literature; a naif is aware or unaware of different things, depending on the particular interpretation. A naif is said to:

- choose at each stage an option which seems currently the best (Strotz, 1956; Hammond, 1976);
- fail to realize that future selves will have different preferences (O'Donoghue and Rabin, 2001; Sarafidis, 2004; DellaVigna and Malmendier, 2006; Herings and Rohde, 2006; Heidhues and Köszegi, 2009);
- believe that – though her preferences might change – she has perfect self-control about the future, allowing her to commit to a strategy (O'Donoghue and Rabin, 1999a; Gruber and Köszegi, 2001; Ali, 2011).

It is easy to see that these are genuinely alternative interpretations of naiveté. A common aspect is that something is amiss with the beliefs held by the agent. We argue that these troubles arise from the way the naif determines her beliefs. In particular, for a naive agent, her current preferences determine her intentions, which in turn determine her beliefs on future actions. Thus, it does not matter whether the agent holds an explicit belief on the lack of change in her preferences, or whether she believes she will simply fail to act on such changes, or that she has strong beliefs in her own will- or pre-commitment power. The essential features of naiveté are the directions of determination seen in Figure 3.4. All the above cases are described by this model.

Preferences  $\rightarrow$  intentions  $\rightarrow$  beliefs  $\rightarrow$  strategy.

Figure 3.4: The forming of intentions and beliefs by a naif.

**Definition 29.** A strategy  $\tilde{s}^h$  of an agent at  $h$  is *naively optimal*, if it maximizes expected utility based on intentions:

$$\tilde{s}^h \in \arg \max_{s \in S^h} U_i^h(s),$$

and it is coherent:

$$\tilde{b}^h(h') = \tilde{i}^h(h'), \text{ for all } h' \in H^{\succ h}.$$

A frame  $\tilde{f}$  is naively optimal if the strategy  $\tilde{f}(h)$  is naively optimal at each history  $h$ .

Working in a continuous-time discounted utility framework, Strotz (1956) shows that only when the discount function is exponential does the decision maker possess a consistent naively optimal frame for all decision problems. For any non-exponential discount functions there are decision problems for which there is no consistent naively optimal frame. However, the existence of a stationary naively optimal frame is guaranteed under our assumptions.

**Theorem 7.** *For any decision problem, there exists a stationary naively optimal frame.*

*Proof.* For each  $h$ , the set  $\arg \max_{s \in S^h} U_i^h(s)$  is nonempty, because the set  $S^h$  is nonempty and closed for pointwise limits, and  $U_i^h$  is continuous (Assumption 1). Therefore, the set of strategies where the maximum is, in fact, reached is nonempty. But note that the optimality condition in the definition of naively optimal strategies only determines the intention-component of strategies, thus, beliefs can be constructed freely. This means that we can ensure coherence, i.e., we can choose a naively optimal strategy at each  $h$ .

Now, to guarantee that the generated frame is stationary, we need to choose the same truncated strategy for each set of histories where the end-state is identical. This is always possible, since whenever the final state is identical for two histories, both the truncated strategy set and the utility function defined at those histories are identical (Assumption 2), and therefore so are the set of truncated optimal strategies.  $\square$

Is the stationary naively optimal frame unique? For a naive decision maker, this depends on whether for each history, there is a unique naively optimal strategy; this latter problem can be reduced to whether for each *state* that can be reached, there is a unique naively optimal strategy (defined at any history where the current state is that state). For generic decision problems, it seems likely that this is indeed the case, though the scope of the proof is beyond this work. Now, if there is a unique naively optimal strategy for each history, then there is only one naively optimal frame – and it is stationary, too. In degenerate cases, where multiple naively optimal strategies can be assigned to at least one state, we get stationary naively optimal frames, along with non-stationary ones. It should

also be noted that – because of the possibility of inconsistency – multiplicity of naively optimal frames also leads to a multiplicity of induced utilities. In particular, the agents facing multiple naively optimal strategies believe it does not matter which strategy they choose, as they expect they will stick to those strategies – but they can be wrong.

As an example, consider the decision problem on Figure 3.5. Discounting is quasi-hyperbolic with  $\beta = \delta = 0.5$ . Since naively optimal strategies are coherent, we will only consider such strategies. There are only two histories,  $h_0 = (\rho)$  and  $h_1 = (\rho, A, \sigma)$ , where the action choice is not trivial, so the naive root agent at  $h_0 = (\rho)$  has to consider only four coherent strategies, which we will denote, by  $s_{AC}^{h_0}, s_{AD}^{h_0}, s_{BC}^{h_0}$  and  $s_{BD}^{h_0}$ . They are defined as:

1.  $s_{AC}^{h_0}(h_0) = (A, A)$  and  $s_{AC}^{h_0}(h_1) = (C, C)$ ;
2.  $s_{AD}^{h_0}(h_0) = (A, A)$  and  $s_{AD}^{h_0}(h_1) = (D, D)$ ;
3.  $s_{BC}^{h_0}(h_0) = (B, B)$  and  $s_{BC}^{h_0}(h_1) = (C, C)$ ;
4.  $s_{BD}^{h_0}(h_0) = (B, B)$  and  $s_{BD}^{h_0}(h_1) = (D, D)$ .

It can be easily seen that from the perspective of the root agent,  $U_i^{h_0}(s_{AC}^{h_0}) = U_i^{h_0}(s_{BC}^{h_0}) = U_i^{h_0}(s_{BD}^{h_0}) = 0 > -\frac{1}{4} = U_i^{h_0}(s_{AD}^{h_0})$ . Thus, the naive root agent is indifferent between choosing action  $A$  first, and  $C$  afterwards, or simply  $B$ .

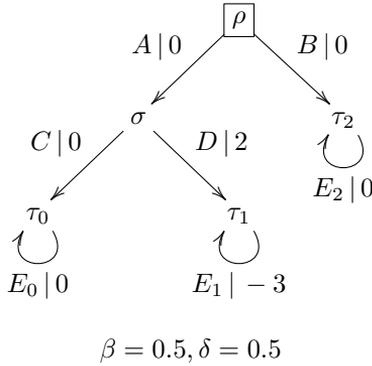


Figure 3.5: Multiplicity of induced utility for naively optimal frames.

Now we focus on the agent at  $h_1$ . She has two strategies available, which we will denote by  $s_C^{h_1}$  and  $s_D^{h_1}$ . They are defined as:

1.  $s_C^{h_1}(h_1) = C$ ;
2.  $s_D^{h_1}(h_1) = D$ .

For the agent at  $h_1$ ,  $U_i^{h_1}(s_D^{h_1}) = 0.5 > 0 = U_i^{h_1}(s_C^{h_1})$ . Therefore, the naive agent at  $h_1$  prefers to choose action  $D$ . Now, consider the two naively optimal frames  $f_{BCD}$  and  $f_{ACD}$ , defined as:

- $f_{BCD}(h_0) = s_{BC}^{h_0}$  and  $f_{BCD}(h_1) = s_D^{h_1}$ ;
- $f_{ACD}(h_0) = s_{AC}^{h_0}$  and  $f_{ACD}(h_1) = s_D^{h_1}$ ;

From the previous calculations, we see that both  $f_{BCD}$  and  $f_{ACD}$  are naively optimal frames. However, the induced utilities are not equal:  $U_r(f_{BCD}) = 0$  and  $U_r(f_{ACD}) = -\frac{1}{4}$ . The underlying reason is as follows: When an agent picks a naively optimal strategy in  $h_0$ , she expects to earn 0 by either choosing strategy  $s_{BC}^{h_0}$  or  $s_{AC}^{h_0}$ . But if she chooses  $s_{AC}^{h_0}$ , after getting to history  $h_1$  – on account of being present-biased – she will not stick to her previous strategy, which would prescribe her choosing action  $C$ ; instead, she chooses  $D$ , which leads to a decrease in her induced utility.

To sum up our discussion of naiveté: If we interpret an optimal frame as predictive for a naive decision maker’s behavior, then in the generic case, we get a unique prediction of actions, intentions and belief for each history; and we expect stationary behavior in realization, and a single prediction for the induced utility. On the other hand, in degenerate cases we can get multiple predictions of actions, intentions and beliefs for some agents; we do not necessarily expect stationary behavior; and we do not necessarily get a unique expectation for induced utility.

### 3.6 Sophistication

Similarly to naiveté, there are several definitions of sophistication:

- optimality under a credibility constraint: following a feasible optimal strategy, or a plan that she will actually follow (Strotz, 1956; Yaari, 1978);
- game-theoretic notion: an intra-personal subgame-perfect equilibrium, sometimes also referred to as ‘Strotz-Pollak equilibrium’ (Peleg and Yaari, 1973; Kocherlakota, 1996; Vieille and Weibull, 2009);

- rational expectations: perfectly anticipating future behavior (O’Donoghue and Rabin, 2001; Gilpatric, 2008);
- self-awareness: being aware of future changes in discount rates and preferences (Hammond, 1976; Heidhues and Kőszegi, 2009; Ali, 2011).

These definitions do not match precisely. For instance, the notion of sophistication as an intra-personal equilibrium does not guarantee the satisfaction of rational expectations in case when multiple such equilibria exist. It is possible that sophisticated agents at  $h$  and  $h'$  with  $h \prec h'$  each choose a strategy  $s^h$ ,  $s^{h'}$  supported by an intra-personal equilibrium, but the action  $b^h(h') \neq b^{h'}(h')$ ; that is, sophistication in itself is no guarantee for consistency.<sup>18</sup> Just like for naiveté, we would like to offer a new interpretation of sophistication. We regard sophistication as primarily a characteristic of agents. In our framework, the defining feature of sophisticated agents is that they first consider their beliefs about the future, and only then do they form intentions. For now, our sophisticated agents assume that all future agents will be sophisticated, too – we will relax this assumption for hybrid agents in Section 3.7.

A sophisticatedly optimal strategy is made up, first, by beliefs about future agents’ choices: She believes each future agent will pick a best response to future agent’s choices. So we implicitly have to consider second-order beliefs, the beliefs of each agent about the beliefs of agents about the future. However, we assume that in a sophisticated strategy, the second-order beliefs coincide with first-order beliefs.<sup>19</sup> Thus, the current agent believes that future agents will believe what she currently believes.

Beliefs and preferences  $\rightarrow$  intentions  $\rightarrow$  action  $\rightarrow$  strategy.

Figure 3.6: The forming of intentions and beliefs by a sophisticated.

The second component of a sophisticatedly optimal strategy concerns the intentions, which are set to match the beliefs; as the agent knows she has no control over her future selves, she can just as well intend future actions that future selves are choosing anyway.

Finally, the current action is chosen to be a best response to future actions.

---

<sup>18</sup> See also case 1.b.ii in Section 3.A.2.

<sup>19</sup> O’Donoghue and Rabin (2001) makes the assumption on second-order beliefs explicit for partially naive agents.

**Definition 30.** A strategy  $\hat{s}^h$  of an agent at  $h$  is *sophisticatedly optimal*, if:

$$\hat{b}^h(h') \in \arg \max_{a \in A_{\omega(h')}} U_b^{h'}(\hat{s}^h[a : h']), \text{ for all } h' \in H^{\succ h},$$

and it is coherent:

$$\hat{i}^h(h') = \hat{b}^h(h'), \text{ for all } h' \in H^{\succ h},$$

where  $\hat{s}^h[a : h']$  denotes the strategy where the action taken at  $h'$  is replaced with action  $a$  in strategy  $\hat{s}^h$ .<sup>20</sup>

A frame  $\hat{f}$  is sophisticatedly optimal if the strategy  $\hat{f}(h)$  is sophisticatedly optimal at each history  $h$ .

One remark about the intention component of a sophisticatedly optimal strategy is in order. It could be argued that instead of intending to give a best response to future beliefs at  $h'$  based on the preferences at  $h'$ , the agent at  $h$  should intend something else at  $h'$ ; namely, to give a best response to the choices of future agents based on the preferences at  $h$ , not the ones at  $h'$ . However, we want to guarantee the coherence of sophisticatedly optimal strategies to ensure that a consistent sophisticated strategy always exists. Moreover, it is psychologically plausible that the sophisticated agent wants to maintain this coherence – indeed, this is why she is reasoning about future agents after all, realizing that the preferences of future agents might be different from her current ones. We will see in Section 3.7 that for hybrid agents intentions and beliefs might not match.

**Theorem 8.** *For any decision problem, there exists a consistent sophisticatedly optimal frame.*

*Proof.* Fix an enumeration of all histories, that is, a bijection  $\rho : \mathbb{N} \rightarrow H$ . Although it is not necessary for the proof, we can assume that the root history is taken first, then all histories at stage 1 are enumerated, then all histories at stage 2, and so on. Let  $\mathcal{A} = \times_{h \in H} A_{\omega(h)}$ , where the product is taken in the order according to  $\rho$ . So, an element  $a = (a^h)_{h \in H}$  of  $\mathcal{A}$  prescribes, for every history  $h \in H$ , an action  $a^h$  for the agent at  $h$ . Take an arbitrary  $\hat{a} \in \mathcal{A}$ .

Now, for every  $n \in \mathbb{N}$ , we construct an  $a_n \in \mathcal{A}$  as follows. For every history  $h$  beyond stage  $n$ , that is, with  $t(h) > n$ , let  $a_n^h = \hat{a}^h$ . Then, we proceed by

<sup>20</sup> When calculating  $U_b^{h'}(s^h)$ , only the payoffs generated by  $s^h$  for the histories succeeding  $h'$  should be taken into account; i.e., we consider the expected utility induced by  $s^h$  for the subtree starting at  $h'$ .

backwards induction. For every history  $h$  at stage  $n$ , that is  $t(h) = n$ , let  $a_n^h$  be an action for the agent at  $h$  that maximizes her utility if all agents that succeed her play the action according to  $a_n$  (or equivalently, according to  $\hat{a}$ ). In general for  $k \in \{1, \dots, n\}$ , if we have defined  $a_n^h$  for all histories  $h$  with  $t(h) > k$ , then for every history  $h$  with  $t(h) = k$ , we choose  $a_n^h$  to be an action for the agent at  $h$  that maximizes her utility if all agents that succeed her play the action according to  $a_n$ .

So, we obtain a sequence  $(a_n)_{n \in \mathbb{N}}$  in the space  $\mathcal{A}$ . Note that  $\mathcal{A}$  with the product topology is compact, and because  $H$  is countable, it is metrizable too. Consequently, it is sequentially compact, which implies that the sequence  $(a_n)_{n \in \mathbb{N}}$  has a subsequence  $(a_{n_k})_{k \in \mathbb{N}}$  which converges to some  $\hat{a} \in \mathcal{A}$ . This means that, for every  $m \in \mathbb{N}$ , there exists an  $K_m \in \mathbb{N}$  such that, for every  $k \geq K_m$ , the actions  $a_{n_k}^h$  and  $\hat{a}^h$  coincide for all histories  $h$  with  $t(h) \leq m$ .

Let  $s_n^h$  be the strategy of the agent at history  $h$  which intends to play the action  $a_n^h$  at all histories  $h' \succeq h$  and believes that these actions will be chosen – that is, a coherent strategy. Similarly, we define the strategy  $\hat{s}^h$  with respect to  $\hat{a}$ . Now consider the frame  $\hat{f}$  that assigns strategy  $\hat{s}^h$  to the agent at  $h$ , for every history  $h$ .

First, we show that  $\hat{f}$  is consistent. Take any  $h, h'$ , the corresponding strategies  $\hat{s}^h$  and  $\hat{s}^{h'}$  and some history  $h''$  with  $h'' \succ h$  and  $h'' \succ h'$ . Then

$$\hat{s}^h(h'') = (\hat{a}^{h''}, \hat{a}^{h''}) = \hat{s}^{h'}(h'').$$

Thus,  $\hat{f}$  is a consistent frame.

We now prove that  $\hat{f}$  is sophisticatedly optimal. For this purpose, consider an arbitrary agent, say at history  $h$ , and an agent at a history  $h' \succeq h$ . By construction, for every  $n \geq t(h')$ , the action  $a_n^{h'}$  maximizes the utility of the agent at  $h'$  if all agents that succeed her play the action according to  $a_n$ . Thus,

$$b_n^h(h') = a_n^{h'} \in \arg \max_{a \in A_\omega(h')} U_b^{h'}(s_n^h[a : h'])$$

and as we have seen,  $s_n^h$  is coherent:

$$i_n^h(h') = b_n^h(h').$$

By taking the limit along the subsequence  $(n_k)_{k \in \mathbb{N}}$  and using continuity, we obtain:

$$\hat{b}^h(h') = \hat{a}^{h'} \in \arg \max_{a \in A_\omega(h')} U_b^{h'}(\hat{s}^h[a : h'])$$

and

$$\hat{i}^h(h') = \hat{b}^h(h').$$

Thus,  $\hat{f}$  is a consistent sophisticatedly optimal frame indeed.  $\square$

Again, we might ask whether there is a unique consistent sophisticatedly optimal frame. In contrast with naively optimal frames, this is not the case even for generic decision problems. In the problem of underinvestment (Section 3.A.2, case 1.b.ii) we get multiple sophisticatedly optimal strategies for each history. Moreover, the same example shows that there might be no stationary sophisticatedly optimal frame when we only allow for pure actions. Also, the induced utilities of optimal frames can differ. Thus, only in some cases do we get a unique sophisticatedly optimal frame, and with it, unique predictions for a sophisticated decision maker's actions, intention, beliefs and induced utility.<sup>21</sup> For these cases Theorem 8 implies that the frame will be consistent, too.

Overall, our understanding of sophistication in terms of sophisticated agents first working through their beliefs, then deriving their intentions is closest to the self-awareness interpretation. However, sophisticated optimality can indeed be regarded as a notion of intra-personal subgame-perfection.<sup>22</sup> As there might be several such equilibria, agents at various histories might pick actions corresponding to different equilibria; thus, there is no a priori guarantee for the satisfaction of rational expectations, or that the subgame-perfect equilibrium chosen by the root agent will actually be followed through. Thus, the equilibrium aspect of sophisticated optimality on the *strategy* level does not imply consistency or stationarity for the sophisticated *frame*. However, the above theorem shows that a consistent sophisticatedly optimal frame exists.

The next natural question concerns the relative advantages of sophistication against naiveté.<sup>23</sup> The most commonly analyzed examples in the literature of dynamic inconsistency are stories of procrastination, impulsiveness, underinvestment, addiction and bingeing behavior. We can classify these problems into two groups: In the first group, there are decisions that concern the execution of a single task, like finishing an academic paper or ending a marriage in a sudden burst

<sup>21</sup> On the non-uniqueness of sophisticatedly optimal strategies, see Phelps and Pollak (1968); Peleg and Yaari (1973); Blackorby et al. (1973). More recently, Vieille and Weibull (2009) show that non-uniqueness is a generic property for hyperbolic discounting, and also give sufficient conditions for uniqueness. For a refinement concept, see Kocherlakota (1996).

<sup>22</sup> This also highlights why in Definition 16 we defined a strategy as an intention-belief pair for *all* future histories, since – as in subgame-perfection – the actions of agents at histories off the optimal path are relevant.

<sup>23</sup> For a the detailed comparison and the calculations, we refer the reader to the Appendix. Here we summarize the main findings of the analysis contained therein.

of anger (procrastination and impulsiveness). In the other conceptual box, we put problems concerning the *repeated* execution of a task, like not saving enough for retirement, and drug addiction (underinvestment and binging). It turns out that we can model these as Markov decision problems with only two states (Figures 3.9 and 3.10), if we assume a decision maker who uses quasi-hyperbolic discounting, arguably the simplest form of non-exponential discounting.

In the Appendix, we derive conditions for the payoffs, the discount factor, and the present-biasedness parameter to classify these problems. We find that in some decision problems involving either a single or a repeated task, namely, impulsivity and binging problems, sophisticates commit the same mistake as naifs, and sophistication brings no benefits. This should not come as a surprise, as we cannot expect a notion of intra-personal subgame-perfect equilibrium to solve all welfare problems of decision makers. In some single and also in some repeated decision problems, however, sophistication brings clear benefits, and from the perspective of the root agent, the induced utility of a sophisticatedly optimal frame can be strictly higher than that of a naively optimal one. Again, this is something to be expected – sophistication would hardly deserve the attention it receives if it never brought any improvement over naiveté.

Coincidentally, among our examples, decision problems in which sophistication is beneficial are also the ones where there exists no sophisticated frame of stationary strategies. Thus, these cases serve as examples to the observation that a sophisticated agent might be forced to use a non-stationary optimal strategy when mixed actions are disallowed. For the underinvestment problem, we also analyze the possibility of using mixed actions, and calculate the mixing probability that leads to a consistent sophisticated frame in stationary strategies.

The final section of the Appendix presents the indulgence problem, discovered by O’Donoghue and Rabin (1999a). This is the simplest known decision problem where naiveté outperforms sophistication from the perspective of the root agent. Thus, the advantages of sophistication are shown not be equivocal across all decision problems, and changing to sophistication might harm the decision maker. We make use of this finding of O’Donoghue and Rabin (1999a) to construct our final example of hybrid decision making in the next section.

### 3.7 Hybrid decision makers

This section introduces a new type of decision maker. So far, we have only considered decision makers who are either always naive, or always sophisticated.

However, such purity is quite rare, perhaps even non-existent in the real world. Even the most naive individual realizes after a while that her intentions might not be credible; and even the most consistently sophisticated individual can slip into wishful thinking about her future actions. Therefore, we try to model this duality of an individual via hybrid types. In contrast to the partially naifs of O'Donoghue and Rabin (2001) who (in the context of quasi-hyperbolic discounting) are aware of future present-biasedness, but underestimate its magnitude, our hybrid decision maker flip-flops between being sophisticated and naive, like Ulysses before and after hearing the sirens sing.

We now extend our model to capture hybrid types. Our type space includes naifs and sophisticates.

**Definition 31.** A *Markov decision problem with agent types* is made up of:

- the set of time periods  $\{0, 1, 2, \dots\}$ ;
- $\Theta \subseteq \Omega \times X$ , where  $\Omega$  is a finite state space and  $X = \{N, S\}$  is the finite type space<sup>24</sup>; we denote a state-type pair by  $\theta$ ; the initial state-type pair is  $\bar{\theta} \in \Theta$ ; the state component is denoted by  $\omega(\theta)$ ; while the type component is denoted by  $x(\theta)$ ;
- a finite and nonempty set of pure actions  $A_\omega$  that the decision maker can choose from in  $\omega$ ;
- a payoff function  $u_\omega : A_\omega \rightarrow \mathbb{R}$  that assigns a payoff to every action in state  $\omega$ ;
- transition probabilities  $m_\theta : A_{\omega(\theta)} \rightarrow \Delta(\Theta)$ , with  $m_\theta(\theta' | a_{\omega(\theta)})$  denoting the probability to transit from the state-type pair  $\theta$  to the state-type pair  $\theta'$  when action  $a_{\omega(\theta)}$  is chosen.

This definition keeps the Markovian properties of the original model, and adds a specification of naiveté or sophistication to each state. Also,  $\Theta$  is common knowledge among the agents.

Our model is quite general. Before moving on to illustrate its use in detail for our motivating example from Section 3.1, we list a few kinds of decision problems for which it could be used:

- exogenous types: All state-type combinations are allowed ( $\Theta = \Omega \times X$ ). Moreover, in each new state, the agent is naive (or sophisticated) with the

---

<sup>24</sup>  $N$  stands for naive, and  $S$  for sophisticated.

same probability:  $m_\theta((\omega', x)|a_{\omega(\theta)}) = m_\theta((\omega'', x)|a'_{\omega(\theta)})$  for all  $\theta, \omega, \omega', x$ . Such a model assumes no correlation between state and type.

- fixed type for each state: Here, we have  $(\omega, x), (\omega, x') \in \Theta \Rightarrow x = x'$ . This is the opposite of the previous scenario, as there is perfect correlation between state and type. It can be easily seen that any Markov decision problem with agent types can be reformulated in this manner by expanding the state space; however, it can make the model less illuminating, possibly hiding structural similarities between the problems faced by a naive and a sophisticated agent.
- deterministic type determination: This requires that for all  $\theta, a_{\omega(\theta)}$ , there is some  $x \in X$ , such that  $\sum_{\omega'} m_\theta((\omega', x)|a_{\omega(\theta)}) = 1$ . This means that whatever the agent chooses, her type (but not necessarily her state) in the next period is fully determined.
- full control over type: For all  $\theta, x \in X$ , there is some  $a_{\omega(\theta)}$  so that  $\sum_{\omega'} m_\theta((\omega', x)|a_{\omega(\theta)}) = 1$ . Each agent can always ensure her type to be whatever she wants for the next period.

Of course, these are merely extreme cases, and many interesting situations lie in the middle, having some, but imperfect correlation between state and type; and giving some, but less than total control for agents over their future types. Indeed, it could be argued that the drinking problem we present below should involve stochastic rather than deterministic type determination. However, for simplicity of analysis, we abstract from such complications.

To understand optimal strategies for hybrid agents, we first have to re-define the notion of history:

**Definition 32.** A *type-dependent history*  $h$  has the form  $h = (\theta_0, a_{\omega(\theta_0)}, \dots, \theta_{t-1}, a_{\omega(\theta_{t-1})}, \theta_t)$ , with:

- $\theta_i \in \Theta$ , for  $i \in \{0, 1, \dots, t\}$ , and  $\theta_0 = \bar{\theta}$ ;
- $a_{\omega(\theta_i)} \in A_{\omega_i}$ , for  $i \in \{0, 1, \dots, t-1\}$ ;
- $m_{\theta_i}(\theta_{i+1}|a_{\omega(\theta_i)}) > 0$ , for  $i \in \{0, 1, \dots, t-1\}$ .

Extending the previous notation,  $x(h)$  refers to the current type. We keep the association between histories and agents – each history now corresponds to an agent, and it also includes the agent's type.

Next, we define optimal type-dependent strategies for the Markov decision problem with agent types. We make use of Definitions 29 and 30 for naively and sophisticatedly optimal strategies. We first present the formal definition, and then explain the intuitions below.

**Definition 33.** A type-dependent strategy  $\check{s}^h$  for a Markov decision problem with agent types is *optimal* at history  $h$ , if it satisfies the following conditions:

- for  $x(h) = N$ :

$$\check{s}^h \in \arg \max_{s \in S^h} U_i^h(s),$$

and

$$\check{b}^h(h') = \check{i}^h(h'), \text{ for all } h' \in H^{\succ h}.$$

- for  $x(h) = S$ :

$$\check{b}^h(h') \in [\arg \max_{s \in S^{h'}} U_i^{h'}(s)](h'), \text{ for all } h' \in H^{\succ h} \text{ with } x(h') = N;$$

$$\check{b}^h(h') \in \arg \max_{a \in A_\omega(h')} U_b^{h'}(\check{s}^{h'}[a : h']), \text{ for all } h' \in H^{\succ h} \text{ with } x(h') = S;$$

and

$$\check{i}^h(h') \in \arg \max_{a \in A_\omega(h')} U_b^{h'}(\check{s}^{h'}[a : h']), \text{ for all } h' \in H^{\succ h} \text{ with } x(h') = N;$$

$$\check{i}^h(h') = \check{b}^h(h'), \text{ for all } h' \in H^{\succ h} \text{ with } x(h') = S.$$

A type-dependent frame  $\check{f}$  is optimal, if  $\check{f}(h)$  is an optimal type-dependent strategy for all  $h$ .

Although this definition is rather lengthy, it captures our basic intuitions for the two types. A naive agent at  $h$  does not reason about future agents, as her intentions determine her beliefs, and as a result, her whole optimal strategy in the standard way. However, a sophisticated agent at  $h$  is able to reason about future agents in the following manner: If a future agent at  $h'$  is naive, then the agent at  $h$  believes the agent at  $h'$  will act in a naive way, maximizing her expected utility based on intentions. If, on the other hand, a future agent at  $h'$  is sophisticated, then the agent at  $h$  (correctly) believes that the agent at  $h'$  will act in a sophisticated way, being able to reason about future agents just as well as  $h$  herself does. So a sophisticated agent at  $h$  intends to choose in a sophisticated

manner at all nodes, giving a best response to the choices of future agents. This implies that the intention and belief component of an optimal type-dependent strategy match for all future histories where the agent is sophisticated, but they might not match for future histories where the agent is naive. Coherence is therefore not a necessary property of optimal type-dependent strategies.

Since naively optimal strategies are, in general, not made up of stationary strategies, the belief-component of an optimal type-dependent frame can also be non-stationary. Moreover, it is easy to see that such a frame is not consistent: the intentions of sophisticated agents will not, in general, correspond to the actions taken by naive agents. In both examples of hybrid decision making we present below, we will see such inconsistencies.<sup>25</sup> The lack of coherence, stationarity and consistency represent the inner conflicts that arise within a hybrid decision maker.

**Theorem 9.** *For any decision problem, there exists a type-dependent optimal frame.*

*Proof.* First, start with determining the intentions and beliefs of naive agents, i.e., those at histories where  $x(h) = N$ . For these, we can simply use the first part of the proof of Theorem 7. So, we have  $\check{s}^h = (\check{i}^h, \check{b}^h)$  defined for all  $h$  with  $x(h) = N$ . Let  $\check{f}(h) = \check{s}^h$  for all such  $h$ .

Moving now to histories at which the agent is sophisticated, our construction is analogous to that of Theorem 8. Let  $\mathcal{A} = \times_{h \in H} A_{\omega(h)}$ , and take an arbitrary  $a = (a^h)_{h \in H} \in \mathcal{A}$  that assigns an action to each history.

Recall that a type-dependent optimal frame will not necessarily be coherent. Therefore, both intentions and beliefs have to be constructed; we start with beliefs. First, transform  $a$  into  $\hat{a}$  by fixing actions assigned to histories where the agent is naive, i.e., where  $x(h) = N$ , so that they are actions corresponding to those agents acting in a naive way:

$$\begin{aligned} \hat{a}^h &= \check{i}^h(h) & , \text{ if } x(h) = N; \\ \hat{a}^h &= a^h & , \text{ if } x(h) = S. \end{aligned}$$

Let  $a_n \in \mathcal{A}$  be defined as follows:  $a_n^h = \hat{a}^h$  for all  $h$  with  $t(h) > n$ , or with  $t(h) \leq n$ , and  $x(h) = N$ . For the remaining histories with  $t(h) \leq N$  and  $x(h) = S$ , we move by backwards induction from  $n$  to 0, and let  $a_n^h$  be the best response to the future actions, which are already all defined.

<sup>25</sup> Recall that when an action is actually taken by an agent, it is both believed and intended by the agent for the current history.

Thus, we obtain a sequence  $(a_n)_{n \in \mathbb{N}}$  in  $\mathcal{A}$ . Since  $\mathcal{A}$  is sequentially compact (see the proof of Theorem 8), the sequence  $a_n$  has a subsequence  $a_{n_k}$  converging to some  $\hat{a} \in \mathcal{A}$ .<sup>26</sup> Thus, for all possible choices of a horizon  $m$ , we can find a  $K_m$  such that  $\hat{a}^h = a_{n_k}^h$  for all  $k \geq K_m$  and  $h$  with  $t(h) \leq m$ .

Now, let the beliefs of a sophisticated agent – i.e., at a history  $h$  with  $x(h) = S$  – be  $\check{b}^h(h') = \hat{a}^{h'}$ .

Finally, we construct the intentions of sophisticated agents. Set sophisticated agents' beliefs about future nodes as:  $\check{i}^h(h') = \check{b}^h(h')$  for all  $h'$  with  $x(h) = x(h')$ . For sophisticated agents' intentions assigned to future naive nodes, we construct  $\bar{a}$  by modifying  $\hat{a}$  in a way that whenever  $x(h) = N$ ,  $\bar{a}^h$  is a best response to the future actions in  $\hat{a}$ . We keep actions at other, sophisticated histories unchanged:  $\bar{a} = \hat{a}$ . We set the intentions of a sophisticated agent to be  $\check{i}^h(h') = \bar{a}^{h'}$ .

We have thus defined both  $\check{b}^h$  and  $\check{i}^h$  for sophisticated agents. Let  $\check{s}^h = (\check{i}^h, \check{b}^h)$  and finally,  $\check{f}(h) = \check{s}^h$  for all  $x(h) = S$ . This completes our construction of a type-dependent optimal frame, as we have provided the strategies assigned to histories where the agent is naive, and also to the ones where she is sophisticated.

We will now review our construction again to confirm that  $\check{f}$  is indeed a type-dependent optimal frame. For  $x(h) = N$ , this is immediate. For  $x(h) = S$ , we will first check the beliefs, and then the intentions.

First, suppose that  $x(h') = N$ . We have:

$$\check{b}^h(h') = \hat{a}^{h'} = \check{i}^{h'}(h') \in \arg \max_{s \in S^{h'}} U_i^{h'}(s)(h'),$$

which is what is required in the definition of  $\check{f}$ . For the other case, take  $x(h') = S$ . Now, from our construction of  $a_n^{h'}$ , for all  $n \geq t(h')$ :

$$\check{b}_n^{h'}(h') = a_n^{h'} \in \arg \max_{a \in A_{\omega(h')}} U_b^{h'}(b_n^h[a : h']).$$

Taking the limit along the subsequence  $n_k$ , using continuity, we get:

$$\check{b}^h(h') = a^{h'} \in \arg \max_{a \in A_{\omega(h')}} U_b^{h'}(b^h[a : h']).$$

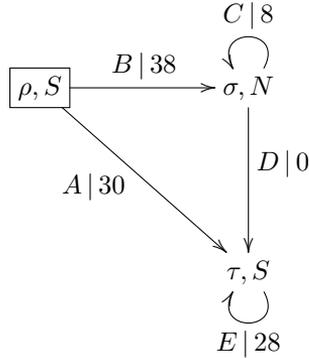
For the intentions of sophisticated agents for future sophisticated nodes, we set these directly to be  $\check{i}^h(h') = \check{b}^h(h')$ . The last thing we need to check is the

<sup>26</sup> There might be multiple subsequences converging to different  $\hat{a}$ -s; in that case, we can select any one of them.

intentions of sophisticated agents for future naive agents. These were defined as:

$$\check{i}^h(h') = \bar{a}^{h'} \in \arg \max_{a \in A_{\omega(h')}} U_b^{h'}(\check{b}^{h'}[a : h']).$$

□



$$\beta = 0.5, \delta = 0.5$$

Figure 3.7: A drinking problem.

We now return to the problem raised in the introduction, displayed in Figure 3.7. The decision maker is sitting in a pub, having finished her second beer (in state  $\rho$ ), and is of type  $S$  (sophisticated). She can either go home directly, by choosing action  $A$ , transiting to state  $\tau$ , where she does not need to make any more decisions. Alternatively, she can drink ‘one more beer’ by choosing action  $B$ . This, however, transitions her to a ‘drunken’ state  $\sigma$ , where she becomes type  $N$  (naive). In this drunken state, she can choose between drinking one more beer by choosing  $C$  (thus, maintaining her drunkenness and staying at  $\sigma$ ), or going home to state  $\tau$  by choosing  $D$ .

Let  $h_0 = ((\rho, S))$  be the root agent. Our goal is to construct  $\check{s}^{h_0}$ , the optimal type-dependent strategy for the root agent. Since  $x(h_0) = S$ , we have to deal with the interesting case, that of a sophisticated root agent.

We start the analysis of this situation by focusing on the state  $\sigma$ . Take any  $h'$  such that  $\omega(h') = \sigma$ . As  $x(h') = N$  if  $\omega(h') = \sigma$ , we get  $\check{b}^h(h') \in \arg \max_{s \in S^{h'}} U_i^{h'}(s)(h')$  according to the definition; the root agent believes that an agent at history  $h'$  with a current state  $\sigma$  is naive. What is the naive choice in

state  $\sigma$ ? It is easy to see<sup>27</sup> that the (subgame-optimal) naively optimal strategy is  $(C, C)(D, D)(D, D) \dots$ , i.e., ‘drink one more beer, and then go home’. Thus,  $\check{b}^h(h') = C$  whenever  $\omega(h') = \sigma$ . The root agent thus believes she would continue drinking after becoming naive.

What about the intentions of the root agent for the agent at  $\sigma$ ? She believes that at history  $h'$ , the continuation actions will be  $C$ . The agent at history  $h'$  could only choose between  $C$  and  $D$ . Going for  $C$  yields  $8 + \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{8}{1-\frac{1}{2}} = 12$ , whereas picking  $D$  gives  $0 + \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{28}{1-\frac{1}{2}} = 14$ . Therefore,  $\check{v}^h(h') = D$  whenever  $\omega(h') = \sigma$ . Together with the result of the previous paragraph, we get that  $\check{s}^h(h') = (D, C)$  whenever  $\omega(h') = \sigma$ . We see that the intentions and the beliefs of the root agent do not match: She would like future agents to pick  $D$  at every  $h'$ , but correctly anticipates that future agents will be unable to do so, and would actually choose  $C$ . The sober, sophisticated root agent realizes that if she drinks just one more beer, she will end up drinking much more than what she actually wishes for.

So what should the root agent choose at  $h$ ? She can pick  $A$ , going home directly, earning her  $U_b^{h_0}(\check{s}^{h_0}[A : h_0]) = 30 + \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{28}{1-\frac{1}{2}} = 44$ . Or, she can pick  $B$ , drink one more beer, and end up still being in the pub. This would earn her  $U_b^{h_0}(\check{s}^{h_0}[B : h_0]) = 38 + \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{8}{1-\frac{1}{2}} = 42$ . Going home seems best. Thus,  $\check{s}^h(h) = (A, A)$ . The optimal type-dependent strategy is:

$$\begin{aligned}\check{s}^h(h) &= (A, A), \\ \check{s}^h(h') &= (D, C), \text{ whenever } \omega(h') = \sigma, \\ \check{s}^h(h') &= (E, E), \text{ whenever } \omega(h') = \tau.\end{aligned}$$

Note that a fully naive root agent would expect that she can resist the temptation of drinking additional bottles of beer, and would expect a utility of  $38 + \frac{1}{2} \cdot \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{28}{1-\frac{1}{2}} = 45$ . The sophisticated root agent realizes that this is unattainable, as the incentives *and the type* of the agents changes by transiting to  $\sigma$ . So in the drinking problem, the sober, sophisticated root agent avoids becoming naive, and thus is better off. Sophistication thus can help avoiding the trap of naiveté. But can sophistication help in avoiding the pitfalls of sophistication?

<sup>27</sup> The decision problem reduced to the states  $\sigma$  and  $\tau$  is that of a procrastination of a single task, discussed in Section 3.A.1, case 1.b. Indeed, with these payoffs,  $8 < (1 - \frac{1}{2}) \cdot 0 + \frac{1}{2} \cdot 28 = 14$ , and  $8 > (1 - \frac{1}{2} \cdot \frac{1}{2}) \cdot 0 + \frac{1}{2} \cdot \frac{1}{2} \cdot 28 = 7$ . So a non-present-biased agent would go home, but a present-biased agent prefers to postpone it.

Section 3.A.3 shows that in some decision problems – in particular, in the indulgence problem – a naive decision maker is strictly better off than a sophisticated one. The psychological intuition behind this is that being aware of one’s inability to resist a temptation in the future can render one unable to resist the temptation in the present, too. This observation allows us to construct the decision problem in Figure 3.8, which we call the indulgence problem with hybrid type.

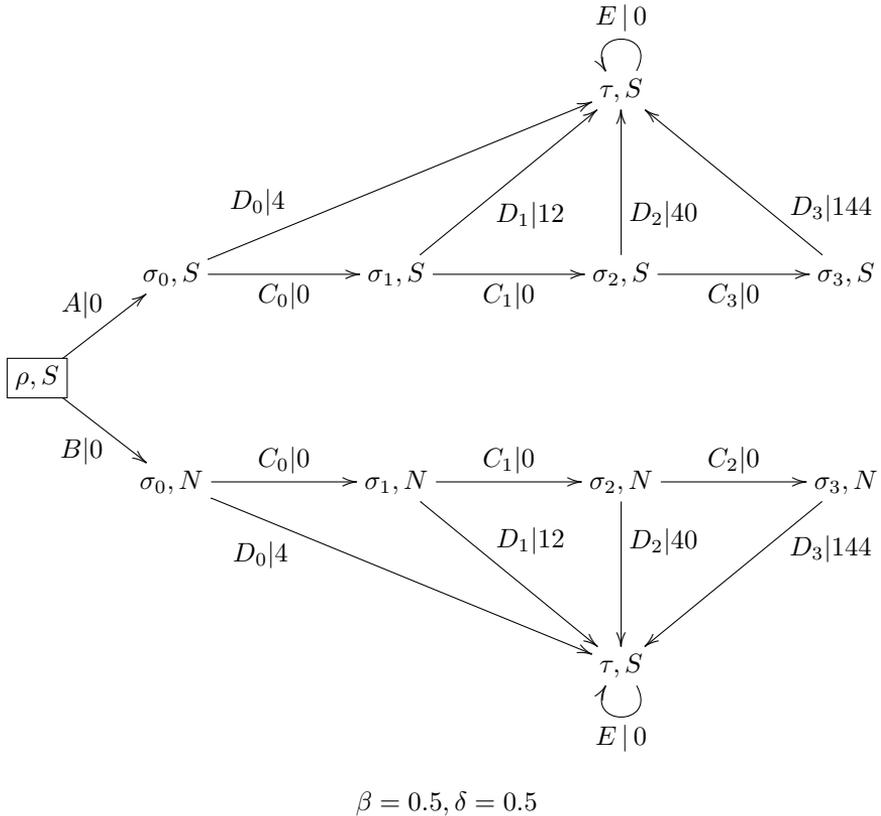


Figure 3.8: Optimal self-deception in the indulgence problem with hybrid type.

The agent at the root history  $h_0 = ((\rho, S))$  is sophisticated. Therefore, she is able to reason about future agents in the following way: at history  $h_1 = ((\rho, S), A, (\sigma_0, S))$ , and at all succeeding histories, she will be sophisticated.

Thus, she concludes – after an analysis similar to the one in Section 3.A.3 – that she will choose the action  $D_0$ . On the other hand, at history  $h'_1 = ((\rho, S), B, (\sigma_0, N))$ , and at all succeeding histories, she will be naive. Thus, she concludes that in the subgame starting at  $h'_1$ , she will choose action  $D_2$ . From the perspective of the root agent:

$$U_b^{h_0}(s^{h_0}[A : h_0]) = \frac{1}{2} \cdot \frac{1}{2} \cdot 4 = 1 < 2.5 = \frac{1}{2} \cdot \left(\frac{1}{2}\right)^3 \cdot 40 = U_b^{h_0}(s^{h_0}[B : h_0]).$$

Therefore, the best response of the root agent is to choose action  $B$ . Notice that this means that a sophisticated agent chooses to face the indulgence problem as a naïf, thereby intentionally causing the agent at  $h'_1$  to have wrong beliefs. In particular, the naïf agent at  $h'_1$  believes that she will be able to wait until the wine fully matures, and take action  $D_3$ . The sophisticated agent at the preceding history  $h_0$  knows that this is not the case, that in fact, action  $D_2$  will be taken. Thus, when choosing an optimal type-dependent strategy, the root agent realizes that she is better off with false beliefs, and decides to deceive herself.

By what means such self-deception might be effectively achieved, or whether it can be achieved intentionally at all is, of course, a difficult problem. But it seems like self-deception has its virtues, which might, in itself, challenge ethical arguments on the inherent immorality of self-deception.<sup>28</sup>

### 3.8 Concluding remarks and future research

This work attempts to play a foundational role for future discourse in multi-self models of dynamic inconsistency. It establishes that the basic epistemic concepts to be considered are beliefs and intentions, and the main levels of analysis should be those of strategies and frames. We would now like to provide some remarks and outline some directions for follow-up research in this area.

An obvious limitation of the current framework is that it only allows for pure actions. This limitation is introduced to ease the presentation, but the technical adaptations required for dealing with mixed actions can be accomplished rather straightforwardly. Mixed actions should play a particularly important role when moving from decision-theoretic models to a game setting.

One might wonder how flexible this model is with regards to increasing the state or action space of the decision problem. Countably infinite states and actions can be allowed for without much difficulty, as long as the set of payoffs

<sup>28</sup> For an overview on the philosophical problems of self-deception, see Deweese-Boyd (2012).

for each action remains compact (and hence, bounded). However, handling continuous time would require a fundamentally different framework, along with a reinterpretation of the notion of ‘agent’.

Whereas our focus was the two most common types of decision makers facing dynamic inconsistency, naifs and sophisticates, there have been arguments in the literature for taking seriously other types as well. In particular, McClennen (1990) argues for the possibility of resolute decision making. Actually, resoluteness can easily be incorporated into our framework. It would be interesting to expand the type space in Markov decision problems with hybrid types to include resolutes.

The horizon of sophisticated decision makers requires further investigation. If agents possess only a finite horizon, reasoning about future agents can be based on two assumptions: Either the *length*, or the *endpoint* of the horizon of that future agent is the same as that of the current agent. In the former case, we are talking about a *moving*, in the latter, about a *fixed* horizon. The implications of these two assumptions on the optimal strategies (generated, for instance, via backward induction) are not yet understood. For example, it seems that a moving horizons approach might be more appropriate to capture the consistent stationary sophisticated strategy that is composed of mixed actions of Section 3.A.2.

Finally, the most interesting application of the framework presented above will be for game theory. How can players reason about the intentions and beliefs of other players, as well as their types? How can one exploit the naiveté (or sophistication) of others? What kind of equilibria are generated when (naive, sophisticated, or hybrid) players are pitted against each other? We hope that through this work, we have broken the ground for such questions.

### 3.A Appendix – Comparison of naifs and sophisticates

This Appendix discusses a few particular decision problems to compare naive and sophisticated decision makers. In Sections 3.A.1 and 3.A.2, our analysis involves simple decision problems with only two states. We show that many classical examples of inconsistent behavior can already be formulated with such decision problems. In Section 3.A.1, we tackle a class of decision problems where the decision maker faces the choice between executing an action eventually or never. We find that in the procrastination problem, but not in the impulsiveness one, sophisticates might outperform naifs, depending on the exact parameters. In Section 3.A.2 we focus on problems where an action can be chosen an arbitrary number of times, and it has both short- and long-term effects. Here, we find similar results: Sophisticates can be better off than naifs in the underinvestment, but not in the binge problem. Although a lot of ground can be covered with just two states, in each of these cases sophisticates are at least as well off as naifs. To show that this is not the case for all decision problems, Section 3.A.3 reviews a decision problem with five states adapted from O’Donoghue and Rabin (1999a), where naiveté dominates sophistication from a welfare perspective.

Common to the analyses is the choice of quasi-hyperbolic discounting for the utility function:

$$U^h(u_{t(h)\rightarrow}) = u_{t(h)} + \beta \sum_{t=t(h)+1}^{\infty} \delta^{t-t(h)} u_t.$$

Here, we assume  $0 < \beta, \delta < 1$ . The main advantages of this particular discount function are that it captures dynamic inconsistency without much technical ado, and also that the parameter  $\beta$  can be interpreted straightforwardly as the present-biasedness factor.

When the discount function is quasi-hyperbolic, discounting is effectively exponential from the next period onwards. Therefore, the analysis of optimal strategies boils down to two questions: Under what conditions will choosing the action be optimal in the future, and under what conditions will it be optimal to perform it right away? Whenever there is a mismatch between these two optimality conditions, the agent’s incentives will change from the current period to the next one, and dynamic inconsistency will arise.

As our examples deal with purely naive or sophisticated decision makers, and the optimal strategies for these types are all defined to be coherent, we

know that for each agent, their beliefs and intentions match whenever they choose optimal strategies. Therefore, we specify only one action for each future history. Whenever we present utility calculations, it should be clear from the agent type whether we are talking about utility based on intentions or beliefs; therefore, we drop the subscript of the utility function. Moreover, notice that our definition of a naively optimal strategy does not require perfection for the strategy – it can specify any action for future agents that will not be reached. To simplify the analysis, we therefore require our naive agents to adopt such strategies.

### 3.A.1 Single tasks – procrastination and impulsiveness

The following decision problems model situations where the decision maker can perform a single task once. The state space contains only two elements: in the – initial – state  $\rho$  the task has not been chosen (yet), while in state  $\sigma$  it has already been performed (see Figure 3.9).

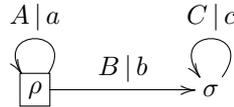


Figure 3.9: Single task problems.

We work with the assumption that whenever the agent is indifferent between choosing  $A$  and  $B$ , she will execute the task by choosing  $B$ . Checking first whether the agent has the incentives to execute the task in the *next* period, we get that she would choose to do so when  $a \leq (1 - \delta)b + \delta c$ , and would not otherwise. This yields two cases:

1.  $a \leq (1 - \delta)b + \delta c$ .

The agent would execute the task in the next period, but will she execute it immediately? If she does, she gains  $b + \beta \frac{\delta}{1-\delta} c$ ; if she does not, she gains  $a + \beta \delta b + \beta \frac{\delta^2}{1-\delta} c$ . This again yields two cases, where after some algebra we get:

$$(a) \quad a \leq (1 - \beta \delta)b + \beta \delta c.^{29}$$

Here, executing the task is the optimal immediate choice, as well as

---

<sup>29</sup> E.g.,  $a = 4$ ,  $b = 0$ ,  $c = 20$ ,  $\beta = \delta = 0.5$ .

the optimal choice for all future periods. Both naively and sophisticatedly optimal strategies prescribe choosing  $B$  at all histories where the end-state is  $\rho$ ; there is no dynamic inconsistency.

- (b)  $a > (1 - \beta\delta)b + \beta\delta c$ .<sup>30</sup>

In this case, executing the task is optimal in the next period, but not immediately. This is the situation of procrastinating a single task.

A naively optimal strategy prescribes postponing  $B$  one period by choosing  $A$  first. A naively optimal frame would involve choosing such a strategy at every history. This implies that in the induced strategy, action  $B$  is in fact never chosen, as each naive agent keeps postponing  $B$  by one period. The induced utility over the naive frame  $\bar{f}$  is thus:  $\bar{U}_r(\bar{f}) = a + \beta \frac{\delta}{1-\delta} a$ .

Although the decision problem looks simple enough, finding the sophisticatedly optimal strategy is not straightforward. Suppose that the sophisticated agent at  $\bar{h}$  believes that no future agent will choose  $B$ . If she chooses to also not do  $B$  herself, she gains  $U^{\bar{h}}(AAA\dots) = a + \beta \frac{\delta}{1-\delta} a$ . On the other hand, if she takes on  $B$ , she gains  $U^{\bar{h}}(B\dots) = b + \beta \frac{\delta}{1-\delta} c$ . It so happens that our two conditions thus far do not determine which of these values is larger.<sup>31</sup> Therefore, we need two more subcases:

- i.  $a + \beta \frac{\delta}{1-\delta} a \geq b + \beta \frac{\delta}{1-\delta} c$ .

This means that choosing  $B$  is a best response to the belief that no future agent would choose  $A$ . Condition 1.b implies that choosing  $A$  is also a best response to the belief that in the next period  $B$  will be chosen. Furthermore, we will now show that choosing  $A$  is also a best reply to the belief that any future agent chooses  $B$ . Suppose the agent at  $\bar{h}$  believes that  $k$  is the earliest period when an agent would choose  $B$ . Then, we get:

$$\begin{aligned} U^{\bar{h}}(AA\dots AB\dots) &= a + \beta \frac{\delta(1-\delta^{k-1})}{1-\delta} a + \beta\delta^k b + \beta \frac{\delta^{k+1}}{1-\delta} c \\ &= a + \beta \frac{\delta}{1-\delta} a + \beta \frac{\delta^k}{1-\delta} (-a + (1-\delta)b + \delta c) \\ &\geq a + \beta \frac{\delta}{1-\delta} a \geq b + \beta \frac{\delta}{1-\delta} c = U^{\bar{h}}(B\dots). \end{aligned}$$

<sup>30</sup> E.g.,  $a = 4$ ,  $b = 0$ ,  $c = 10$ ,  $\beta = \delta = 0.5$ .

<sup>31</sup> For instance, take  $a = -1$ ,  $b = -1$ ,  $c = 4$ ,  $\delta = 0.5$ ,  $\beta = 0.5$ . These values satisfy Conditions 1:  $a \leq (1 - \delta)b + \delta c$ , 1.b:  $a > (1 - \beta\delta)b + \beta\delta c$ , with  $a + \beta \frac{\delta}{1-\delta} a > b + \beta \frac{\delta}{1-\delta} c$ . Changing only the assignment of  $c$  to  $c = 6$ , the parameters still satisfy Conditions 1 and 1.b, but now  $a + \beta \frac{\delta}{1-\delta} a < b + \beta \frac{\delta}{1-\delta} c$ . Thus Conditions 1.b.i/1.b.ii are independent from 1 and 1.b.

Note that we only get equality if both Conditions 1 and 1.b.ii are satisfied with equality. It can be derived that such a scenario is impossible, given Conditions 1.b and  $\beta < 1$ . Therefore, at least one of them is a strict inequality. This implies that  $A$  is a best response to any belief about the future, and the only sophisticatedly optimal strategy for an agent is believing and intending  $A$  for every future history. Thus, in this case, there is no difference between the induced strategy of the naively and sophisticatedly optimal frame, and the induced utilities are equal.

- ii.  $a + \beta \frac{\delta}{1-\delta} a < b + \beta \frac{\delta}{1-\delta} c$ .<sup>32</sup>

Choosing  $B$  is the best response for an agent that believes all future agents will choose  $A$ , and thus the strategy ‘all- $A$ ’ is not sophisticatedly optimal. In what follows, we show that if the moment at which  $B$  is taken in the future is sufficiently far, it is optimal to choose  $B$  immediately.

Take the strategies  $s_k^{\bar{h}}$  and  $s_{k+1}^{\bar{h}}$ , where the index  $k \geq 1$  indicates the first period in which action  $B$  is taken.<sup>33</sup> For the utilities generated by these two strategies, we get that:

$$\begin{aligned} \Delta_k &= U^{\bar{h}}(s_k^{\bar{h}}) - U^{\bar{h}}(s_{k+1}^{\bar{h}}) \\ &= \left( a + \beta \frac{\delta}{1-\delta} (1 - \delta^{k-1}) a + \beta \delta^k b + \beta \frac{\delta^{k+1}}{1-\delta} c \right) \\ &\quad - \left( a + \frac{\beta \delta}{1-\delta} (1 - \delta^k) a + \beta \delta^{k+1} b + \beta \frac{\delta^{k+2}}{1-\delta} c \right) \\ &= \beta \delta^k (-a + (1 - \delta)b + \delta c) \geq 0. \end{aligned}$$

So we find, by Condition 1, that  $U^{\bar{h}}(s_k^{\bar{h}})$  is decreasing in  $k$ .

Taking the difference between the utility generated by strategy  $s_1^{\bar{h}} = AB\dots$  and an arbitrary strategy with  $k > 1$ , we get:

$$U^{\bar{h}}(s_1^{\bar{h}}) - U^{\bar{h}}(s_k^{\bar{h}}) = \sum_{i=1}^{k-1} \Delta_i = \beta \delta \frac{1-\delta^{k-1}}{1-\delta} (-a + (1 - \delta)b + \delta c).$$

<sup>32</sup> We ignore the possibility of mixed actions here, and instead choose to discuss it only for the underinvestment problem. Moreover, we assume that when a sophisticated decision maker can attain consistency, she will do so. For both issues, see Section 3.A.2 Condition 1.b.ii below.

<sup>33</sup> E.g.,  $s_1^{\bar{h}} = AB\dots$ ,  $s_2^{\bar{h}} = AAB\dots$

Hence, we obtain:

$$\begin{aligned} \lim_{k \rightarrow \infty} U^{\bar{h}}(s_k^{\bar{h}}) &= U^{\bar{h}}(s_1^{\bar{h}}) - \lim_{k \rightarrow \infty} (U^{\bar{h}}(s_1^{\bar{h}}) - U^{\bar{h}}(s_k^{\bar{h}})) \\ &= a + \beta\delta b + \beta \frac{\delta^2}{1-\delta} c - \beta \frac{\delta}{1-\delta} (-a + (1-\delta)b + \delta c) \\ &= a + \beta \frac{\delta}{1-\delta} a. \end{aligned}$$

We know from Condition 1.b that  $U^{\bar{h}}(s_1^{\bar{h}}) = U^{\bar{h}}(AB\dots) > U^{\bar{h}}(B\dots) = U^{\bar{h}}(s_0^{\bar{h}})$ . However, we see that  $\lim_{k \rightarrow \infty} U^{\bar{h}}(s_k^{\bar{h}}) = a + \beta \frac{\delta}{1-\delta} a < b + \beta \frac{\delta}{1-\delta} c = U^{\bar{h}}(B\dots)$ . Therefore, as  $k$  grows larger, playing  $B$  will at some point become the best response to believing that  $A$  will be played until period  $k$ . In particular, there exists a smallest such  $\hat{k}$ , for which  $U^{\bar{h}}[A : s_{\hat{k}-1}^{\bar{h}}] > U^{\bar{h}}[B : s_{\hat{k}-1}^{\bar{h}}]$ , but  $U^{\bar{h}}[A : s_{\hat{k}}^{\bar{h}}] \leq U^{\bar{h}}[B : s_{\hat{k}}^{\bar{h}}]$ .

Thus, an optimal strategy requires an agent to choose  $A$ , if she believes  $B$  will be chosen after at most  $\hat{k} - 1$  periods, but pick  $B$  otherwise. This leads to cyclical beliefs. Table 3.4 represents a frame with such strategies in a compact manner: Each row represent a strategy for which the row label identifies the agent's temporal distance from the root agent. Similarly, each column identifies the history that is a certain temporal distance away from the root agent.

		$t(h) - t(h_0) =$						
		0	1	2	3	4	5	...
$t(\bar{h}) - t(h_0) =$	0	$\boxed{A}$	$A$	$B$	$A$	$A$	$B$	...
	1	-	$\boxed{A}$	$B$	$A$	$A$	$B$	...
	2	-	-	$\boxed{B}$	$A$	$A$	$B$	...
	3	-	-	-	$\boxed{A}$	$A$	$B$	...
	4	-	-	-	-	$\boxed{A}$	$B$	...
	5	-	-	-	-	-	$\boxed{B}$	...
	...	-	-	-	-	-	-	...

Table 3.4: A consistent sophisticatedly optimal frame for the procrastination problem, Condition 1.b.ii.

It follows that a sophisticated agent will choose a non-stationary,

cyclical strategy. When precisely she will pick  $B$  depends on the beliefs of the root agent, but a consistent sophisticatedly optimal frame will definitely induce action  $B$  to be chosen sometime in the first  $\hat{k}$  periods (that is, up to period  $\hat{k} - 1$ ). It can be checked that the expected utility of the induced strategy of the consistent sophisticatedly optimal frame will be higher than that of the naively optimal one. Thus, a sophisticated decision maker will be better off in the procrastination problem.

2.  $a > (1 - \delta)b + \delta c$ .

This condition means that the agent has no incentives to execute the task in the next (and subsequent) periods. Thus, her utility is  $b + \beta \frac{\delta}{1-\delta} c$  if she decides to perform it immediately, while just  $a + \beta \frac{\delta}{1-\delta} a$  if she does not do so (thus not performing the task at all). The two sub-cases are:

(a)  $(1 - \delta + \beta\delta)a \leq (1 - \delta)b + \beta\delta c$ .<sup>34</sup>

Here, the agent has the incentives to do it right away given that no agent would execute the task in a later period. A naively optimal strategy simply prescribes choosing  $B$  right away, generating a utility of  $\bar{U} = b + \beta \frac{\delta}{1-\delta} c$  for the naive agent.

The sophisticatedly optimal strategy depends on the beliefs of the agent. Because of Condition 2.a, choosing  $B$  right away is optimal, if the agent believes that no future agent will take it. Now, suppose the sophisticated agent at  $\bar{h}$  believes that some future agent will pick  $B$ . Let  $k$  be the earliest future period in which  $B$  will be chosen, according to the strategy  $s^{\bar{h}}$ . We get that:

$$\begin{aligned} U^{\bar{h}}[B : s^{\bar{h}}] &= b + \frac{\beta\delta}{1-\delta} c \geq \frac{1-\delta+\beta\delta}{1-\delta} a \\ &= a + \frac{\beta\delta}{1-\delta} (a - \delta^{k-1}a + \delta^{k-1}a) \\ &> a + \frac{\beta\delta}{1-\delta} ((1 - \delta^{k-1})a + (1 - \delta)\delta^{k-1}b + \delta^k c) \\ &= U^{\bar{h}}[A : s^{\bar{h}}] \end{aligned}$$

We conclude that taking  $B$  right away is optimal, whatever the agent's beliefs. This means that the sophisticated decision maker cannot avoid impulsivity, and that there is no difference between the

<sup>34</sup> E.g.,  $a = -4$ ,  $b = 0$ ,  $c = -10$ ,  $\beta = \delta = 0.5$ .

utility gained by naifs and sophisticates. In either case, if the decision maker later reflects on this decision from a non-present-biased perspective, she will regret taking  $B$ .

(b)  $(1 - \delta + \beta\delta)a > (1 - \delta)b + \beta\delta c$ .<sup>35</sup>

Now, picking  $A$  and staying in  $\rho$  is always optimal both for naifs and sophisticates, and dynamic inconsistency does not arise.

This shows that for single task problems, there is no difference in the behavior of sophisticated and naive agents for the problem of impulsiveness. In the procrastination problem, it can happen that sophistication brings no benefits, but – depending on the parameters of the problem – it might also be the case that it overperforms naiveté.

The asymmetry found between problems of impulsiveness and procrastination might look counter-intuitive at first, as the structure of the Markov decision problem looks quite similar. However, our finding is consistent with Thaler (1981), who presents some behavioral evidence that implicit discount rates are much higher for gains than for losses. If we interpret taking an impulsive action as a choice of an immediate reward, whereas taking the action that is to be procrastinated as accepting an immediate loss, our result that sophisticated decision makers can overcome the problem of procrastination are in line with their findings.

### 3.A.2 Repeated tasks – underinvestment and binges

The following decision problems are very similar to the previous ones in that there are two states, with only one of them requiring a decision, whether to perform a task or not. The main difference is that the task can be executed repeatedly, as the decision maker always returns to the initial state. Figure 3.10 presents the parametrized version of the repeated task problem.

To simplify the analysis, we restrict our attention to strategies where the action assigned to a history  $h$  can be conditioned only on the current time and state, i.e., Markovian strategies. This allows for a very compact representation of coherent strategies, namely, as a sequence of  $A$ -s and  $B$ -s. For an agent at  $\bar{h}$  the  $k^{th}$  element in such a sequence specifies what intention-belief pair the agent assigns to period  $k - 1$ . For instance, when we write  $s^{\bar{h}} = ABAA\dots$  it can be read as: ‘The agent at  $\bar{h}$  (in state  $\rho$ ) intends and believes  $A$  to be her current action, intends and believes action  $B$  if she is in state  $\rho$  after one period, etc.’

<sup>35</sup> E.g.,  $a = 4$ ,  $b = 0$ ,  $c = 0$ ,  $\beta = \delta = 0.5$ .

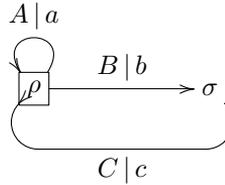


Figure 3.10: Repeated task problem.

We also assume that whenever the agent is indifferent between the two actions available in state  $\rho$ , she will execute task  $B$ .

Starting with the *next* period, the agent has the incentives to pick  $B$  whenever  $a + \delta a \leq b + \delta c$ . We get the following cases:

1.  $a + \delta a \leq b + \delta c$ .

The agent will prefer executing  $B$  from the second period onwards. Whether she prefers to do it right away depends on the relationship between the payoff of picking  $A$  immediately, and then starting to do  $B$ ,  $a + \beta \frac{\delta}{1-\delta^2}(b + \delta c)$ , or starting to execute  $B$  right away,  $b + \beta \delta c + \beta \frac{\delta^2}{1-\delta^2}(b + \delta c)$ . After some algebra, we get the following two cases:

- (a)  $(1 + \delta)(a - b) \leq \beta \delta(c - b)$ .<sup>36</sup>

Here  $B$  is optimal whenever the state is  $\rho$ , regardless whether the agent is naive or sophisticated. Dynamic inconsistency does not arise.

- (b)  $(1 + \delta)(a - b) > \beta \delta(c - b)$ .<sup>37</sup>

When this is the case, the decision revolves around choosing an action that brings benefits in the longer run, but these benefits are not sufficient to offset immediate costs when the agents are present-biased. Depending on the particular interpretation, this can be called a problem of slacking or underinvestment.

Let us focus on a naive decision maker first, supposing she is at history  $\bar{h}$ . Our Condition 1 implies that the optimal action for all future histories is  $B$ , while Condition 1.b implies that the optimal action for the current period is  $A$ . Therefore, the unique naively optimal strategy is  $\tilde{s}^{\bar{h}} = ABBBBB\dots$ , irrespective of  $\bar{h}$ . The naively optimal frame is thus  $\tilde{f}(h) = ABBBBB\dots$  for all  $h$ . This frame

<sup>36</sup> E.g.,  $a = 0$ ,  $b = 0$ ,  $c = 4$ ,  $\beta = \delta = 0.5$ .

<sup>37</sup> E.g.,  $a = 4$ ,  $b = 0$ ,  $c = 16$ ,  $\beta = \delta = 0.5$ .

is represented in a succinct form in Table 3.5. Each row represents a set of strategies for which the row label identifies the current agent's temporal distance from the root agent. Similarly, each column identifies histories that are a certain temporal distance away from the root agent. From the representation it should be both clear that the naively optimal frame is inconsistent and that the induced naive strategy is  $\Lambda(\tilde{f}) = AAAAAA\dots$ , which would generate a utility of  $U_r(\tilde{f}) = a + \beta \frac{\delta}{1-\delta} a$  for the (naive) root agent.

		$t(h) - t(h_0) =$					
		0	1	2	3	4	...
$t(\bar{h}) - t(h_0) =$	0	A	B	B	B	B	...
	1	-	A	B	B	B	...
	2	-	-	A	B	B	...
	3	-	-	-	A	B	...
	4	-	-	-	-	A	...
	...	-	-	-	-	-	...

Table 3.5: Naively optimal frame for the underinvestment problem, Condition 1.b.

The analysis of the sophisticated decision maker is more complicated. Let us start with checking whether  $A$  or  $B$  is the best response of a sophisticated agent at some  $\bar{h}$  whose belief is that in the next period action  $A$  will be chosen. This implies that after two periods, the current state will be  $\rho$ , so the utility generated by the rest of strategy ( $x$ ) is irrelevant. The agent at  $\bar{h}$  needs to compare  $U^{\bar{h}}(AA\dots) = a + \beta\delta a + x$  and  $U^{\bar{h}}(BA\dots) = b + \beta\delta c + x$ . It turns out that Conditions 1 and 1.b do not determine which of the two is larger.<sup>38</sup>

- i.  $a + \beta\delta a > b + \beta\delta c$ .

This means that  $U^{\bar{h}}(AA\dots) > U^{\bar{h}}(BA\dots)$ ; thus, if any  $A$  appears in the optimal strategy, it must be preceded by  $A$ -s. So, any candidate for a sophisticatedly optimal strategy has the following form:  $A\dots AB\dots B$ . However, because of Condition 1.b above,  $U^{\bar{h}}(ABBBBB\dots) > U^{\bar{h}}(BBBBBB\dots)$  and con-

<sup>38</sup> To show this, take  $a = -1$ ,  $b = -1.6$ ,  $c = 0$ ,  $\delta = 0.75$ ,  $\beta = 0.75$ . These values satisfy Conditions 1:  $a + \delta a \leq b + \delta c$ , and also 1.b:  $(1 + \delta)(a - b) > \beta\delta(c - b)$ , with  $a + \beta\delta a > b + \beta\delta c$ . Changing only the assignment of  $b$  to  $b = -1.5$ , the parameters still satisfy Conditions 1 and 1.b, but now  $a + \beta\delta a < b + \beta\delta c$ . Thus, Conditions 1.b.i/1.b.ii are independent from 1 and 1.b.

sequently there is only one candidate of such a strategy: all- $A$ . As we have already seen,  $A$  is a best-response whenever in the next period  $A$  will be played. Therefore, for any  $\bar{h}$  the sophisticatedly optimal strategy is  $\hat{s}^{\bar{h}} = AAAAAA\dots$ . While the sophisticated decision maker avoids inconsistency, her behavior will be the same as that of naive decision maker, also earning  $U_r(\hat{f}) = U_r(\tilde{f}) = a + \beta \frac{\delta}{1-\delta} a$ .

ii.  $a + \beta\delta a \leq b + \beta\delta c$ .

Now, we get  $U^{\bar{h}}(BA\dots) \leq U^{\bar{h}}(AA\dots)$ , so  $B$  is the best response to choosing  $A$  in the next period. What about one period before  $B$ , i.e., is  $U^{\bar{h}}(ABA\dots) > U^{\bar{h}}(BBA\dots)$ ? This is indeed the case – to show it, we again need to realize that since the strategy prescribes action  $A$  in the second period from the current time, the sophisticated agent knows that the current state three periods from now will be  $\rho$ , and that therefore the utility generated by the rest of the strategy, denoted by  $x$ , is irrelevant to this comparison. We get:

$$\begin{aligned}
 U^{\bar{h}}(ABA\dots) &= a + \beta\delta b + \beta\delta^2 c + x \\
 &= a - b + b + \beta\delta b + \beta\delta^2 c + x \\
 &> \beta \frac{\delta}{1+\delta} (c - b) + b + \beta\delta b + \beta\delta^2 c + x \\
 &= \beta \frac{\delta}{1+\delta} (c + \delta b + \delta c + \delta^2 c) + b + x = \\
 &= \beta \frac{\delta}{1+\delta} (\delta b + \delta^2 c) + b + \beta\delta c + x \\
 &\geq b + \beta\delta c + \beta\delta^2 a + x \\
 &= U^{\bar{h}}(BBA\dots).
 \end{aligned}$$

This implies that  $A$  is the best response to playing  $B$  next period. But as we are in Condition 1.b.ii,  $B$  itself is a best response to playing  $A$  next period. So, for each  $\bar{h}$  we have two candidates<sup>39</sup> for a sophisticatedly optimal strategy:  $\hat{s}_{AB}^{\bar{h}} = ABABAB\dots$  and  $\hat{s}_{BA}^{\bar{h}} = BABABA\dots$ . Notice that none of these is a stationary strategy. It is then clear that a sophisticatedly optimal frame should be made up of strategies  $\hat{s}_{AB}$  or  $\hat{s}_{BA}$ . Two consistent sophisticatedly optimal frames can be made up of these, one being represented in Table 3.6, the other having the entries  $A$  and  $B$  exchanged. The induced strategies are  $\hat{s}_{AB}$  and  $\hat{s}_{BA}$ . But

<sup>39</sup> Just like in the previous case, strategy  $BBBBBB\dots$  is ruled out by Condition 1.b.

from the perspective of the root agent, these induced strategies generate different utilities:  $U^{h_0}(\hat{s}_{AB}) = a + \beta \frac{\delta}{1-\delta^2}(b + \delta c) > b + \beta \delta c + \beta \frac{\delta^2}{1-\delta^2}(b + \delta c) = U^{h_0}(\hat{s}_{BA})$ . However, it can be easily checked that both of these generate a higher payoff than the induced strategy of the naive agent.

		$t(h) - t(h_0) =$					
		0	1	2	3	4	...
$t(\bar{h}) - t(h_0) =$	0	A	B	A	B	A	...
	1	-	B	A	B	A	...
	2	-	-	A	B	A	...
	3	-	-	-	B	A	...
	4	-	-	-	-	A	...
	...	-	-	-	-	-	...

Table 3.6: A consistent sophisticatedly optimal frame for the underinvestment problem, case 1.b.ii.

Not only is there no unique consistent sophisticatedly optimal frame, there are numerous other *inconsistent* frames.<sup>40</sup> The problem is that the beliefs of the agents are underdetermined by the best-response condition imposed by sophistication. In fact, we can create any induced strategy for the sophisticated decision maker by manipulating these agents' beliefs. For example, since  $U^{\bar{h}}(\hat{s}_{AB}) > U^{\bar{h}}(\hat{s}_{BA})$  for all  $\bar{h}$ , maybe each sophisticated agent believes that the agent in the next period will pick *B*, so that they should choose *A* at  $\bar{h}$ !

We see that sophistication is no guarantee for either the existence of pure stationary strategies, uniqueness of an induced sophisticated strategy, or unique payoff for the root agent. It seems that some kind of consistency of beliefs across agents is necessary to acquire a unique sophisticatedly optimal frame, and thereby a unique induced strategy. But even this might not guarantee stationarity.

One way to get a consistent sophisticatedly optimal frame of stationary strategies is to allow for randomized actions.<sup>41</sup> In our

<sup>40</sup> For example, consider the stationary frame  $\hat{f}$  with  $\hat{f}(h) = ABAB \dots$  for all  $h$ .

<sup>41</sup> Note that this is a short detour from our original framework, where we only allowed for

strategy, at each history with current state  $\rho$ , action  $A$  is played with probability  $p$ , and  $B$  with probability  $1 - p$ . Take an agent at history  $\bar{h}$ , with  $\omega(\bar{h}) = \rho$ . First, calculate the expected utility from the next period onwards in state  $\rho$ , denoted by  $x = x(p)$ :

$$x = p(a + \delta x) + (1 - p)(b + \delta c + \delta^2 x).$$

We also know that the agent should be indifferent between playing  $A$  and  $B$  in  $\bar{h}$ , since she mixes between them. Therefore, we have:

$$a + \beta\delta x = b + \beta\delta c + \beta\delta^2 x.$$

Solving this system of equations for  $p$ , we get:

$$p = \frac{(1 + \delta)(a - b) - \beta\delta(c - b)}{\delta(1 - \beta)(a - b)}.$$

It can be shown that with the given conditions,  $p$  always falls in the  $[0, 1]$  interval. By using mixed actions, we can thus construct a consistent stationary sophisticatedly optimal frame, which involves mixing between  $A$  and  $B$  with the given probability  $p$  at all histories  $h$  with  $\omega(h) = \rho$ . Despite the technical and conceptual difficulties, our result is a positive one for sophistication: the sophisticated decision maker can earn a higher payoff than the naive one in the underinvestment problem, given that the parameters fall into the right range.

2.  $a + \delta a > b + \delta c$ .

In this case, the agent will prefer picking  $A$  from the second period onwards. She will prefer to execute  $B$  immediately whenever  $a + \beta\delta a \leq b + \beta\delta c$ .

(a)  $a + \beta\delta a \leq b + \beta\delta c$ .<sup>42</sup>

Here, the agent has incentives to choose  $B$  once immediately, but never thereafter. A shopping binge or the consumption of an addictive substance might serve as examples for such a decision problem: present-biasedness induces the enjoyable action  $B$ , for which the costs  $c$  only need to be paid later. A naive agent thus chooses  $B$  at the outset, but intends to refrain from choosing it again later. Obviously,

---

pure actions.

<sup>42</sup> E.g.,  $a = -4$ ,  $b = 0$ ,  $c = -16$ ,  $\beta = \delta = 0.5$ .

the frame generated by such strategies will be inconsistent, and the naive decision maker will stay addicted or keep on binging, executing  $B$  whenever in state  $\rho$ . The induced utility of this naively optimal frame is  $U_r(\tilde{f}) = b + \beta\delta c + \beta \frac{\delta^2}{1-\delta^2} (b + \delta c)$ .

For finding sophisticatedly optimal frames, first consider an agent at history  $h$  who believes the agent in the next period in state  $\rho$  would choose  $A$ . This implies that regardless of what the agent at  $h$  chooses, the current state will be  $\rho$  two periods from now. As the payoffs from that period onwards are fixed, the agent at  $h$  only needs to contrast the payoffs gained in this and the next period. By choosing  $A$ , she would gain  $a + \beta\delta a$ ; by choosing  $B$ , she would gain  $b + \beta\delta c$ . With our assumption on the parameters, it is optimal for the agent at  $h$  to choose  $B$ . This means that in a sophisticated strategy, an action  $A$  is always preceded by an action  $B$ .

Now, take an agent at a history  $\bar{h}$  that chooses action  $A$  at that history. There are two possibilities: either her strategy does not assign action  $A$  to any future history where the current state is  $\rho$ , or it assigns  $A$  to at least one. Suppose the first possibility; then, the strategy of the agent assigns  $B$  to all future histories where the current state is  $\rho$ . She can then choose between adopting strategy  $s_{\bar{A}}^{\bar{h}} = ABBBBB\dots$  and strategy  $s_{\bar{B}}^{\bar{h}} = BBBBBB\dots$ . Her utilities for these are:

$$\begin{aligned}
 U^{\bar{h}}\left(s_{\bar{B}}^{\bar{h}}\right) &= b + \beta\delta c + \beta \frac{\delta^2}{1-\delta^2} (b + \delta c) \\
 &\geq a + \beta\delta a + \beta \frac{\delta^2}{1-\delta^2} (b + \delta c) \\
 &= \beta \frac{\delta}{1+\delta} (a + \delta a - (b + \delta c)) + a + \beta \frac{\delta}{1-\delta^2} (b + \delta c) \\
 &> a + \beta \frac{\delta}{1-\delta^2} (b + \delta c) \\
 &= U^{\bar{h}}\left(s_{\bar{A}}^{\bar{h}}\right).
 \end{aligned}$$

Thus, if the agent believes that all future agents will choose  $B$ , she also should choose  $B$ . Therefore, one sophisticatedly optimal strategy – based on the belief that all future agents will choose  $B$  – is playing and intending to play  $B$  in all periods.

The second possibility is that the agent at  $\bar{h}$  believes that at least one future agent will choose  $A$ . Denote the earliest such period by  $k + 1$ . The strategy of the agent at  $\bar{h}$  can then be either  $s_{\bar{A}}^{\bar{h}} =$

$ABB\dots BBA\dots$  or  $s_B^{\bar{h}} = BBB\dots BBA\dots$ , where the first and last represented action are separated by  $k$  periods of choosing  $B$ . It is clear that after period  $k + 1$ , the agent will definitely be in state  $\rho$  – therefore, the payoffs from that point on can be ignored for the purposes of choosing an action for the agent at  $\bar{h}$ . We show that  $U^{\bar{h}}(s_B^{\bar{h}}) > U^{\bar{h}}(s_A^{\bar{h}})$ , so that choosing  $B$  is optimal. For simplicity, we assume  $k$  is even; the proof for odd  $k$  is analogous.

$$\begin{aligned}
 U^{\bar{h}}(s_B^{\bar{h}}) &= b + \beta\delta c + \beta\delta^2 \frac{1-\delta^k}{1-\delta^2} (b + \delta c) \\
 &\geq a + \beta\delta a + \beta\delta^2 \frac{1-\delta^k}{1-\delta^2} (b + \delta c) \\
 &> a + \beta\delta \frac{1-\delta^k}{1-\delta^2} (b + \delta c) + \beta\delta^{k+1}a \\
 &= U^{\bar{h}}(s_A^{\bar{h}}).
 \end{aligned}$$

The latter inequality follows from Condition 2. This shows that given the belief that some future agent will choose  $A$ , a sophisticated agent should choose  $B$ , which implies that no sophisticated strategy can include action  $A$ . As we have already shown that  $BBBBB\dots$  is a sophisticatedly optimal strategy, it is clear that sophisticates cannot outperform naifs in the binging problem.

(b)  $a + \beta\delta a > b + \beta\delta c$ .<sup>43</sup>

With this parameter configuration, picking  $A$  is always optimal for both naifs and sophisticates, and there is no issue of dynamic inconsistency.

Repeated task problems with two states involving dynamic inconsistency – i.e., underinvestment and binging problems – show a lot of similarity to single task problems. The dangers of binging behavior and impulsiveness affect sophisticates just as much as they do naifs. On the other hand, depending on the exact parameters, sophisticates can avoid underinvestment, just like they can overcome procrastination. The similarity between underinvestment and procrastination problems is even deeper, as in both cases there might be no sophisticatedly optimal frame of stationary strategies when we allow only for pure actions.

<sup>43</sup> E.g.,  $a = 4$ ,  $b = 0$ ,  $c = 0$ ,  $\beta = \delta = 0.5$ .

### 3.A.3 Indulgence

In the previous two decision problems, the utility induced by a sophisticatedly optimal frame was never strictly lower than that induced by a naively optimal frame. Is this always the case? In this section, we review a decision problem presented by O'Donoghue and Rabin (1999a), displayed in Figure 3.11.

The problem is that of performing a single task in either period 0, 1, 2 or 3 by a present-biased decision maker with  $\beta = 0.5$  and  $\delta = 0.5$ . Think of consuming a bottle of valuable wine that gains in taste for up to three years as it ages, but becomes undrinkable afterwards. Will she indulge in drinking it right away, or will she wait for it to fully mature? In Figure 3.11, 'D' and 'C' stand for 'delay' and 'consume'. We ignore specifying the choice of an action for histories  $h$  with  $\omega(h) = \sigma$ .

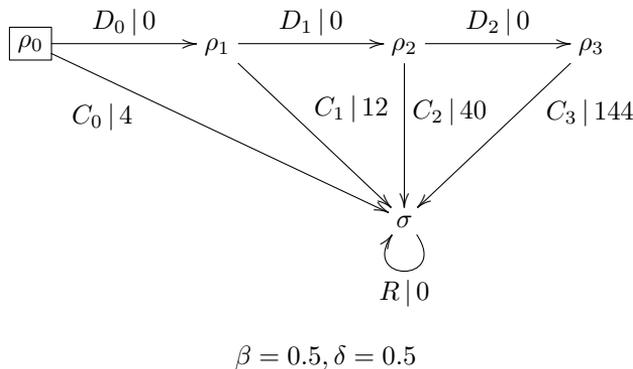


Figure 3.11: Indulgence problem.

The root agent is  $h_0 = (\rho_0)$ . Then, a representation of a coherent strategy for the root agent is a vector of four elements that specifies whether to open the bottle if the wine has not been consumed yet. Since only the first choice of  $C_i$  matters for induced utility, we get four relevant classes of strategies for the root agent:  $S_0^{h_0} = C_0 \dots$ ;  $S_1^{h_1} = D_0 C_1 \dots$ ;  $S_2^{h_2} = D_0 D_1 C_2 \dots$ ;  $S_3^{h_3} = D_0 D_1 D_2 C_3$ .<sup>44</sup>

<sup>44</sup> In order to simplify notation, we introduce the following abbreviations:  $h_0 = (\rho_0)$ ;  $h_1 = (\rho_0 D_0 \rho_1)$ ;  $h_2 = (\rho_0 D_0 \rho_1, D_1 \rho_2)$ . Moreover, we use the following notation for strategies: strategy  $s_t^{h_t'}$  should be read as 'the agent at  $h_t'$  intends and believes to drink the wine in period  $t$ '. We will not make a strict distinction between strategies that are members of the same class – the only relevant element of a strategy is when the earliest period in which some action  $C_i$  (drinking the wine) appears. In Table 3.7 we are even sloppier, and omit the reference to the agent from the indication of a strategy class.

The expected utilities calculated from the perspective of each agent for these strategy classes are presented in Table 3.7. For instance:

$$U^{h_1}(s_3^{h_1}) = 0 + \left(\frac{1}{2}\right)^2 \cdot 0 + \left(\frac{1}{2}\right)^3 \cdot 144 = 18.$$

$U^h(s)$		$h$		
		$h_0$	$h_1$	$h_2$
$s \in$	$S_0 = (C_0, \dots)$	4	–	–
	$S_1 = (D_0, C_1, \dots)$	3	12	–
	$S_2 = (D_0, D_1, C_2, \dots)$	5	10	40
	$S_3 = (D_0, D_1, D_2, C_3)$	9	18	36

Table 3.7: Expected utilities of coherent strategies from the various agents for the decision problem in Figure 3.11.

The naively optimal strategy at  $h_0$  and  $h_1$  is thus choosing  $s_3^{h_0}$  and  $s_3^{h_1}$  respectively, but at  $h_2$  is choosing  $s_2^{h_2}$ . Therefore, for the naively optimal frame  $\tilde{f}$ , we get  $\tilde{f}(h_0) \in S_3^{h_0}$ ,  $\tilde{f}(h_1) \in S_3^{h_1}$ , and  $\tilde{f}(h_2) \in S_2^{h_2}$ . The naive decision maker believes and intends postponing consumption right until the end, but one year before the wine fully matures, she indulges herself. The decision maker thus waits two periods, and this generates a total utility of  $U_r(\tilde{f}) = 5$  for the root agent.

To calculate the sophisticatedly optimal frame, we use backward induction. The sophisticated agent at  $h_2$  chooses  $s_2^{h_2}$ , for the same reasons as the naive agent, since there are no more decisions to make afterwards. Therefore, the sophisticated agent at  $h_1$  can only consider two possibilities:  $s_1^{h_1}$  or  $s_2^{h_1}$ . Since  $U^{h_1}(s_1^{h_1}) = 12 > 10 = U^{h_1}(s_2^{h_1})$ , she chooses to hasten the indulgence, since she correctly believes that she is unable to hold on until the end anyway. Following the same reasoning, the root agent can choose between  $s_0^{h_0}$  and  $s_1^{h_0}$ , and opts for  $s_0^{h_0}$ , consuming immediately. For the sophisticatedly optimal frame  $\hat{f}$ , we get  $\hat{f}(h_0) \in S_0^{h_0}$ ,  $\hat{f}(h_1) \in S_1^{h_1}$  and  $\hat{f}(h_2) \in S_2^{h_2}$ . The sophisticated decision maker indulges herself earlier, precisely because the root agent is able to correctly reason about future agents. As the sophisticated frame generates a utility of  $U_r(\hat{f}) = 4$  for the root agent, we get  $U_r(\hat{f}) > U_r(\tilde{f})$ ; thus, sophistication can leave somebody worse off than naiveté.

This shows that there are decision problems in which a sophisticated decision maker is worse off than a naive one. We emphasize that this is not a new result<sup>45</sup>,

<sup>45</sup> Gruber and Köszegi (2001) also found that sophistication can exacerbate drug consump-

but shows that our framework can easily be put to use to make such welfare comparisons between agents of various types, as well as to generate interesting insights about self-deception 3.7.

Overall, by merely using the framework of quasi-hyperbolic discounting, it is possible to construct extremely simple Markov decision problems that cover most classical examples of dynamically inconsistent behavior. With these models, we can illustrate that sophistication is indeed a double-edged sword: when compared to naiveté, it may provide benefits, but it may also generate welfare losses.

Moreover, our approach shows some interesting aspects of sophistication: Even for decision problems with only two states, it can lead to cyclical beliefs and non-stationary strategies for all agents. For such cases, our method of representing decision makers via tables does indeed provide useful.

---

tion for unaddicted people. Furthermore, in a scenario that involves punishment, sophistication can actually lead to worst possible outcome (Heidhues and Kőszegi, 2009).



## Chapter 4

# Building the Leviathan – Voluntary Centralization of Power Sustains Cooperation

**Abstract.** The prevalence of cooperation among humans is puzzling because cooperators can be exploited by free riders. Peer punishment has been suggested as a key to this puzzle, but cumulating evidence questions its effectiveness in sustaining cooperation. Punishment has to be powerful, meaning that the effect of punishment has to be sufficiently larger than its cost. Since group members can refrain from punishing non-cooperators, peer punishment poses a social dilemma in itself. By exploring a novel experimental setup, we show that the voluntary transfer of punishment power enables groups to overcome the problem of free riding when peer punishment can not. Participants are willing to empower individuals who act in the interest of the group. The endogenous establishment of power centralization solves the social dilemma inherent to peer punishment, sustains cooperation, and increases welfare. Our results could explain why hierarchical power structures are widespread among animals and humans.

## 4.1 Introduction

The scale of cooperation observed among humans constitutes a puzzle for the social and biological sciences. Cooperative efforts bear the risk of exploitation as selfish agents can reap their benefits without contributing to the common good. Nevertheless, cooperation is frequently sustained in human societies (Trivers, 1971, Alexander, 1987, Fehr and Fischbacher, 2003, Nowak and Sigmund, 2005). Peer punishment has been proposed as a possible means to sustain cooperation (Güth et al., 1982, Yamagishi, 1986, Boyd and Richerson, 1992, Ostrom, Walker and Gardner, 1992, Fehr and Gächter, 2000, Fehr and Gächter, 2002, de Quervain et al., 2004, Walker and Halloran, 2004, Falk, Fehr and Fischbacher, 2005). Experiments on public goods dilemmas show that cooperation deteriorates quickly in the absence of sanctioning mechanisms, but can stabilize when free riders can be punished (Yamagishi, 1986, Yamagishi, 1988, Fehr and Gächter, 2000, Fehr and Gächter, 2002, Gächter et al., 2008).

However, peer punishment does not guarantee cooperation. Recent research documented crucial limitations of peer punishment in its ability to sustain cooperation and foster welfare. The cost of punishment has to be sufficiently low relative to its effect on the punished (Masclot et al., 2003, Casari, 2005, Falk, Fehr and Fischbacher, 2005, Egas and Riedl, 2008, Nikiforakis and Normann, 2008, Ohtsuki, Iwasa and Nowak, 2009). Further, non-cooperators sometimes punish cooperators out of spite or retribution, thereby undermining cooperation (Dreber et al., 2008, Herrmann, Thöni and Gächter, 2008, Janssen and Bushman, 2008, Nikiforakis, 2008, Gächter and Herrmann, 2009, Rand et al., 2010, Rand and Nowak, 2011, Dreber and Rand, 2012). The excessive use of punishment can stabilize cooperation but leads to adverse results for group welfare (Masclot et al., 2003, Dreber et al., 2008, Gächter et al., 2008, Janssen, Holahan and Ostrom, 2010, Engelmann and Nikiforakis, in press). Finally, not all group members are willing to participate in the punishment of non-cooperators. It is frequently observed that some, while willing to cooperate, are not willing to punish non-cooperators. Hence, peer punishment produces a second-order social dilemma in which cooperators not willing to punish can second-order free ride on those who do engage in costly punishment (Boyd et al., 2003, Panchanathan and Boyd, 2004, Fowler, 2005, Kiyonari and Barclay, 2008, Perc, 2012).

Social institutions provide an alternative for upholding cooperation through centralized punishment mechanisms (Yamagishi, 1986, Ostrom, 1990, Casari and Plott, 2003, Rockenbach, Guererck and Irlenbusch, 2006, Sefton et al., 2007, Kosfeld, Okada and Riedl, 2009, Boyd, Gintis and Bowles, 2010, Sigmund et al., 2010, Guala, 2012). Laws are issued to tackle ‘tragedies of the commons’, like

problems of over-fishing, littering or air pollution. Contracts are made between individuals to prevent exploitation in mutual agreements like rentals, insurances or investments. Authorities, like courts or the police force, enforce these formal institutions. An essential characteristic of these executive institutions is that they embody a centralized power to punish (Flew, 1954).

Centralized power is also an important feature of groups that are not fully governed by laws or contracts, ranging from hunter-gatherer chiefdoms to modern internet communities. For example, Wikipedia provides a global public good to which everybody can contribute, but only a small share of its editors holds the right to enforce policy and sanction antisocial behaviour.

We propose that this centralization is created and maintained through the voluntary transfer of power to a few trusted individuals. In particular, the transfer of punishment power, fundamental to executive institutions, could solve the problems associated with decentralized peer punishment. We test this hypothesis experimentally by introducing a new experimental setup, which we call the ‘power transfer game’.

## 4.2 Methods

### 4.2.1 The power transfer game

The power transfer game consists of three stages: the power transfer, the contribution, and the punishment stage. In our experiment, participants played the game in groups of five.

In the power transfer stage, group members can give up and transfer punishment power to other group members at no cost. Each group member has a power of 1 at her disposal. Power can be transferred in units of 0.1 and can be distributed among multiple group members. The sum of power kept and received from others determines one’s punishment effectiveness (see below). After all power transfer decisions are made, everyone is informed about how much punishment power each group member has. Power transfer does not change the total punishment power in the group but (may) change its distribution among group members.

In the contribution stage, representing the classic public goods game, participants receive an endowment of 20 monetary units (20 MUs = €0.50) and decide how much to contribute to a ‘group project’. Group members keep any MUs not contributed. The sum of MUs contributed to the group project is then multiplied by 1.5, and distributed equally among all five group members, regardless of

how much each individual contributed. This poses a social dilemma: The return of each contributed MU is 1.5 MU for the group, but only 0.3 MU for the individual. Therefore, if participants are selfish payoff-maximisers, they should not contribute at all. However, if everyone contributes their entire endowment to the group project, participants will earn 30 MUs ( $20 \text{ MUs} \times 5 \text{ group members} \times 1.5 \text{ multiplier} / 5 \text{ group members}$ ), whereas if no one contributes, participants will only earn 20 MUs each. Contribution decisions are made simultaneously. Afterwards, everyone is informed about how many MUs each of the other group members contributed and how many they kept for themselves.

Finally, in the punishment stage, group members are able to punish their peers. Punishment is dealt out by assigning between 0 and 10 punishment points. Punishment decisions are also made simultaneously. Subsequently, group members see how many punishment points each group member assigned, and to whom. Punishment reduces the earnings of both the punisher and the punished. For each assigned punishment point, the punisher pays 1 MU. Importantly, the amount of MUs that are deducted from the punished is determined by the power of the punisher. For example, assume that in the power transfer stage, group member A decided to transfer all of her power to group member B, and no one else transferred any power. B would now have a power of 2, whereas A would have a power of 0. Suppose both A and B decide to punish group member C. Because of her increase in power, each point B uses to punish C will lead to a reduction of 2 MUs in earnings for C. On the other hand, A, who transferred all her power to B, and did not receive any power from others, would have a punishment effectiveness of 0. Thus, A would still pay 1 MU for each punishment point that she assigns to C, but those would not decrease the earnings of C. A game theoretic description and a detailed presentation of the computer interface can be found in Appendix A.

### 4.2.2 Experiment

Participants were recruited from the subject pool of the Behavioural and Experimental Economics lab (BEElab) at Maastricht University and were invited via e-mail. Each experimental session comprised at least 3 and at most 5 groups. Participants ( $n = 350$ , mean age = 21.1,  $SD = 2.6$ , 204 female) were allocated to one of three conditions that differed in how punishment power was determined. The experiment lasted for a total of 20 rounds in fixed groups of five. In the endogenous condition ( $n = 135$ ) participants played the power transfer game described above. Voluntary transfer of power was only possible in this condition.

---

Power transfer decisions made in the previous round served as the default option for the current round. However, participants could freely change their allocation each round.

The remaining participants were allocated to one of two control conditions: the fixed condition ( $n = 80$ ), or the exogenous condition ( $n = 135$ ). In the fixed condition, each participant had a punishment power of 1 and participants were not able to transfer any power. Thus, each group member had a 1:1 effectiveness-to-cost ratio of punishment during the whole experiment. In the exogenous condition, power transfer was not voluntary. Instead, for each group in the endogenous condition a twin group was created in the exogenous condition, in which the history of power transfers and, hence, punishment effectiveness, was mirrored at the individual level. Thus, each group member followed the same change in punishment power across rounds as its twin. The purpose of the exogenous condition is to test whether it is important that power transfer is voluntary. In the endogenous condition, power transfers can make some group members more effective punishers. At the same time, group members are also free to determine who these effective punishers are going to be. The exogenous condition therefore disentangles the effects of increasing punishment effectiveness and voluntarily selecting group members to hold punishment power.

In all three conditions, the different stages of the game were introduced sequentially to the participants (Figure 4.1). The experiment started with a round consisting of only a contribution stage (public goods game). The second round consisted of a contribution and a punishment stage (public goods game with punishment). In the third round, the power mechanism was introduced to the experiment according to the condition. Subsequent rounds followed the structure of the third round.

### 4.3 Results

Participants in all three conditions transferred roughly half of their endowment to the group project in the first round. In the fixed condition with decentralized 1:1 punishment, cooperation decreased steadily (Figure 4.2a; mixed effect regression, round coefficient =  $-0.28$ , 95% CI =  $[-0.51, -0.05]$ ). In contrast, in the endogenous condition, with voluntary transfer of power, initial cooperation was not only sustained, but even increased slightly over time (Figure 4.2a; mixed effect regression, round  $\times$  endogenous condition coefficient =  $0.46$ , 95% CI =  $[0.16, 0.74]$ ). This was not the case for groups in the exogenous condition. Lack-

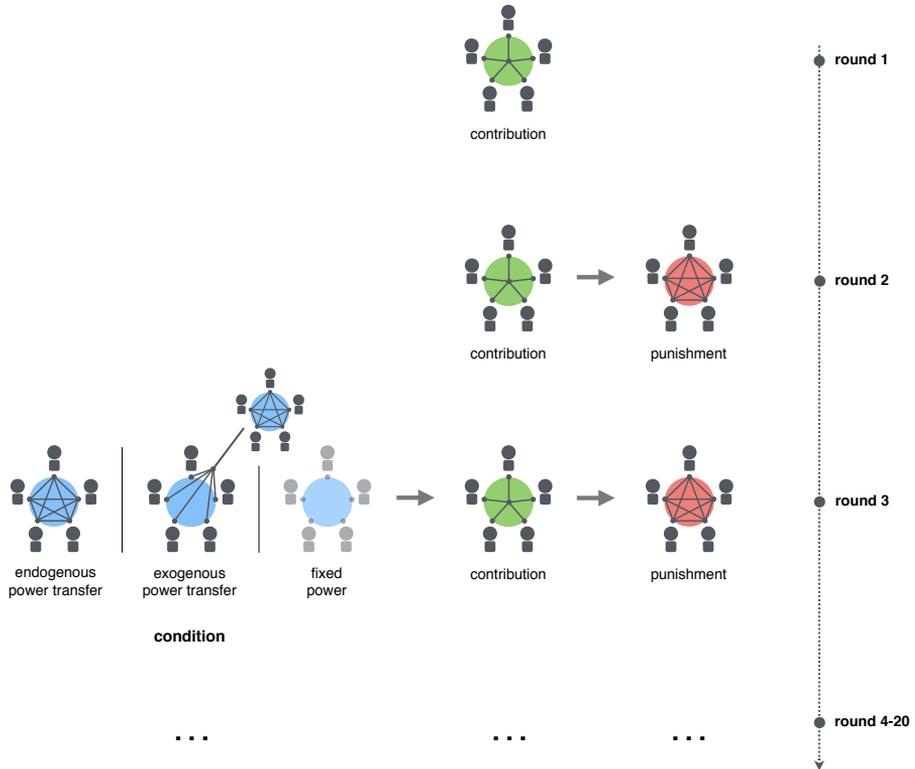


Figure 4.1: Timeline of the experiment. Groups in all three conditions started with one round of a contribution stage, followed by round 2, consisting of a contribution stage and a punishment stage. In round 3, the experimental manipulation was introduced. In the endogenous condition, representing the power transfer game, group members were able to transfer power to other group members before the contribution and punishment stages. Each exogenous condition group mirrored the power transfers of one endogenous condition group and thus group members were not able to transfer power voluntarily. In the fixed condition, power transfers were not possible, and everyone's power was fixed to 1. Rounds 4 to 20 followed the structure of round 3, according to the assigned condition.

ing the freedom to decide whom to transfer power to, these groups showed a decline in cooperation that was not significantly different from that in the fixed condition (Figure 4.2a; mixed effect regression, round  $\times$  exogenous condition coefficient = 0.22, 95% CI = [-0.06, 0.51]). Thus, only the voluntary transfer of power could solve the social dilemma.

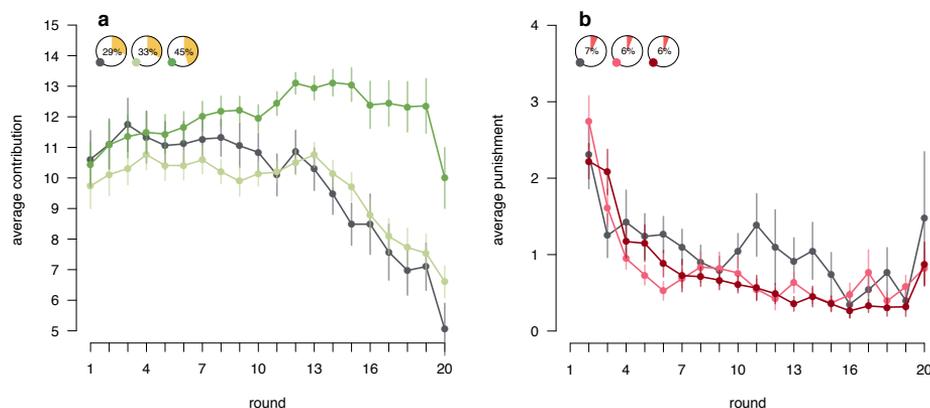


Figure 4.2: Cooperation and punishment over rounds. (a) Mean contributions to the group project for endogenous (dark green), exogenous (light green) and fixed (grey) conditions. Yellow pie charts show overall earnings as a percentage of the social optimum (maximum cooperation without punishment, 30 MUs per group member = 100%), compared to the selfish outcome (minimal cooperation without punishment, 20 MUs per group member = 0%) for each condition. (b) Average amount of MUs spent on punishment in the endogenous (dark red), exogenous (light red), and fixed (grey) conditions. Red pie charts show the average amount of MUs lost due to punishment dealt and received as a percentage of the total earnings for each condition. Error bars show the within-subject standard errors of the mean.

The punishment histories for all three conditions are displayed in Figure 4.2b. Overall, average MUs assigned for punishment declined over the course of the experiment. This decline was the strongest in the endogenous condition (mixed effect regression, round  $\times$  endogenous condition coefficient = -0.07, 95% CI = [-0.13, -0.00]). Higher levels of cooperation, and the more pronounced decline in punishment in the endogenous condition paid off for participants in terms of earnings. Participants with the ability to transfer power earned progressively more compared to participants in the two control conditions (Figure

4.18, mixed effect regression, round  $\times$  endogenous condition coefficient = 1.47, 95% CI = [0.56, 2.41]; difference between round  $\times$  endogenous and round  $\times$  exogenous condition coefficient = 0.79, 95% CI = [0.32, 1.26]). In contrast, there was no significant difference in earnings over rounds between the fixed and exogenous condition (mixed effect regression, round  $\times$  exogenous condition coefficient = 0.68, 95% CI = [-0.24, 1.59]). Thus, voluntary power transfer enabled participants to achieve an outcome much closer to the social optimum, thereby enhancing group welfare.

To understand how groups in the endogenous condition solved the cooperation dilemma, we first looked at the power centralization that emerged over time. Power was already transferred in the first round with a power transfer stage (round 3, see Figure 4.1). A substantial fraction of participants (37%) were willing to transfer power in this round. The amount of power held by the most powerful group member increased significantly over rounds (Figure 4.3a, mixed effect regression, round coefficient = 0.02, 95% CI = [0.00, 0.04]). Thus, power became more and more centralized. Centralization of power was positively related to cooperation in the endogenous condition, but not in the exogenous condition. For each group, we computed the correlation across rounds between power held by the most powerful group member and average cooperation. For groups who could transfer power voluntarily, higher power centralization was associated with higher average group cooperation (Figure 4.3b, mean Pearson's  $r = 0.24$ , one-sample t-test,  $t(25) = 2.9$ ,  $p < .01$ , two-sided). In contrast, for groups in the exogenous condition who experienced exactly the same power centralization but without the ability to transfer power voluntarily, correlations between power centralization and cooperation were not significantly different from zero (Figure 4.3b, mean Pearson's  $r = 0.08$ ; one-sample t-test,  $t(26) = 1.3$ ,  $p = 0.21$ , two-sided).

Next, we analyzed decisions in the endogenous condition on the individual level in order to understand who transferred and who received power, how it was used, and what effect it had on group members. Interestingly, even though participants were unaware of the subsequent introduction of the power transfer mechanism, behaviour in the first two rounds reliably predicted the average power status of a group member: Initial cooperators, i.e., those who contributed at or above the group average in the first round, received significantly more power over the course of the experiment than initial free riders, defined as group members who contributed less than the group average (Mann-Whitney U-test,  $U = 2847.5$ ,  $p < .01$ , two-sided). Similarly, group members who punished free riders in the first punishment stage (round 2) received significantly more power from other group members than those who were not willing to punish (Mann-

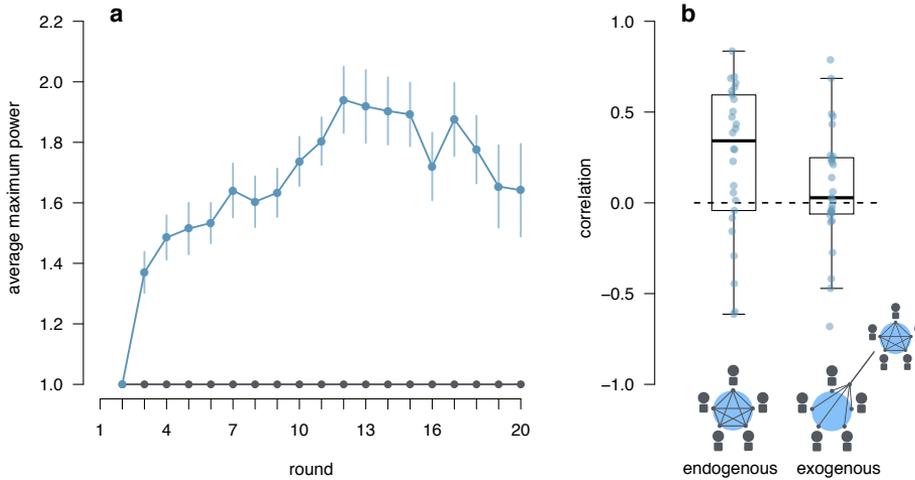


Figure 4.3: Power and cooperation. (a) Change of average power of the most powerful group member over rounds in the endogenous condition (blue). In the exogenous condition, power transfers were identical to the endogenous condition, and thus, the average power of the most powerful group member was the same, too. In the fixed condition, power was fixed to 1 (grey). Error bars show the within-subject standard errors of the mean. (b) Distribution of correlations across rounds between maximum power and cooperation for each group in the endogenous and exogenous condition. Thick horizontal bars represent the medians.

Whitney U-test,  $U = 2294$ ,  $p = 0.02$ , two-sided).

Looking at power transfers from round to round, power transfers were mostly initiated by non-punishers. Group members with a lower than average punishment expenditure were significantly more likely to give up power (mixed effect logistic regression, t-1 punishment difference coefficient = 0.51, 95% CI = [0.17, 0.84]). In line with the first two rounds, the likelihood of receiving power was significantly increased by being a cooperator or spending MUs on punishing free riders in the previous round (mixed effect logistic regression, t-1 cooperator coefficient = 0.52, 95% CI = [0.18, 0.86]; t-1 punishing free rider coefficient = 0.78, 95% CI = [0.39, 1.19]). In turn, gaining power further increased the odds of punishing free riders (mixed effect logistic regression, power coefficient = 1.55, 95% CI = [0.85, 2.23]). Since those willing to engage in costly punishment and

cooperating above group average were more likely to gain power, and, in turn, gaining power further increased the likelihood of spending own MUs on punishment, powerful group members earned less than the group average (correlation of power and earnings, Spearman's rank correlation  $r = -0.24$ ,  $p < .01$ , Figure 4.20). This indicates that the behaviour of powerful group members was not driven by selfish payoff-maximisation.

Group members increased their contributions in response to punishment and power changes. The more MUs someone lost due to receiving punishment in the previous round, the more she increased her contribution to the group project (mixed effect regression, earning reduction coefficient = 0.31, 95% CI = [0.25, 0.37]). Furthermore, the higher the increase in power centralization from the previous round, the more group members increased their contributions compared to the previous round (mixed effect regression, power change coefficient = 4.76, 95% CI = [3.06, 6.48]). Thus, group members reacted not merely to actual punishment, but also to the threat of powerful punishment.

However, not all groups in the endogenous condition were able to solve the social dilemma. Out of 27 groups, cooperation increased steadily over time in 17 (cooperative groups), while it decreased in 10 (non-cooperative groups). This increase or decrease in cooperation is not explained by initial propensities to cooperate: Cooperation in the first round was not significantly different between cooperative and non-cooperative groups (t-test,  $t(24) = 0.8$ ,  $p = 0.43$ , two-sided). To reach a solution to the free rider problem from a state of weak and decentralized punishment, a punisher needs to be sufficiently empowered by her peers. Therefore, some group members have to be willing to give up power and believe that their power will be used in their interest. We measured this willingness by the total amount of power transferred within the group. Further, they have to select the right person to shift power to, somebody who will use this power for the benefit of the group. To evaluate whether this selection was successful, we calculated the share of rounds in which the group member most willing to punish free riders in the past became the most powerful.

Power centralization, the willingness to give up power, as well as selection success were not different in the first third of the experiment across cooperative and non-cooperative groups (Figure 4.4). However, power centralization increased more sharply in cooperative groups and remained stable towards the end of the experiment (Figure 4.4a). This result supports the findings reported above (see Figure 4.2b). In non-cooperative groups, power centralization decreased towards the end of the experiment. This observed difference was not driven by willingness to give up power. The average amount of power transferred was similar in the first two thirds of the experiment (Figure 4.4b). Instead,

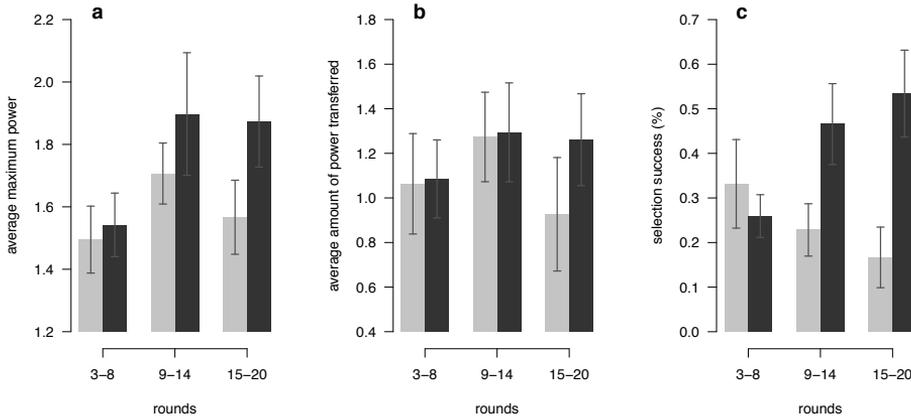


Figure 4.4: Characteristics of cooperative and non-cooperative groups across time intervals. Bars depict groups in which cooperation declined (light grey), or increased (dark grey). (a) Power centralization, measured by the power of the most powerful group member; (b) average amount of power transferred; (c) selection success, measured by the share of rounds in which the most active punisher of non-cooperators of past rounds was the most powerful.

cooperative and non-cooperative groups diverged in their success to centralize power in the hands of a group member who reliably punished free riders over past rounds (Figure 4.4c). Thus, groups that transferred sufficient power to the right group member could maintain cooperation. Figure 4.5 shows the power transfer networks. Although the initial networks had a similar structure, non-cooperative groups diverted more power away from the centre, and also transferred it along circles, leading to less power centralization. On the other hand, cooperative groups directed more and more power to one group member over time.

## 4.4 Discussion

Power hierarchies help foster cooperation and lead to a better social outcome in environments where decentralized peer punishment is unable to sustain cooperation. The transfer of power solves the social dilemma by enabling group members who do not punish (second-order free riders) to empower cooperators who are willing to sacrifice private resources to bring free riders in line. Free

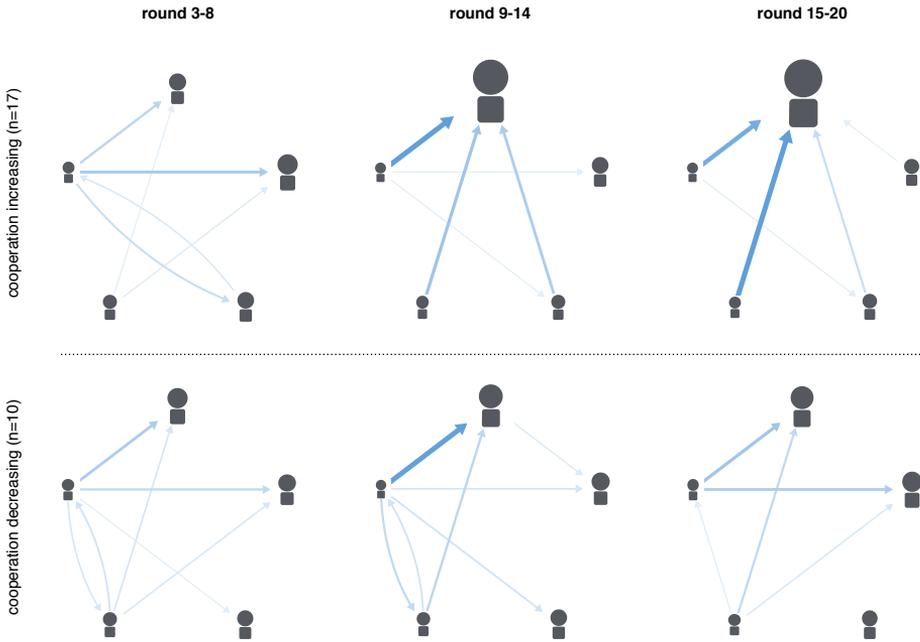


Figure 4.5: Power networks by time interval and cooperation success. Each network shows the average power transfers (blue arrows) of groups in which either cooperation increased (top), or declined (bottom) in a given third of the experiment. The thickness of the line is proportional to the amount transferred. The size of the group members (nodes) is proportional to the amount of accumulated power.

riders anticipate this behaviour and raise their cooperation when they observe that a powerful group member is emerging.

Our work demonstrates the emergence of centralized punishment out of a ‘state of nature’, characterized by weak and decentralized punishment. The resulting power hierarchy overcomes the problems of peer punishment. As pro-social punishers gain power, anti-social punishment becomes more risky. Interestingly, the most powerful group members earned the least, indicating that their behaviour was not driven by financial incentives. They were instead willing to use their power for the sake of the group. Conversely, it is essential that power is concentrated in the right hands. When groups did not have the freedom

to decide whom to direct power to, or failed to select the right group member, cooperation could not be sustained.

Social structures that are characterized by an unequal distribution of power are not only prevalent in human societies but also in other social animals (Ellis, 1995). For example, many nonhuman primates live in complex social groups organized in dominance hierarchies (de Waal, 1987, Cheney and Seyfarth, 1990). The emergence of social structures in which some group members have more power than others to enforce shared goals may thus be a crucial step in the evolution of cooperation. In human societies, institutions such as elected representative bodies, legal courts and law enforcement agencies govern much of social life. These institutions are built upon, and embody the centralization of power. The willingness to give up, transfer and centralize power, demonstrated here, can be seen as an important intermediary step and prerequisite to the constitution of such complex institutions.

## 4.A Appendix – Supplementary materials

### 4.A.1 Game description

We introduce a new game, the ‘power transfer game’. In the power transfer game, a punishment stage and a power transfer stage are added to the standard public goods game. The game-theoretic description given below first presents the standard public goods game, and then adds the punishment and the power transfer mechanisms sequentially. For equilibrium analysis, we focus on subgame perfect and trembling hand perfect equilibria.

#### Public goods game

The public goods game is played by  $n > 2$  players. Each player has an initial endowment of  $E$  monetary units (MUs). Players decide on how much of the endowment to contribute to a public good, keeping the remaining amount. Contributions to the public good are multiplied by  $m$ , and then distributed equally among all players. We assume  $m > 1$  and  $\frac{m}{n} < 1$ ;  $\frac{m}{n}$  is called the ‘marginal per capita return’. Decisions on contribution are made simultaneously. The contribution of player  $i$  is denoted by  $c_i$ , with  $0 \leq c_i \leq E$ . After the contribution decisions are made, each player is informed about the contributions of all players, and the level of the public good.

In the public goods game, player  $i$ ’s payoff is given by:

$$\pi_i = E - c_i + \frac{1}{n}m \sum_{j=1}^n c_j.$$

Selfish payoff-maximisation prescribes a contribution of  $c_i = 0$  for each individual  $i$ , as  $\frac{m}{n} < 1$ , and thus the private return on each invested MU is negative:  $\frac{m}{n} - 1 < 0$ . However, since  $m > 1$ , the social return of contributions is positive:  $m - 1 > 0$ . Therefore, social optimum, defined as the sum of all monetary payoffs, is reached when all the players contribute their full endowment, i.e., at  $c_i = E$ . With full contributions, the earnings of everyone will be  $mE$ . However, at individually rational contributions of zero, each player earns just their endowment  $E$ . Thus, everyone contributing their full endowment constitutes a Pareto-improvement over zero contributions. As selfish motives clash with group interests, the public goods game is an example of a social dilemma.

Because contributing zero is a strictly dominant strategy for every player,  $c_i = 0$  for all players constitutes the only Nash-equilibrium of the game. When the public goods game is played for a finite number of periods, we get through

a backward induction argument that contributions are zero for every period in the only subgame perfect equilibrium of the game.

In our experiment, the game was played with  $n = 5$  players, and the contribution multiplier was set to  $m = 1.5$ . The initial endowment was  $E = 20$ , and only integer contributions were allowed.

### Public goods game with punishment

A punishment option is added to the simple public goods game. Following contributions and after being informed about their peers' contributions, players simultaneously make a punishment decision. Player  $i$  can assign  $0 \leq d_{ij} \leq d_{\max}$  deduction points (DP) to each other player  $j \neq i$ . For each DP that player  $i$  assigns to player  $j$ , player  $i$ 's own payoff is reduced by the cost of punishment,  $pc$ , while the punished player  $j$ 's payoff is reduced by the effectiveness of punishment  $pe$ . After punishment decisions are made, players are informed about who punished whom, and by how much (i.e., all DP assignments  $d_{ij}$ ).

In the public goods game with punishment, player  $i$ 's payoff is given by:

$$\pi_i = E - c_i + \frac{1}{n}m \sum_{j=1}^n c_j - pc \sum_{\substack{j=1 \\ j \neq i}}^n d_{ij} - pe \sum_{\substack{j=1 \\ j \neq i}}^n d_{ji}.$$

In the only subgame perfect equilibrium of the one-shot public goods game with punishment, punishments and contributions are zero. The reasoning is as follows: Once the contribution decisions are made, no selfish payoff-maximiser has any incentive to punish, because doing so would only reduce her payoffs. Therefore, it can be known from the start that  $d_{ij} = 0$ . Therefore, later punishment is not credible, and cannot raise contributions above  $c_i = 0$ . In the finitely repeated public goods game with punishment, the same backwards induction argument can be used, starting with the last period, to show that in a subgame perfect equilibrium, punishment and contribution is zero in all periods.

Almost ubiquitously in the literature, the cost of punishment is set at  $pc = 1$ . In the majority of experiments, the effectiveness of punishment is higher than its cost, with the most commonly used value of  $pe = 3$ . In our experiment, we set the *initial* effectiveness to the same level as its cost at  $pe = pc = 1$  (see the next subsection). We limited the number of deduction points that could be assigned to  $d_{\max} = 10$ . We imposed no lower constraint on the earnings of players. Thus, players could punish and be punished below zero income.

### Power transfer game

We modify the public goods game with punishment by adding an additional decision *before* the contribution decision where players may transfer their power. Power determines the effectiveness of punishment. Players control an equal amount of power  $\widehat{pe}$ , and can decide to transfer it to other players. Thus, player  $i$  decides on how much power  $pt_{ij}$  to transfer to each player  $j$  (allowing for  $j = i$ ), with  $pt_{ij} \geq 0$  and  $\sum_{j=1}^n pt_{ij} = \widehat{pe}$ . Power transfer is free. Moreover, transferring power has no impact on the cost of punishment  $pc$ . The power  $pe_i$  of player  $i$  will be the total power she keeps for herself and receives from others:  $pe_i = \sum_{j=1}^n pt_{ji}$ . Players are informed about every player's power before they decide on their subsequent contributions.

In the power transfer game, player  $i$ 's payoff is given by:

$$\pi_i = E - c_i + \frac{1}{n}m \sum_{j=1}^n c_j - pc \sum_{\substack{j=1 \\ j \neq i}}^n d_{ij} - pe_j \sum_{\substack{j=1 \\ j \neq i}}^n d_{ji}$$

To recapitulate, the items of the sum are, in turn, the initial endowment, the contribution to the public good, the individual return from the public good, the cost of punishment dealt to others, and the punishment received from others.

The cost of punishment  $pc$  is not affected by power transfers. Thus, the same argument showing that both punishments and contributions are zero in the only equilibrium of the public goods game with punishment can be extended to the power transfer game straightforwardly. Therefore,  $c_i = 0$  and  $d_{ij} = 0$  in the subgame perfect equilibria of the finitely repeated power transfer game. However, we get a multiplicity of equilibria, because any level of power transfers is compatible with these choices. We can get a unique equilibrium by focusing instead on trembling hand perfection. In this case, 'slips of hand' – strategies that assign a positive probability to every pure strategy – should be taken into account. Thus, a player needs to consider the possibility that she might get punished by another by mistake. If the player has transferred any power to whomever punishes her, her payoff will be lower than if she had chosen not to transfer power. Because of this potentially harmful effect of power transfers, players should not transfer power to any other player. Therefore, in the only trembling hand perfect equilibrium, we get  $pe_{ij} = 0$  for each player  $i$  and other player  $j \neq i$ .

In our experiment, each player controlled a total power of  $\widehat{pe} = 1$ , so the constraint on power transfers was  $\sum_{j=1}^n pt_{ij} = 1$ .

### 4.A.2 Timeline and parameter setup

We assign each simultaneously made decision, together with the feedback/information provided after the decision to a *stage*. The timeline of the power transfer game is thus the following.

1. Power transfer stage: power transfer decision;  
players are informed about every player's effectiveness.
2. Contribution stage: contribution decision;  
players are informed about every player's contribution.
3. Punishment stage: DP assignment;  
players are informed about all DP assignments.

To summarise, our choice of parameters for the experiment was as follows:

- $n = 5$  players;
- contribution multiplier  $m = 1.5$ ;
- initial endowment  $E = 20$  MUs;
- maximum number of deduction points  $d_{\max} = 10$ ;
- punishment cost  $pc = 1$ ;
- own power  $\hat{p}e = 1$ .

### 4.A.3 Participants

Participants were recruited from the subject pool of the Behavioural and Experimental Economics lab (BEElab) at Maastricht University and were invited via e-mail. Participants were randomly assigned to groups of five. Each experimental session comprised at least 3 and at most 5 groups.

Participants were paid the sum of monetary units (MUs) earned over the 20 rounds (10 MUs = €0.25) plus a show-up fee of €3 and a small sum based on an incentivised social value orientation questionnaire. In three groups the experiment was aborted due to a network malfunction. In these groups data are available for at least 14 rounds and used as such in the analysis.

#### 4.A.4 Experimental setup

Participants were seated in separate cubicles and stayed there for the duration of the whole experiment. Each participant had a notepad and a pen to make notes.

The experiment started with one round of a public goods game (PG). Participants read instructions explaining the rules of the PG on the computer screen. Instructions used neutral labels for describing the social dilemma: Participants were told that they would receive 20 MUs each round and that they have to decide how many MUs to contribute to a ‘project’. The output of the project would then be distributed equally among all group members, irrespective of how much each member contributed. After reading the instructions, participants had to answer a set of comprehension questions about the rules of the PG.

##### **Round 1: contribution**

At the beginning of the first round each participant was asked to indicate how much of the 20 MUs to contribute to the project. When entering a number, the participant saw a graphical representation of the share of MUs she would contribute. After each participant in the group made a contribution decision, they saw how much each group member contributed. Group members were associated with a unique symbol by which they could be identified throughout the whole experiment. Participants saw a summary of the earnings of each group member and the outcome of the group project for this round. This summary was provided at the end of every round.

##### **Round 2: punishment**

In the second round, punishment was introduced. Participants received instructions on the computer screen and answered a set of comprehension questions about the punishment rules. Instead of ‘punishment’ we used the neutral label ‘deduction’ and ‘deduction points (DP)’ in the instructions and throughout the experiment. Upon answering all questions correctly, participants entered the contribution stage of round 2. After getting informed about the contribution of each group member, participants simultaneously assigned between 0 and 10 deduction points to each other group member. When entering a number, participants saw the DP costs as well as the effect the punishment would have on the punished.

The deduction stage outcome summary showed which participants were punished and by whom using a graphical matrix representation. By going through

the matrix by columns, participants could see how much a group member contributed in the contribution stage, how many DPs in total this group member spent on punishing others, and how many DPs were assigned to her by the others as well as the effect it had on her earnings. By going through the matrix by rows, participants could see how much the corresponding group member punished other group members. On this screen, participants also had the possibility to look at behaviour of previous rounds. Thus, it was possible to review past contribution and punishment decisions of each group member. The round ended with the outcome screen.

In the second round, for each assigned DP, the punisher had to pay 1 MU and the punished would lose 1 MU. With the power transfer mechanism introduced in the third round and explained below, this effectiveness-to-cost ratio of punishment could change from round to round.

### **Round 3: power transfer**

Before entering the next round, participants received the third and final set of instructions. In the fixed condition, participants were told that the contribution stage and the punishment stage would be repeated for another 18 rounds. In the endogenous condition, the power transfer stage was explained to the participants. Neutral labels for power was used. Instead of ‘power’, we used ‘deduction effectiveness’ and the stage was called ‘shifting stage’. After answering a set of comprehension questions about the power transfer mechanism, participants started the third round with the power transfer stage. In this stage, participants had the possibility to transfer power to other group members. Each participant had a power of 1 and could transfer power in units of 0.1 to other group members. Power could also be distributed among multiple group members (e.g., it was possible to transfer 0.5 to one participant, 0.2 to another participant and keep 0.3 to oneself).

After all participants made their power transfer decisions, they were shown the power each group member had for this round, based on the five transfer decisions. Note that participants could not see the individual transfer decisions of the other group member but only the outcome of these decisions, i.e., the total power of each group member for this round.

In the exogenous condition, instead of transferring power by themselves, power changed exogenously based on the power transfer decisions that participants made in one of the endogenous groups. In the instructions set, participants were told that “deduction effectiveness of you and the other group members can change” from round 3 to 20. Therefore, they could not transfer power voluntar-

ily, but only saw the changes in power at the beginning of each round, starting from round three.

Changes in power modified the effectiveness of punishment. In the fixed condition, power was fixed to 1 and transferring power was not possible, just like in round two for all conditions. In the other two conditions power could change as explained above.

### **Rounds 4–20**

In the exogenous and endogenous condition the power transfer, the contribution and the punishment stage were repeated for the consecutive 17 rounds, with the difference that in the exogenous condition, participants saw the change in power from round to round, without being able to influence it themselves.

Each round began with the power transfer stage. The transfer decisions made in the previous round served as the status quo for the current round. When entering round 4, participants would see the power status each group member had in the previous round together with the transfer decisions made by the participant in the previous power transfer stage. Thus, by default, the participant would make the same power allocation as th one she chose in the previous round. However, the participant could also decide to reverse or change the previous decision by changing the numbers below each bar accordingly (with the only constraint that the total amount of power transferred could never exceed 1).

By introducing punishment and power transfer stage round by round to the participants, we were able to measure baseline contribution and punishment rates across conditions. There should be no significant differences in average contributions in the first and average contributions and punishment in the second round between conditions, since the actual experimental manipulation started in the third round.

### **Computer interface**

Figures 4.6 to 4.16 show what a group member would see on the computer screen in the different stages from round 3 to the beginning of round 4 in the endogenous treatment in a hypothetical round. What is shown is meant as an example to explain the computer interface and does not represent real data.

At the beginning of round 3, our group member sees the first power transfer stage screen (Figure 4.6). In this example, our group member decides to transfer

---

0.5 of her power to group member 3 and 0.2 of her power to group member 4 (Figure 4.7).

After all group members made their power transfer decision, the power transfer outcome screen is shown. Due to power transfers of our and the other group members, group member 3 now holds the most power, while our group member is the least powerful in the group (Figure 4.8).

The next screens show the contribution stage (Figure 4.9). Our group member decides to contribute 15 MUs to the group project (Figure 4.10). After all group members made their contribution decision, the contribution outcome screen is shown (Figure 4.11).

After the contribution stage, the group enters the deduction stage (punishment). Figure 4.12 shows the input screen of the deduction stage. In the first row, the contributions of each group member in the previous contribution stage is shown. By entering numbers between 1 and 10 below each column, our group member is able to assign deduction points. This is shown in Figure 4.13. Our group member decides to assign 2 DPs to group member 2 and 5 DPs to group member 5.

After all group members made their punishment decisions, the deduction outcome screen is shown. In this example, both our group member and group member 3 decides to punish group member 5 by 5 DPs for free riding in this round's contribution stage (Figure 4.14). Since our group member only has a power of 0.3, her punishment is rather ineffective. The 5 DPs costs her 5 MUs but will only reduce the earnings of group member 5 by 1.5 MUs (0.3 effectiveness times 5 DPs). In contrast, group member 3 received power from other group members and has a total power of 2. The 5 DPs assigned to group member 5 also reduces her earnings by 5 (costs of punishment), but will reduce the earnings of group member 5 more substantially. Group member 5 would lose 10 MUs due to this punishment (2.0 effectiveness times 5 DPs). The same logic applies to the punishment of group member 2 by our group member and group member 3 (Figure 4.14).

On the next screen the outcome of this round is shown (Figure 4.15). Group members see the earnings of each group member as well as the outcome of the group project.

By pressing the button at the bottom of the screen, our group member can enter the next round, starting with the power transfer stage. Previously, she decided to transfer some of her power to group members 3 and 4. This decision serves as the default option in this round's power transfer stage (Figure 4.16). By pressing 'accept & proceed', she would again transfer 0.5 and 0.2 of her power to these group members, respectively.

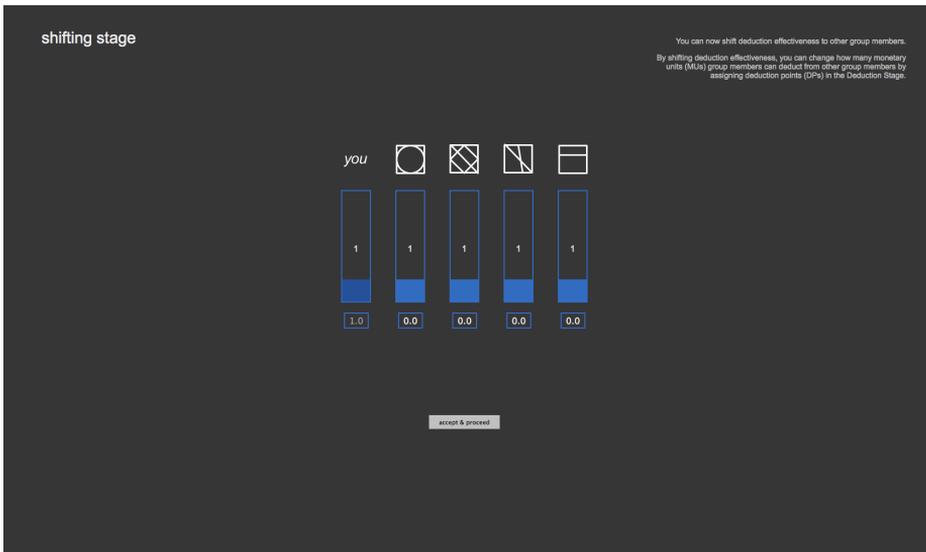


Figure 4.6: Power transfer stage before input, round 3. Since this is the first round with power transfers, the status quo option, visualised here, is not to transfer any power to others. Each group member is associated with a unique symbol throughout the whole experiment. Note that this screen did not appear in the fixed and exogenous condition.

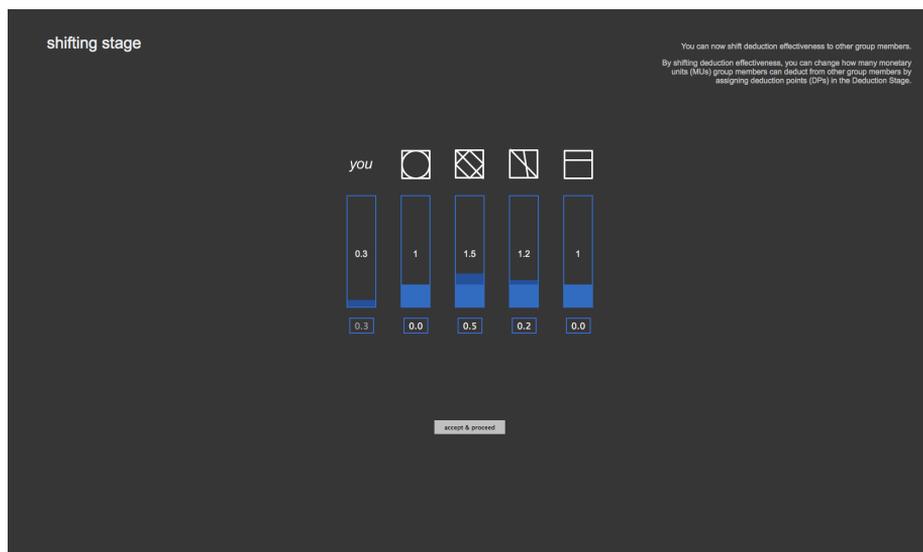


Figure 4.7: Power transfer stage screen, round 3. Input to each field is followed by a graphical representation of how the power status of the respective group member will change (if nobody else transfers any power). Own power is shown in dark blue. A button labelled 'accept & proceed' allows the participant to finalize her decision. Note that this screen did not appear in the fixed and exogenous condition.

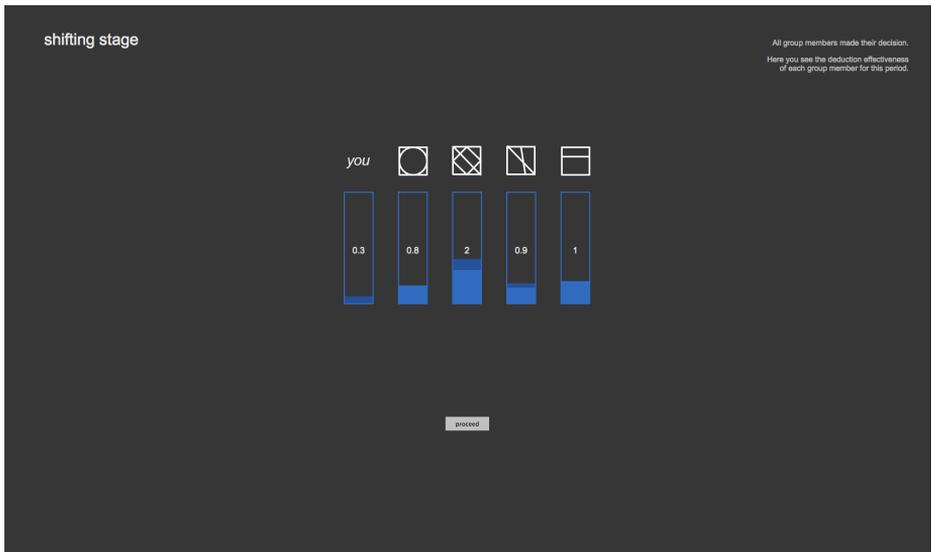


Figure 4.8: Power transfer outcome screen, round 3. After every group member made her power transfer decision, the power transfer outcome screen is shown. Bars represent the power of each group member for this round. Power changes due to transferring own power are shown in dark blue. Note that this screen did not appear in the fixed condition.

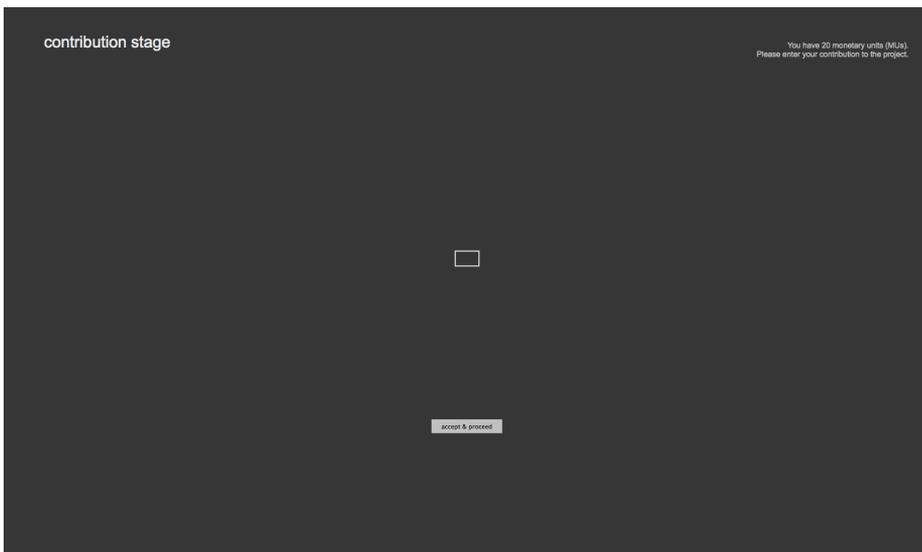


Figure 4.9: Contribution stage before input, round 3. Participants decide simultaneously how much of their endowment of 20 monetary units (MUs) to contribute to the group project.

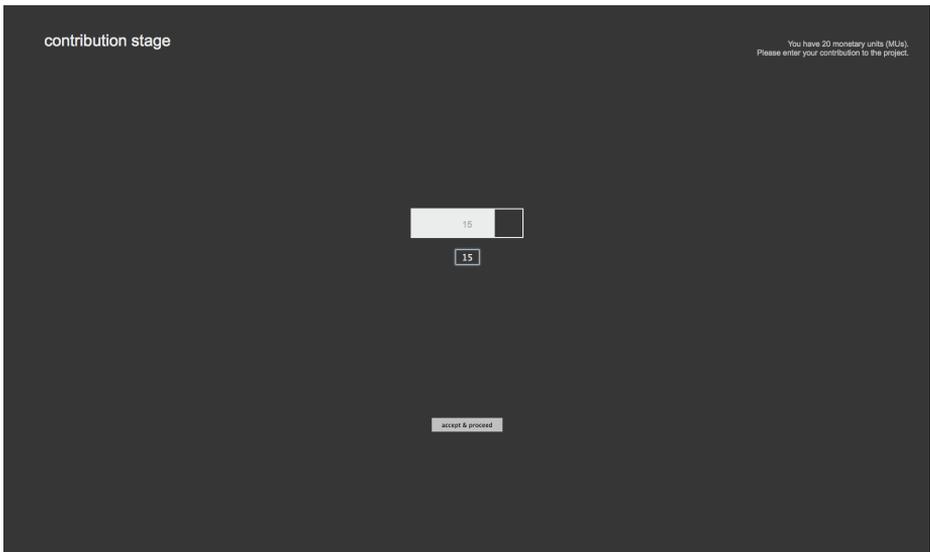


Figure 4.10: Contribution stage input screen, round 3. After inputting a number, the participant sees a graphical representation of the fraction of her endowment she would contribute. A button labelled “accept & proceed” allows the participant to finalize her decision.

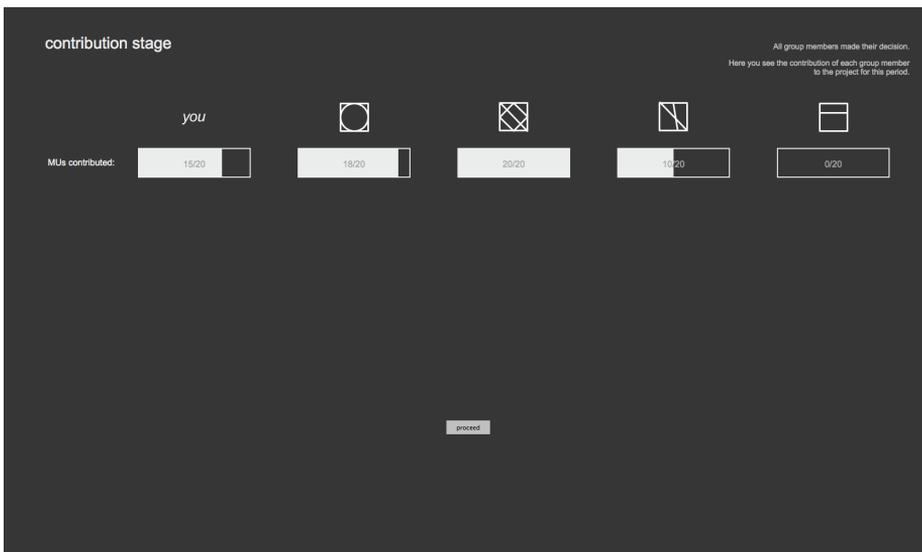


Figure 4.11: Contribution stage outcome screen, round 3. After all participants make their contribution decision, the outcome screen shows how much each group member contributed.

deduction stage

You can now assign deduction points (DPs).  
You can maximally assign 10 DPs per group member.

	you				
MUs contributed:	15 / 20	18 / 20	20 / 20	10 / 20	0 / 20
DP costs:	?	?	?	?	?
you (0.3)		?	?	?	?
 (0.8)	?		?	?	?
 (2.0)	?	?		?	?
 (0.9)	?	?	?		?
 (1.0)	?	?	?	?	

accept & proceed

Figure 4.12: Punishment stage screen before input, round 3. Each participant sees a  $5 \times 5$  punishment matrix (yellow and grey rectangles) and assigns between 0 and 10 deduction points (DPs) to each other group member. Information about the power of each group member is available (blue number next to player symbols on the left). In the first row, the contributions of each group member in the contribution stage of this round are shown.

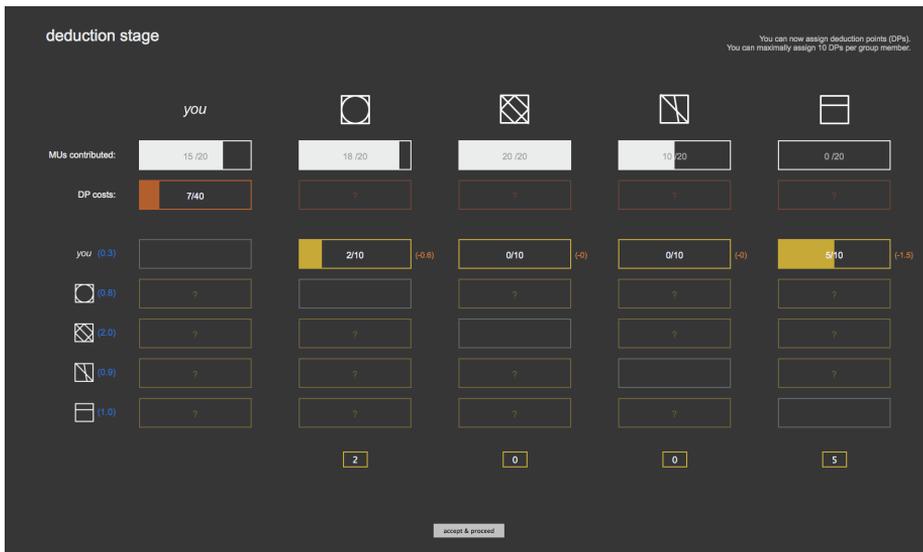


Figure 4.13: Punishment stage screen after input, round 3. The participant sees the total cost of her DP assignments (orange bar on the top left). The current DP assignment, as well as the effect on the punished group member are represented in the  $5 \times 5$  matrix (yellow bars for DP assignments, orange numbers in parentheses next to them representing the effect on the punished).

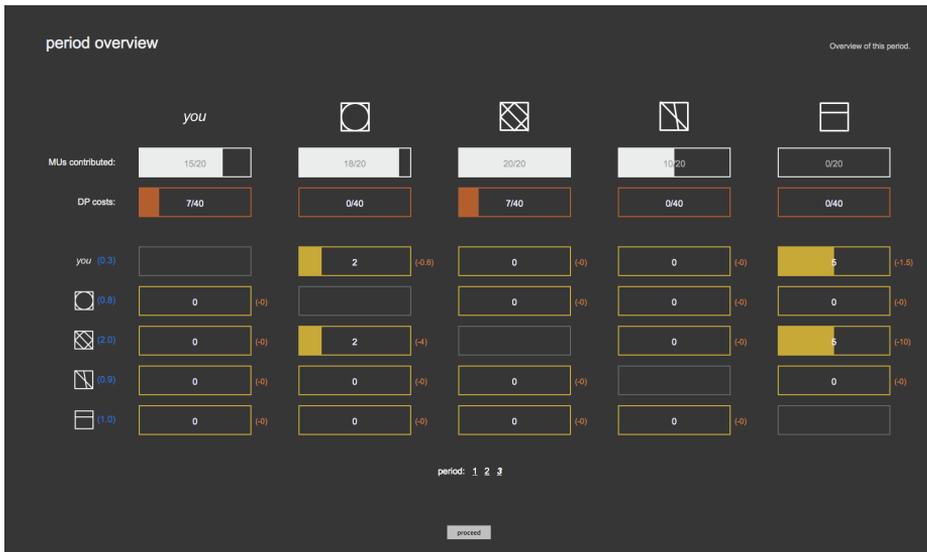


Figure 4.14: Punishment stage outcome screen. Participants see who was punished by whom. Each column indicates by how much a group member was punished, while the rows indicate by whom these deduction points were assigned by.

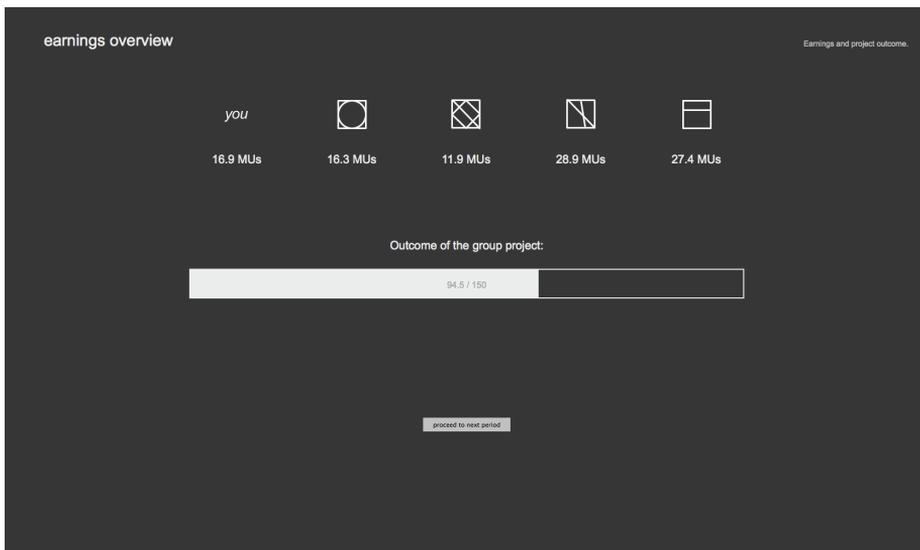


Figure 4.15: Earnings screen, round 3. This screen is shown at the end of each round summarising earnings and the outcome of the group project (sum of contributions  $\times 1.5$ ). The sum of MUs kept, MUs received from the group project, losses due to assigning DPs, and losses due to receiving punishment led to the payoffs seen on the screen.

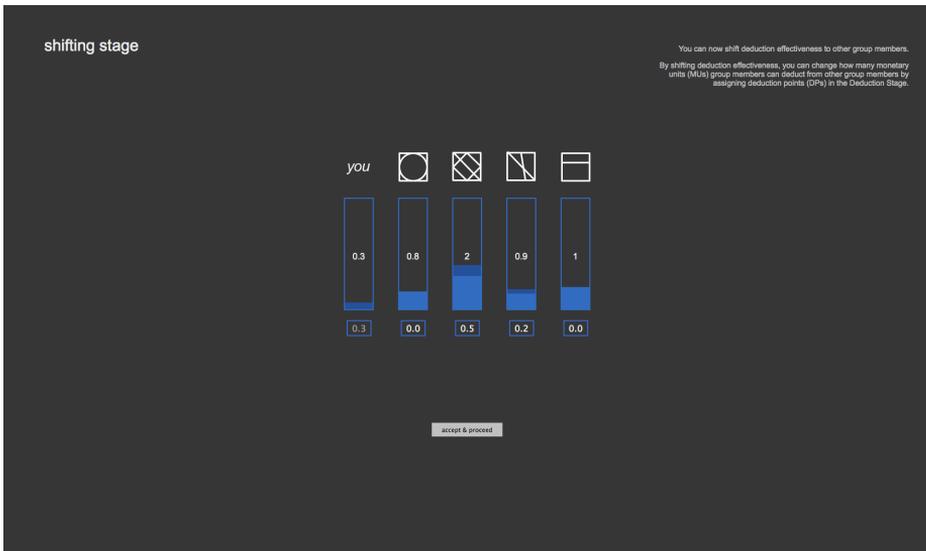


Figure 4.16: Power transfer stage screen before input, round 4. Each bar represents the power of each group member in the previous round. The power transfers made in the previous round serve as the status quo allocation for this round. Dark blue bar segments indicate the power allocated by the active group member. Thus by pressing “accept & proceed”, the participant will again allocate 0.5 of her power to group member 3 and 0.2 of her power to group member 4.

## 4.B Appendix – Supplementary analysis

### 4.B.1 Measurements

For the statistical analysis, we defined proxies and compound measures based on the behavioural data. Each of these measures is defined below.

#### Punishment behaviour

For the receiving power and use of power models (see below), we were interested in how power was used with regards to punishment. We wanted to differentiate the punishment dealt to free riders from that dealt to cooperators (antisocial punishment). Moreover, we opted for a measure that assigned higher value for punishing non-cooperators who deviated more from average contributions. Punishment behaviour was defined as:

$$\text{punishment behaviour}_{it} = \sum_{\substack{j=1 \\ i \neq j}}^5 \left( \frac{\bar{c}_t - c_{jt}}{\sigma_{ct}} \times d_{ijt} \right)$$

where  $c_{jt}$  = contribution of player  $j$  in round  $t$ ,  
 $\bar{c}_t$  and  $\sigma_{ct}$  = mean and standard deviation of contributions  
 in round  $t$ ,  
 $d_{ijt}$  = deduction points assigned by  $i$  to  $j$   
 in round  $t$ , with  $2 \leq t \leq 20$ .

(4.1)

Thus, for each group member in each round, assigned deduction points were weighted by the standardised cooperation of the punished group member and summed up. Negative numbers therefore indicated that punishment was predominantly used to punish cooperators (antisocial punishment), while positive numbers indicated that punishment was predominantly used to punish free riders (e.g., a punishment behaviour value of 1 can be interpreted as using one deduction point to punish a group member with a contribution that was 1 standard deviation below group average in this round's contribution stage). Figure 4.17 shows the percentage of punishment behaviour types across treatments. Antisocial punishment (free riders punishing cooperators) was observed less frequently than no punishment and the punishment of free riders.

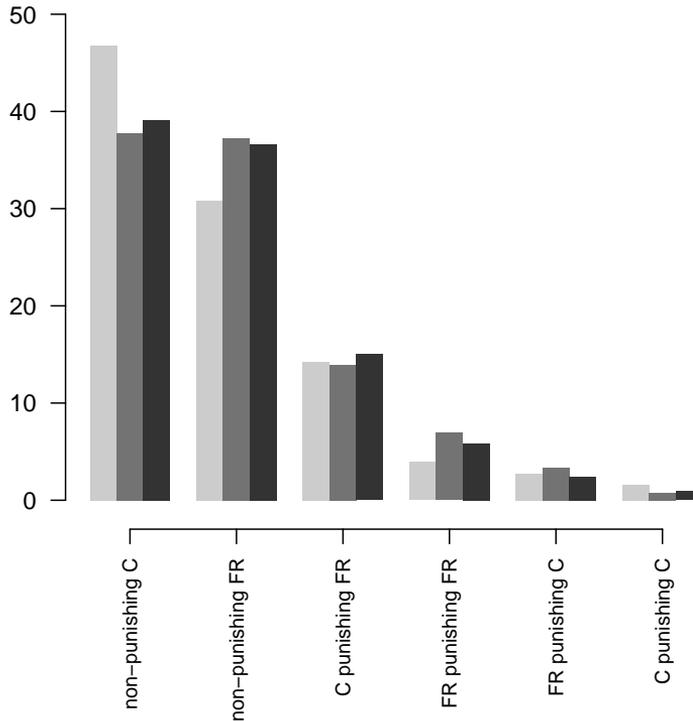


Figure 4.17: Cooperation and punishment decisions. Percentage of punishment decisions observed across treatments (light grey = endogenous condition, grey = exogenous condition, dark grey = fixed condition). C = cooperator, FR = free rider.

### Power centralization

We summarised power centralization through a single indicator: the power of the most powerful group member. In the fixed condition every group member had a fixed power of 1. For the endogenous and exogenous conditions, power centralization was defined by:

$$\text{power centralization}_t = \max_{i \in \{1, \dots, 5\}} pe_{it}$$

$$\text{where } pe_{it} = \sum_{j=1}^5 pt_{jit}, \text{ power (i.e., punishment effectiveness) of } i$$

$$\text{in round } t, \text{ with } 3 \leq t \leq 20.$$
(4.2)

A minimal power centralization of 1 was achieved by, for instance, no power being transferred in a given round. A full power centralization value of 5 would be achieved, if all group members transferred their power to a single individual.

### Willingness to transfer power

We defined our measure for the willingness to transfer power as the sum of all power transferred in a certain round.

$$\text{willingness to transfer}_t = \sum_{i=1}^5 \sum_{\substack{j=1 \\ j \neq i}}^5 pt_{jit}$$
(4.3)

$$\text{where } pt_{jit} = \text{power transferred from } j \text{ to } i$$

$$\text{in round } t, \text{ with } 3 \leq t \leq 20.$$

In rounds when group members did not transfer any power to each other, our willingness to transfer measure took a value of 0. A willingness to transfer of 5 would be achieved, if everyone transferred all their power.

### Selection success

We introduced a measure to gauge a group's ability to select the most active prosocial punishers as the recipient of power transfers. First, building on Definition 4.1, we defined the aggregate (past) punishment behaviour as:

$$APB_{it} \text{ (aggregate punishment behaviour}_{it}) = \sum_{j=2}^{t-1} \left( \text{punishment behaviour}_{ij} \right)$$

in round  $t$ , with  $3 \leq t \leq 20$ .

(4.4)

This variable summarized information about the past behaviour of group members, as potential leaders of the in-group power hierarchy. Within a group, larger values for a certain member  $i$  indicated both the amount of resource  $i$  sacrificed for punishment, as well as her propensity to pick less cooperative group members, as a target for her punishment. If the past is a reliable predictor of an individual's future behaviour, those who showed a willingness to punish non-cooperators should receive the most power. Thus, for groups in the endogenous condition, we defined our selection success in the following manner:

$$\text{selection success}_t = \begin{cases} \frac{1}{|\arg \max_{i=1, \dots, 5} pe_{it}|} & \text{if } APB_{it} \in \arg \max_{i=1, \dots, 5} pe_{it}, \\ 0 & \text{otherwise,} \end{cases} \quad (4.5)$$

in round  $t$ , with  $3 \leq t \leq 20$ .

Specifically, if in a certain round individual  $i$  having the highest aggregate punishment behaviour was strictly more powerful than all other group members, then for that round the selection success of the group would be 1. If, however,  $i$  was less powerful than some other group member, we would assign it a value of 0. In case of a tie, the value would be 1 divided by the number of group members that are equally (most) powerful. In such a case, the group has partially solved the selection problem, and we would assign  $\frac{1}{2}$ ,  $\frac{1}{3}$ , etc. to the selection success variable.

#### 4.B.2 Statistical models

Because of the hierarchical structure of the data (participants clustered in groups and repeated measures over rounds), we fitted Bayesian mixed effects models to the data using R and JAGS.

Non-informative Gaussian priors ( $m=0$ ,  $sd=100$ ) were used for each predictor and non-informative uniform priors (range from 0 to 100) for the error terms.

In every model, random intercepts and slopes were allowed to covary. Therefore, the variance-covariance matrix was estimated alongside the fixed and random coefficients. For the correlations between random effects, non-informative uniform priors (range from  $-1$  to  $1$ ) were used. We used three parallel chains. For every estimated coefficient, the potential scale reduction factor (Gelman and Rubin Diagnostic) was below  $1.05$ , indicating good mixing of the three chains and thus high convergence. Regression tables reported below show estimated coefficients together with the 95% confidence interval (CI, also called highest density interval in the Bayesian framework). Note that, since non-informative priors were used, a 95% CI that only contains negative or positive values can be interpreted as significant at a  $p = .05$  two-sided threshold from the frequentist perspective. Fitting the models using restricted maximum likelihood (REML) as implemented in the `lme4` package in R revealed similar estimates and the same statistical inferences. However, models on the individual subject level failed to converge and also the censoring in the data could not be accounted for in these models.

### Group level analysis

The aim of the group level analysis was to compare cooperation (i.e., contribution to the group project), punishment and earnings across the three different conditions, as well as to analyse the increase of maximum power over rounds in the endogenous condition (and thus also the exogenous condition). For this, we aggregated the data by group members, such that for each group we had one data point for each round (e.g., average contribution).

**Contribution.** The fixed part of the contribution model contained two dummy variables coding the three experimental conditions (with the fixed condition as baseline), a continuous round predictor and the round  $\times$  condition interactions. The random part contained a random intercept as well as a random slope for the round predictor for each group. Thus, for each group a separate baseline cooperation rate in round 1, and a separate slope of cooperation over rounds was estimated (see Equations 4.6). Since average group contribution could not exceed 20 and fall below 0, the data were treated as left and right censored.

$$\begin{aligned}
y_i &\sim N(\mu_y, \sigma_y^2), \text{ for } i = 1, \dots, n \\
\mu_y &= \alpha_{1j} + \beta_{1j}\text{round} + \alpha_2 + \beta_2\text{round} \\
&\quad + \beta_3\text{exogenous} + \beta_4\text{endogenous} \\
&\quad + \beta_5\text{exogenous} \times \text{round} + \beta_6\text{endogenous} \times \text{round} \\
\begin{pmatrix} \alpha_{1j} \\ \beta_{1j} \end{pmatrix} &\sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{\alpha_1}^2 & \rho\sigma_{\alpha_1}^2\sigma_{\beta_1}^2 \\ \rho\sigma_{\alpha_1}^2\sigma_{\beta_1}^2 & \sigma_{\beta_1}^2 \end{pmatrix} \right), \text{ for } j = 1, \dots, J,
\end{aligned} \tag{4.6}$$

where  $n$  = number of observations,  
 $J$  = number of groups.

Table 4.1 shows the estimated coefficients together with the 95% CI. First round cooperation was predicted to be 12.6 for the fixed condition. Groups in the exogenous and endogenous conditions did not significantly differ from this initial level of cooperation ( $\beta_3$  and  $\beta_4$ ). In the fixed condition, there was a significant drop in cooperation over rounds ( $\beta_2$ ). The drop in cooperation was not significantly different in the exogenous condition ( $\beta_5$ ). In contrast, in the endogenous condition, cooperation over time was significantly higher ( $\beta_6$ ).

Examining the posterior distributions of the exogenous  $\times$  round and endogenous  $\times$  round parameter revealed an estimated difference of 0.24 with a 95% CI ranging from 0.08 to 0.39. Thus, also compared to the exogenous condition, cooperation over time was significantly higher in the endogenous condition.

**Punishment** The punishment model followed the same structure as the contribution model (see Equations 4.6), except for average deduction points spent as the dependent variable. Since average group punishment could not fall below 0, the data were treated as left censored.

Table 4.2 shows the estimated coefficients together with the 95% CI. Punishment expenses in the second round (first round with punishment) did not significantly differ across conditions ( $\alpha_2$ ,  $\beta_3$  and  $\beta_4$ ). In the fixed condition, the use of punishment dropped over rounds ( $\beta_2$ ). Groups in the exogenous condition did not deviate significantly from this trend ( $\beta_5$ ). In the endogenous condition, the drop in punishment expenses over rounds was significantly higher compared to the fixed condition ( $\beta_6$ ).

---

		estimate	95% CI
$\alpha_2$	intercept (fixed condition, round 1)	12.60	[10.19, 14.95]
$\beta_2$	round	-0.28	[-0.51, -0.05]
$\beta_3$	exogenous condition dummy	-1.84	[-4.86, 1.01]
$\beta_4$	endogenous condition dummy	-1.47	[-4.56, 1.46]
$\beta_5$	exogenous condition $\times$ round	0.22	[-0.06, 0.51]
$\beta_6$	endogenous condition $\times$ round	0.46	[0.16, 0.74]
$\sigma_{\alpha_1}^2$	error-term random intercepts	0.47	[0.38, 0.56]
$\sigma_{\beta_1}^2$	error-term round slopes	2.41	[2.31, 2.51]
$\sigma_y^2$	error-term y	4.55	[3.76, 5.42]
$\rho$	correlation between random effects	-0.15	[-0.40, 0.09]

---

Table 4.1: Contribution regression model.

Dependent variable: Contribution group average clustered by group.

---

		estimate	95% CI
$\alpha_2$	intercept (fixed condition, round 2)	1.32	[0.70, 1.94]
$\beta_2$	round	-0.08	[-0.13, -0.03]
$\beta_3$	exogenous condition dummy	-0.19	[-0.97, 0.58]
$\beta_4$	endogenous condition dummy	0.10	[-0.66, 0.89]
$\beta_5$	exogenous condition $\times$ round	-0.01	[-0.08, 0.05]
$\beta_6$	endogenous condition $\times$ round	-0.07	[-0.13, -0.002]
$\sigma_{\alpha_1}^2$	error-term random intercepts	0.08	[0.05, 0.10]
$\sigma_{\beta_1}^2$	error-term round slopes	1.47	[1.39, 1.55]
$\sigma_y^2$	error-term y	1.04	[0.78, 1.32]
$\rho$	correlation between random effects	-0.60	[-0.82, -0.35]

---

Table 4.2: Punishment regression model.

Dependent variable: Average punishment expense clustered by group.

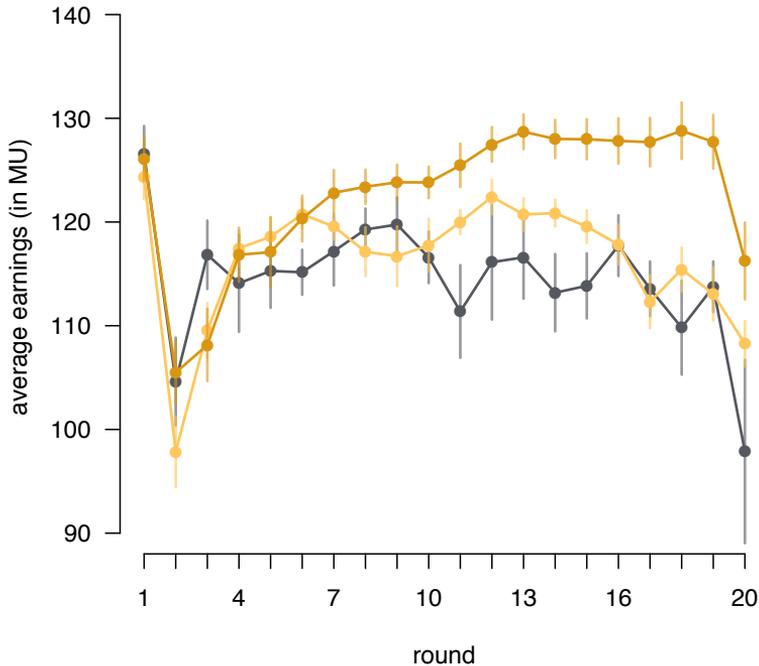


Figure 4.18: Average earnings over rounds. Average earnings in the fixed condition (grey), exogenous condition (light yellow) and endogenous condition (dark yellow). Error bars show the within-subject standard errors of the mean.

**Group earnings** Figure 4.18 shows the average earnings over rounds by condition. There was a substantial drop in earnings in all three conditions after the introduction punishment to the experiment in round 2, as it is often observed in PG experiments with punishment.

The group earnings regression model followed the same structure as the contribution model (see Equations 4.6), except for aggregated earnings (in MUs) as the dependent variable. Since earnings for one round could not exceed 150 MUs ( $20 \text{ MUs} \times 5 \text{ participants} \times 1.5$ ), the data were treated as right censored.

Table 4.3 shows the estimated coefficients together with the 95% CI. Earnings in the first round did not significantly differ between fixed and exogenous condition, as well as fixed and endogenous condition ( $\beta_3$  and  $\beta_4$ ). Earnings were significantly higher over rounds in the endogenous condition compared to the

		estimate	95% CI
$\alpha_2$	intercept (fixed condition, round 1)	118.72	[111.12, 127.07]
$\beta_2$	round	-0.36	[-1.08, 0.36]
$\beta_3$	exogenous condition dummy	-3.31	[-13.29, 6.79]
$\beta_4$	endogenous condition dummy	-3.30	[-13.30, 6.64]
$\beta_5$	exogenous condition $\times$ round	0.68	[-0.24, 1.59]
$\beta_6$	endogenous condition $\times$ round	1.47	[0.56, 2.41]
$\sigma_{\alpha_1}^2$	error-term random intercepts	1.36	[1.06, 1.68]
$\sigma_{\beta_1}^2$	error-term round slopes	13.58	[13.01, 14.15]
$\sigma_y^2$	error-term y	14.91	[12.01, 18.04]
$\rho$	correlation between random effects	-0.24	[-0.50, 0.02]

Table 4.3: Earnings regression model.  
Dependent variable: Average earnings clustered by group.

fixed condition ( $\beta_6$ ).

Examining the posterior distributions of the exogenous  $\times$  round and endogenous  $\times$  round parameter revealed an estimated difference of 0.79 with a 95% CI ranging from 0.32 to 1.26. Thus, also compared to the exogenous condition, earnings over time were significantly higher in the endogenous condition.

**Maximum power** The maximum power regression model followed the same random structure as the contribution model (see Equations 4.6), but since power transfers only happened in the endogenous condition and exogenous condition groups merely mimicked these transfers over time, the model used only data of the endogenous condition and therefore contained only one predictor coding the round. The dependent variable was power centralization, i.e., the amount of power of the most powerful group member in a particular round. Since maximum power could not fall below 1, the data were treated as left censored.

Table 4.4 shows the estimated coefficients together with the 95% CI. Already

		estimate	95% CI
$\alpha_2$	intercept (round 3)	1.455	[1.237, 1.672]
$\beta_2$	round	0.024	[0.004, 0.043]
$\sigma_{\alpha_1}^2$	error-term random intercepts	0.041	[0.023, 0.060]
$\sigma_{\beta_1}^2$	error-term round slopes	0.579	[0.536, 0.624]
$\sigma_y^2$	error-term y	0.499	[0.314, 0.699]
$\rho$	correlation between random effects	-0.063	[-0.574, 0.465]

Table 4.4: Power change regression model.  
Dependent variable: maximum power clustered by group.

in the first power transfer stage (round 3), average power centralization was predicted to be 1.5 by the model ( $\alpha_2$ ). Power accumulation increased over rounds. In each round, the maximum power was estimated to increase by 0.02 on average ( $\beta_2$ ).

### Individual level analysis

The aim of the the individual level analysis was to analyse who received power, who was willing to give up power, and how power affected contributions in the endogenous condition on a round-by-round basis. Therefore, the data were not aggregated on a group level and thus was clustered by groups and by individuals (over time). The regression models accounted for that by having two grouping levels (see below).

In the receiving, giving, and use of power models, the distributions of the dependent variables were non-normal and highly restricted. We therefore transformed the dependent variable in these models into a dichotomous variable and fitted logistic regressions.

Also some predictors were transformed to dichotomous ‘type’ variables. This has the downside of losing some statistical power, as well as the ability to make more detailed quantitative statements (e.g., with a one point increase in the independent variable, the probability of the dependent variable being 1 changes by x). On the other hand, coefficients can be interpreted more easily. For example, by converting contributions to a binary variable (equal to or above/below group

average), participants are classified into free riders (those who contributed less than group average) and cooperators (those who contributed at least or above group average).

**Receiving power** The dependent variable of the receiving power model was power received by other group members (0 = no power received, 1 = power received). The fixed part of the model contained the continuous round predictor, a dummy predictor indicating free riding or cooperation (0 = below average contribution, 1 = equal to or above average contribution) in previous round's contribution stage, and a dummy predictor coding the punishment behaviour (see Equation 4.1; 0 antisocial or no punishment, 1 = punishment of free riders) of previous round's punishment stage as predictors.

The random part of the model contained a random intercept and a random slope for the round predictor for each group, as well as a random intercept and a random slope for the round predictor for each participant. Thus, for each group a separate baseline of the likelihood of power transferring in round 3 (first power transfer stage), and a separate slope in how the likelihood of transferring power changed over rounds was estimated. Separately, the model estimated the likelihood of power being transferred in round 3, as well as the change in the likelihood of receiving power over rounds for each participant (see Equations 4.7).

$$Pr(y_i = 1) \sim \text{logit}^{-1}(\mu_y), \text{ for } i = 1, \dots, n$$

$$\begin{aligned} \mu_y &= \alpha_{1j} + \beta_{1j}\text{round} + \alpha_{2k} + \beta_{2k}\text{round} \\ &+ \alpha_3 + \beta_3\text{round} + \beta_4\text{contribution type}_{t-1} \\ &+ \beta_5\text{punishment type}_{t-1} \end{aligned}$$

$$\begin{pmatrix} \alpha_{1j} \\ \beta_{1j} \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{\alpha_1}^2 & \rho_1 \sigma_{\alpha_1}^2 \sigma_{\beta_1}^2 \\ \rho_1 \sigma_{\alpha_1}^2 \sigma_{\beta_1}^2 & \sigma_{\beta_1}^2 \end{pmatrix} \right), \text{ for } j = 1, \dots, J, \quad (4.7)$$

$$\begin{pmatrix} \alpha_{2k} \\ \beta_{2k} \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{\alpha_2}^2 & \rho_2 \sigma_{\alpha_2}^2 \sigma_{\beta_2}^2 \\ \rho_2 \sigma_{\alpha_2}^2 \sigma_{\beta_2}^2 & \sigma_{\beta_2}^2 \end{pmatrix} \right), \text{ for } k = 1, \dots, K,$$

where  $n$  = number of observations,

$J$  = number of groups,

$K$  = number of subjects.

Table 4.5 shows the estimated coefficients together with the 95% CI as well as the odds ratio (exponential of coefficient). According to the model, cooperators had a 68% increase in odds of receiving power ( $\beta_4$ ). The odds of receiving power more than doubled for participants who punished free riders in the previous punishment stage ( $\beta_5$ ).

		estimate	odds ratio	95% CI
$\alpha_3$	intercept (round 3)	0.16		[-1.13 1.53]
$\beta_3$	round	-0.14	0.87	[-0.25 -0.03]
$\beta_4$	contribution type $_{t-1}$	0.52	1.68	[ 0.18 0.86]
$\beta_5$	punishment type $_{t-1}$	0.78	2.18	[ 0.39 1.19]
$\sigma_{\alpha_1}^2$	error-term random intercepts (group level)	3.23		[2.08, 4.53]
$\sigma_{\beta_1}^2$	error-term round slopes (group level)	2.54		[1.87, 3.26]
$\sigma_{\alpha_2}^2$	error-term random intercepts (individual level)	0.23		[0.13, 0.33]
$\sigma_{\beta_2}^2$	error-term round slopes (individual level)	0.27		[0.19, 0.34]
$\sigma_y^2$	error-term y	49.97		[1.39, 96.31]
$\rho_1$	correlation random effects (group level)	-0.55		[-0.89, -0.14]
$\rho_2$	correlation random effects (individual level)	-0.46		[-0.72, -0.18]

Table 4.5: Receiving power regression model.

Dependent variable: power received (0 = no, 1 = yes) clustered by group and participant.

**Giving away power** The regression model for giving away power followed the same random structure described in Equations 4.7. The dependent variable coded the transfer of power to other group member (0 indicated no transfer of power and 1 indicated transfer of power).

In this analysis, we were interested in whether the willingness to spend points on punishment predicted the likelihood of giving away power. Therefore, we used the difference in the amount of points spent on punishment compared to group average as a dummy predictor (0 = punishment expense equal to or above group average, 1 = punishment expense below group average). The fixed part of the model also contained the continuous round predictor and a dummy predictor indicating free riding or cooperation (0 = below average contribution, 1 = equal to or above average contribution) in the previous round's contribution stage.

Table 4.6 shows the estimated coefficients together with the 95% CI as well as the odds ratio. According to the model, punishing below group average increased the odds of transferring power to other group members in the next round by 67% ( $\beta_5$ ).

**Use of power** The regression model for use of power followed the same random structure described in Equations 4.7. The dependent variable coded the punishment behaviour (0 indicated no punishment or antisocial punishment and 1 indicated punishment of free riders).

As predictors we used the continuous round predictor, the group contribution of the present round and a dummy variable indicating whether the participant had power above 1 in the present round (0 = power below or equal to 1, 1 = power more than 1), as well as the interaction of this dummy with the group contribution.

Table 4.7 shows the estimated coefficients together with the 95% CI as well as the odds ratio. According to the model, having a power greater than 1 increased the odds to punish free riders in the consecutive punishment stage nearly fivefold ( $\beta_5$ ). Higher group contributions decreased these odds slightly. For each additional MU invested by the group, the odds for a powerful group member to punish decreased by 1% ( $\beta_6$ ).

**Effect of power and punishment.** The regression model for the effect of power and punishment followed the same random structure described in Equations 4.7. The dependent variable coded the change in cooperation from the previous round. Since the individual change in cooperation could not exceed 20 and fall below -20, the data were treated as left and right censored.

As predictors we used the continuous round predictor, the reduction in earnings due to punishment in the last round and the change in maximum power from last round, as well as the interaction of change in power and a reduction in earnings due to punishment.

---

		estimate	odds ratio	95% CI
$\alpha_3$	intercept (round 3)	-1.42		[-2.52, -0.41]
$\beta_3$	round	-0.04	0.96	[-0.13, 0.04]
$\beta_4$	contribution difference $_{t-1}$	0.03	1.03	[-0.32, 0.38]
$\beta_5$	punishment difference $_{t-1}$	0.51	1.67	[ 0.17, 0.84]
$\sigma_{\alpha_1}^2$	error-term random intercepts (group level)	1.08		[0.04, 2.17]
$\sigma_{\beta_1}^2$	error-term round slopes (group level)	4.55		[3.48, 5.67]
$\sigma_{\alpha_2}^2$	error-term random intercepts (individual level)	0.07		[0.00, 0.16]
$\sigma_{\beta_2}^2$	error-term round slopes (individual level)	0.34		[0.26, 0.42]
$\sigma_y^2$	error-term y	49.96		[0.39, 95.35]
$\rho_1$	correlation random effects (group level)	-0.19		[-1.00, 0.83]
$\rho_2$	correlation random effects (individual level)	-0.64		[-0.83, -0.44]

---

Table 4.6: Giving away power regression model.

Dependent variable: Power giving (0 = no power was transferred, 1 = power was transferred) clustered by group and participant.

---

		estimate	odds ratio	95% CI
$\alpha_3$	intercept (round 3)	-1.03		[-1.78, -0.28]
$\beta_3$	round	-0.12	0.89	[-0.18, -0.06]
$\beta_4$	group contribution	-0.01	0.99	[-0.02, 0.00]
$\beta_5$	power	1.55	4.71	[ 0.85, 2.23]
$\beta_6$	group contribution $\times$ power	-0.01	0.99	[-0.02, -0.001]
$\sigma_{\alpha_1}^2$	error-term random intercepts (group level)	0.62		[0.02, 1.17]
$\sigma_{\beta_1}^2$	error-term round slopes (group level)	1.66		[1.19, 2.12]
$\sigma_{\alpha_2}^2$	error-term random intercepts (individual level)	0.09		[0.03, 0.16]
$\sigma_{\beta_2}^2$	error-term round slopes (individual level)	0.13		[0.08, 0.18]
$\sigma_y^2$	error-term y	49.97		[0.04, 95.00]
$\rho_1$	correlation random effects (group level)	-0.30		[-1.00, 0.68]
$\rho_2$	correlation random effects (individual level)	-0.64		[-0.88, -0.35]

---

Table 4.7: Use of power regression model.

Dependent variable: punishment behaviour (0 = no punishment or antisocial punishment, 1 = punishment of free riders) clustered by group and participant.

Table 4.8 shows the estimated coefficients together with the 95% CI. According to the model, contribution decisions were influenced by actual punishment, as well as changes in power of the most powerful group member (threat of getting punished). For every MU a participant lost due to getting punished in the last round increased her contribution by 0.3 MUs ( $\beta_4$ ). A change in power of 0.1 of the most powerful group member increased contributions by 0.5 MUs ( $\beta_5$ ). Additionally, getting punished followed by an increase in power also increased contributions significantly ( $\beta_6$ ).

		estimate	95% CI
$\alpha_3$	intercept (change to round 3)	-0.12	[-0.49, 0.25]
$\beta_3$	round	-0.03	[-0.08, 0.01]
$\beta_4$	earnings reduction (from punishment in round $t - 1$ )	0.31	[0.25, 0.37]
$\beta_5$	change in maximum power (from round $t - 1$ )	4.76	[3.06, 6.48]
$\beta_6$	earnings reduction $\times$ change in power	0.54	[0.08, 0.99]
$\sigma_{\alpha_1}^2$	error-term random intercepts (group level)	0.46	[0.00, 0.89]
$\sigma_{\beta_1}^2$	error-term round slopes (group level)	0.11	[0.00, 0.28]
$\sigma_{\alpha_2}^2$	error-term random intercepts (individual level)	0.07	[0.01, 0.12]
$\sigma_{\beta_2}^2$	error-term round slopes (individual level)	0.01	[0.00, 0.03]
$\sigma_y^2$	error-term y	3.73	[3.62, 3.84]
$\rho_1$	correlation random effects (group level)	-0.73	[-1.00, 0.23]
$\rho_2$	correlation random effects (individual level)	-0.27	[-1.00, 0.83]

Table 4.8: Effect of power and punishment regression model. Dependent variable: change in cooperation from round $_{t-1}$  clustered by group and participant.

### Individual level analysis for the exogenous condition

To analyse how gaining power and a change in maximum power affected individual decisions in the exogenous condition in which power transfers were not

possible, we fitted the ‘use of power’ and the ‘effect of power and punishment’ models described above also to the data of the exogenous condition groups. The regression models followed the same structure as described above for the endogenous condition.

Table 4.9 shows the estimated coefficients together with the 95% CI as well as the odds ratio for the use of power regression model. Table 4.10 shows the estimated coefficients together with the 95% CI for the effect of power and punishment regression model.

Like in the endogenous condition, having a power greater than 1 increased the odds to punish free riders in the consecutive punishment stage ( $\beta_5$ , Table 4.9). Individual contributions in turn increased after experiencing a reduction in earnings due to punishment, similarly to the endogenous condition ( $\beta_4$ , Table 4.10). However, contrary to the endogenous condition, a change in maximum power from last round (threat of punishment) did not affect contribution decisions significantly ( $\beta_5$ , Table 4.10).

Thus, the difference in cooperation we observe over rounds between the endogenous and exogenous condition can not be attributed to a lower willingness to punish free riders after receiving power or a lower effect of punishment on cooperation. Instead, observing an increase in power centralization already increased cooperation of group members in the endogenous condition, while power centralization did not affect cooperation of group members in the exogenous condition.

### **Other remarks**

**Power and cooperation.** In Figure 3b of the manuscript, we reported the correlation of maximum power and average cooperation across rounds for each group. For the calculation of these correlations we omitted round 20 from each group. This was done because of the sharp drop of cooperation in the last round (known as the endgame effect). As can be seen in Figure 4.19, with increased maximum power, mean contribution increased. Round 20 can be identified as an outlier.

Including round 20 in the reported analysis did not change the statistical inferences reported in the manuscript. Quantitatively, the average correlation dropped from  $r = 0.24$  to  $r = .21$  for the exogenous condition groups.

**Power and earnings.** As reported in the manuscript, receiving power was correlated with lower earnings compared to the other group members. Figure

		estimate	odds ratio	95% CI
$\alpha_3$	intercept (round 3)	-2.57		[-3.33, -1.78]
$\beta_3$	round	-0.01	0.99	[-0.05, 0.03]
$\beta_4$	group contribution	0.00	1.00	[-0.01, 0.02]
$\beta_5$	power	1.69	5.42	[ 1.03, 2.34]
$\beta_6$	group contribution $\times$ power	-0.01	0.99	[-0.02, 0.00]
$\sigma_{\alpha_1}^2$	error-term random intercepts (group level)	0.79		[0.30, 1.30]
$\sigma_{\beta_1}^2$	error-term round slopes (group level)	1.67		[1.26, 2.07]
$\sigma_{\alpha_2}^2$	error-term random intercepts (individual level)	0.06		[0.01, 0.10]
$\sigma_{\beta_2}^2$	error-term round slopes (individual level)	0.09		[0.05, 0.13]
$\sigma_y^2$	error-term y	50.01		[0.18, 95.02]
$\rho_1$	correlation random effects (group level)	-0.18		[-0.84, 0.80]
$\rho_2$	correlation random effects (individual level)	-0.79		[-0.97, -0.59]

Table 4.9: Use of power regression model for the exogenous condition. Dependent variable: punishment behaviour (0 = no punishment or antisocial punishment, 1 = punishment of free riders) clustered by group and participant.

---

		estimate	95% CI
$\alpha_3$	intercept (change to round 3)	-0.25	[-0.73 0.22]
$\beta_3$	round	-0.03	[-0.08 0.02]
$\beta_4$	earnings reduction (from punishment in round $t - 1$ )	0.38	[ 0.30 0.46]
$\beta_5$	change in maximum power (from round $t - 1$ )	-0.13	[-0.54 0.29]
$\beta_6$	earnings reduction $\times$ change in power	0.19	[ 0.08 0.30]
$\sigma_{\alpha_1}^2$	error-term random intercepts (group level)	0.15	[0.00, 0.38]
$\sigma_{\beta_1}^2$	error-term round slopes (group level)	0.23	[0.00, 0.66]
$\sigma_{\alpha_2}^2$	error-term random intercepts (individual level)	0.02	[0.00, 0.04]
$\sigma_{\beta_2}^2$	error-term round slopes (individual level)	0.03	[0.00, 0.08]
$\sigma_y^2$	error-term y	4.56	[4.42, 4.69]
$\rho_1$	correlation random effects (group level)	-0.26	[-1.00, 0.83]
$\rho_2$	correlation random effects (individual level)	-0.25	[-1.00, 0.82]

---

Table 4.10: Effect of power and punishment regression model for the exogenous condition.

Dependent variable: change in cooperation from round  $t - 1$  clustered by group and participant.

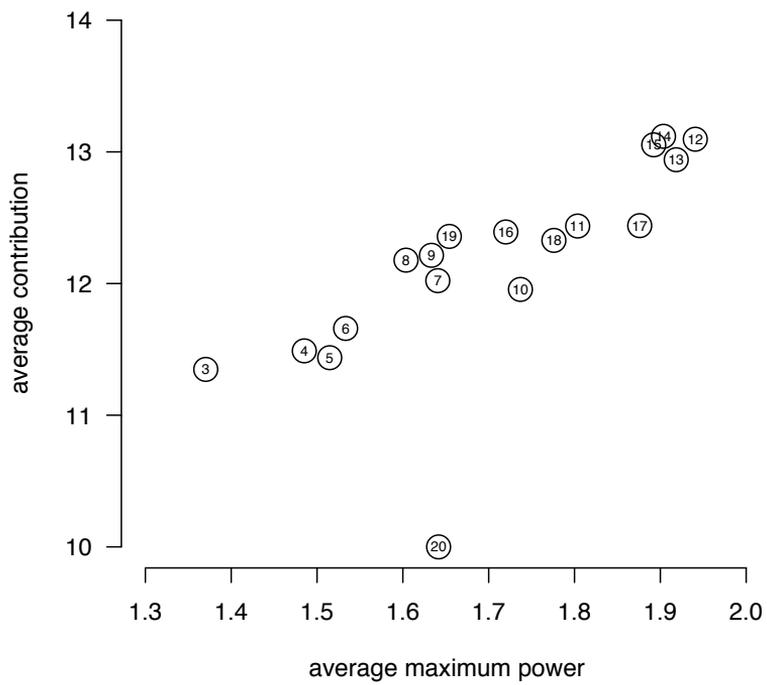


Figure 4.19: Average cooperation and average maximum power in each round of the exogenous condition. The numbers in each circle indicate the round.

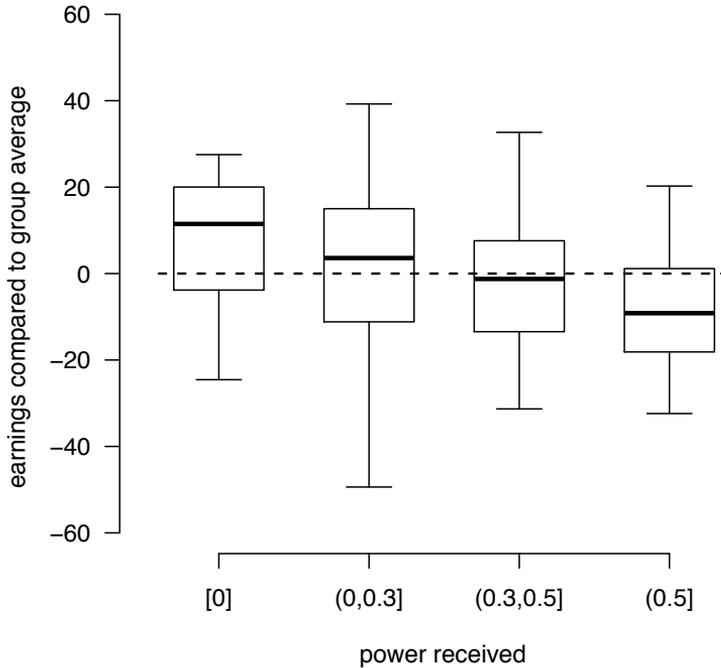


Figure 4.20: Earnings and power. Average earnings of group members compared to the group average for different average amounts of power received over the experiment. Horizontal bars depict the median.

4.20 shows how the earnings of a participant decreased as the power received by her increased.

**First and second round types** When examining the connection between behaviour in the first two rounds and later power status, we omitted an analysis of antisocial punishers, because antisocial punishment was very rare. The distribution of initial punishment behaviour was as follows: 75 participants decided not to punish in the second round, 49 participants punished primarily free riders, and only 11 participants punished antisocially.

## Chapter 5

# Conclusion

We have started with an emphasis on the occasional difficulties in distinguishing between fraudulent and serious science. The previous chapters provided entirely frank investigations of various decision- and game-theoretic problems. While the transitions between chapters might have seemed somewhat rough, at least this thesis can provide an illustration of the state of knowledge in economics today: uncertain, fragmented, but open in many directions. It is almost as if the notion of *tâtonnement* had been transformed into the style of research for the entire field.

The second chapter, perhaps the most technical one, can be read as a detailed examination of the concept of ‘optimality on the limit of finite horizons’. It showed that the meaning of this expression is far from ambiguous. When comparing two strategies on the limit of finite horizons, a distinction should be made between weakly and strongly favoring a strategy over the other. Because the induced ‘favored-on-the-infinite-horizon’ relations are either not transitive, or not complete, ‘optimality’ can also have two possible meanings. Beat-all and not-beaten optimality, combined with the variations of the favored relations yielded a plethora of possible definitions for ‘optimality on the limit of horizons’. However, after the conceptual dust had settled, we were left with a total of five refinements: strict optimality on the infinite horizon, strict and non-strict overtaking optimality, as well as strict and non-strict repeating optimality.

We analyzed and visualized the containment relationships between sets of optimal strategies, and related it to pointwise limits of strategies optimal on finite horizons, studied by Fudenberg and Levine (1983). It has turned out that these limit optimal strategies might fall outside even of our weakest refined set,

that of repeatedly optimal strategies. The next problem was that of existence. Through simple examples, it could be shown that most of our defined sets of optimal strategies can be empty. However, an open question for future research remains: Do repeatedly optimal strategies exist? While we conjectured that it is so, we were unable to find a proof. Section 2.6 presented a ‘valiant effort’ for such a proof through transfinite induction, and have shown why that particular attempt fails.

The overall carriage of the second chapter is both technical and conceptual. While it is impossible to predict when conceptual clarifications will prove to be academically or socially beneficial, nevertheless such exercises are not futile. First, they help us to think more clearly about matters, in particular, when discussing ‘limit optimality’. Moreover, our vocabulary expands, encompassing the previously-introduced Ramsey-Weizsäcker overtaking criterion (von Weizsäcker, 1965). Finally, we have gained a set of examples, some easy and straightforward, some more sophisticated, which can be used for future research.

In the third chapter, we focused on dynamic inconsistency in decision problems. Naiveté and sophistication have been characterized in many ways, albeit usually *en passant*, not as a direct problem for research. Here, we presented a new pair of definitions, based on the distinction between beliefs and intentions. Thus, a naive agent is one that first forms intentions about the future, and then believes them; whereas a sophisticated agent first forms (credible) beliefs about the future, and then makes her intentions match those beliefs. While both procedures yield corresponding intentions and beliefs for the future, the process of determination of the components of agent’s strategies is different for the two types.

The main point of departure for this work was O’Donoghue and Rabin (2001), who were first to explicitly consider the intermediate ground between naiveté and sophistication. However, their model used quasi-hyperbolic discounting because that way the present-biasedness parameter could be handled easily. The framework introduced here is not limited to quasi-hyperbolic discounting, but instead allows for arbitrary utility functions. Our hybrid type also differs from previous approaches in that we encode naiveté or sophistication as part of the state of the decision problem. Thus, we can now allow for situations in which type changes endogenously and stochastically – for example, when one becomes naive by entering an emotionally ‘hot’ state, or through substance abuse.

The Markov decision problems of the Appendix revisited many well-known phenomena from the literature on dynamic inconsistency: procrastination, underinvestment, indulgence, etc. This part also served to exemplify the usage of

the notion of *frame*, novel in the literature, which summarizes all the information about a decision maker present in our model. By comparing the induced utilities of naive and sophisticated frames, it became clear that sophistication is not unequivocally favorable to naiveté. This has led to a further interesting conclusion: A hybrid decision maker might intentionally choose to become naive to achieve a higher expected payoff.

I believe the insights of this chapter are relevant for research on inconsistent decision making, even outside the disciplinary boundary of decision theory. Hybrid decision making could be of interest to psychologists and therapists. For example, patients struggling with addiction could be asked to record their intentions and beliefs about the future in the form of a diary, following our definition of a frame. Our results on the occasional optimality of turning oneself naive can open up interesting questions on self-deception. Is such self-deception conceptually meaningful? If so, can it be achieved in practical terms? And is it ethical to cause oneself to have false beliefs? Finally, our foundational account of naiveté and sophistication opens up the possibility of analyzing dynamically inconsistent players facing each other in a strategic setting.

Chapter 4 shifted ground from theory-driven research on decision making to an experimental study on the emergence and role of power. The ‘power transfer game’ presented therein introduces a novel mechanism in the literature on peer punishment. We allowed groups to not only choose between having a punishment authority or not, but also to determine which group member will become that authority. Moreover, the exact level of the authority’s punishment power was also determined entirely by the participants. The power transfer game thus stands in contrast with research on ‘voting with feet’ models, which only allow the choice between punishment mechanisms already predetermined by the experimenter. In this sense, the power transfer game allows for a truly endogenous emergence of a central punishment authority.

It gives cause for optimism that our subjects managed to overcome the free rider problem starting in an anarchistic state of weak peer punishment. Even the confines of self-organization allowed by our experiment are thus sufficient to enable groups to reach a welfare-enhancing state rather quickly. Nevertheless, such self-organization is not a trivial task, and some of our experimental groups playing the power transfer game failed to sustain cooperation. Naturally, the heterogeneity of motivations, as well as path-dependence play an important part; yet their effects are difficult to measure precisely. On the other hand, this research does identify some crucial factors for reaching a cooperative state: the willingness to give up power; the ability to identify and coordinate on a reliably pro-social punisher; and ultimately, the creation of a centralized punishment

power. Finally, the contrast between the success of the endogenous condition and the failure of the exogenous conditions indicates that the ability to *freely* create a powerful punisher might be crucial. Thus, the perceived legitimacy of the punishment authority could have a strong impact on the efficiency of (peer) punishment to steer free riders towards cooperation.

As it is so often the case, our findings generate more questions than they answer. Some are related to the economics of power and power transfers: Will group members give away power if this choice is costly for them? Can hierarchies emerge when holding power brings extra benefits for the powerful, such as it is often documented in the animal kingdom? Finally, will hierarchies also emerge spontaneously if transferred power cannot be taken back freely, as in many social institutions? Other questions relate to the potentially harmful side of hierarchies. For example, if power transfers are permanent, some group members might choose to act prosocially for a while to gain power, just to abuse their situation later on by free riding. Moreover, if holding power is associated with individual benefits, a competition for power might generate too much punishment, undermining social welfare. Lastly, there are larger differences in peer punishment behavior across cultures. We need to examine the precise cultural milieus that allow for an endogenous emergence of hierarchies empirically, rather than assuming prematurely that our results can serve as a context-independent general picture of social behavior.

Overall, this doctoral thesis presents the most prominent pieces of the research I have conducted with my supervisors and colleagues at Maastricht University since 2010. I do not subscribe to any grand theory of decision making and strategic choice. Indeed, if economists should learn anything from post-modernism, it is that grand narratives are dead (Lyotard, 1984). I believe the concepts included selected for title – beliefs, intentions, and power – should and will play a larger role in economic thinking. However that may be, we as theoreticians should not fall into the trap of self-aggrandizement. We should strive not for an abstract construction of ‘economic science’, but to further the understanding of fellow humans. Only by shedding new light on well-known phenomena, by engaging in games of dialogue, by questioning what we think we already know can we avoid being deceived by tricksters, and also save our minds from a digital conquest.

# Bibliography

- Ainslie, G. W. (1974). Impulse control in pigeons. *Journal of the Experimental Analysis of Behavior*, 21(3), 485–489.
- Ainslie, G. W. (1975). Specious reward: A behavioral theory of impulsiveness and impulse control. *Psychological Bulletin*, 82(4), 463.
- Ainslie, G. W., and Monterosso, J. (2003). Hyperbolic discounting as a factor in addiction: A critical analysis. *Choice, Behavioral Economics and Addiction*, 35–61.
- Akerlof, G. A. (1991). Procrastination and obedience. *The American Economic Review*, 81(2), 1–19.
- Alexander, R. D. (1987). *The biology of moral systems*. Transaction Publishers, New Jersey.
- Ali, S. N. (2011). Learning self-control. *The Quarterly Journal of Economics* 126(2), 857–893.
- Angeletos, G. M., Laibson, D. I., Repetto, A., Tobacman, J., and Weinberg, S. (2001). The hyperbolic consumption model: Calibration, simulation, and empirical evaluation. *The Journal of Economic Perspectives*, 15(3), 47–68.
- Appel, K., and Haken, W. (1980). Every planar map is four colorable. *Bulletin of the American Mathematical Society*, 82(5), 711–712.
- Ariely, D., and Wertenbroch, K. (2002). Procrastination, deadlines, and performance: Self-control by precommitment. *Psychological Science*, 13(3), 219–224.

- Aronowitz, S. (1997). Alan Sokal's Transgression. *Dissent, Winter 1997*, 107–110.
- Asheim, G.B. (2007). *Procrastination, partial naivete, and behavioral welfare analysis*. No. 2007/02. Memorandum, Department of Economics, University of Oslo.
- Bach, C.W., and Heilmann, C. (2011). Agent connectedness and backward induction. *International Game Theory Review*, 13(02), 195–208.
- Baez, J. (2002). Physics bitten by reverse Alan Sokal hoax? Sci.physics Usenet Newsgroup. Retrieved from <https://groups.google.com/forum/?hl=en!msg/sci.physics/KOBNOKq9FvI/ZjV8jeT9-AkJ>.
- Benabou, R., and Pycia, M. (2002). Dynamic inconsistency and self-control: A planner-doer interpretation. *Economics Letters*, 77(3), 419–424.
- Blackorby, C., Nissen, D., Primont, D., and Russell, R. R. (1973). Consistent intertemporal decision making. *The Review of Economic Studies*, 40(2), 239–248.
- Bogdanov, G., and Bogdanov, I. (2001). Topological field theory of the initial singularity of spacetime. *Classical and Quantum Gravity*, 18(21), 4341–4372.
- Bohannon, J. (2013). Who's afraid of peer review? *Science*, 342(6154), 60–65.
- Boyd, R., Gintis, H., and Bowles, S. (2010). Coordinated punishment of defectors sustains cooperation and can proliferate when rare. *Science*, 328(5978), 617–620.
- Boyd, R., Gintis, H., Bowles, S., and Richerson, P. J. (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences of the United States of America*, 100(6), 3531–3535.
- Boyd, R. and Richerson, P. J. (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology*, 13(3), 171–195.
- Brock, W. A. (1970). An axiomatic basis for the Ramsey-Weizsäcker overtaking criterion. *Econometrica*, 38(6), 927–929.
- Brock, W. A. and Haurie, A. (1976). On existence of overtaking optimal trajectories over an infinite time horizon. *Mathematics of Operations Research*, 1(4), 337–346.

- 
- Casari, M. (2005). On the design of peer punishment experiments. *Experimental Economics*, 8(2), 107–115.
- Casari, M. and Plott, C. R. (2003). Decentralized management of common property resources: Experiments with a centuries-old institution. *Journal of Economic Behavior and Organization*, 51(2), 217–247.
- Cheney, D. L. and Seyfarth, R. M. (1990). The representation of social relations by monkeys. *Cognition*, 37(1-2), 167–196.
- Comité national du CNRS (2003). *Rapport des sections 01 et 02 du comité national du CNRS sur deux thèses de doctorat*. Retrieved from [http://www.marianne.net/docs/rapport\\_cnrs\\_bogdanoff.pdf](http://www.marianne.net/docs/rapport_cnrs_bogdanoff.pdf).
- Cowen, T. (1991). Self-constraint versus self-liberation. *Ethics*, 101(2), 360–373.
- Coyne, J. A. (2012) A Sokal-style hoax by an anti-religious philosopher. Retrieved from <http://whyevolutionistrue.wordpress.com/2012/09/25/a-sokal-style-hoax-by-an-anti-religious-philosopher-2/>.
- de Quervain, D. J.-F., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., and Fehr, E. (2004). The neural basis of altruistic punishment. *Science*, 305(5688), 1254–1258.
- de Waal, F. (1987). Dynamics of social relationships. In: Smuts, R., Cheney, D. L., Seyfarth, R., Wrangham, R. W., and Struhsaker, T. T.[eds.], *Dynamics of social relationships*, 421–429. University of Chicago Press, Chicago.
- DellaVigna, S., and Malmendier, U. (2004). Contract design and self-control: Theory and evidence. *The Quarterly Journal of Economics*, 119(2), 353–402.
- DellaVigna, S., and Malmendier, U. (2006). Paying not to go to the gym. *The American Economic Review*, 96(3), 694–719.
- Derrida, J. (1993). Structure, sign, and play in the discourse of the human sciences. *A postmodern reader*, 223–242.
- Deweese-Boyd, I. (2012). Self-deception. *The Stanford Encyclopedia of Philosophy (Spring 2012 Edition)*, Edward N. Zalta (ed.), Retrieved from <http://plato.stanford.edu/archives/spr2012/entries/self-deception>.
- Distler, J. *Bogdanorama*. Retrieved from <https://golem.ph.utexas.edu/~distler/blog/archives/000375.html>.

- Dreber, A., Rand, D., Fudenberg, D., and Nowak, M. A. (2008). Winners don't punish. *Nature*, 452(7185), 348–351.
- Dreber, A. and Rand, D. G. (2012). Retaliation and antisocial punishment are overlooked in many theoretical models as well as behavioral experiments. *Behavioral and Brain Sciences*, 35(1), 24.
- Egas, M. and Riedl, A. (2008). The economics of altruistic punishment and the maintenance of cooperation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 275(1637), 871–878.
- Ellis, L. (1995). Dominance and reproductive success among nonhuman animals: A cross-species comparison. *Ethology and Sociobiology*, 16(4), 257–333.
- Eliasz, K., and Spiegler, R. (2006). Contracting with diversely naive agents. *The Review of Economic Studies*, 73(3), 689–714.
- Elster, J. (1979). *Ulysses and the sirens: Studies in rationality and irrationality*. Cambridge, Cambridge University Press.
- Engelmann, D. and Nikiforakis, N. (in press). In the long-run we are all dead: On the benefits of peer punishment in rich environments. *Social Choice and Welfare*, March 2015, 1–17.
- Epstein, W. M. (1990). Confirmational response bias among social work journals. *Science, Technology & Human Values*, 15(1), 9–38.
- Eriksson, K. (2012). The nonsense math effect. *Judgment and decision making*, 7(6), 746–749.
- Falk, A., Fehr, E., and Fischbacher, U. (2005). Driving forces behind informal sanctions. *Econometrica*, 73(6), 2017–2030.
- Fehr, E. and Fischbacher, U. (2003). The nature of human altruism. *Nature*, 425(6960), 785–791.
- Fehr, E. and Gächter, S. (2000). Cooperation and punishment in public goods experiments. *The American Economic Review*, 90(4) 980–994.
- Fehr, E. and Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415(6868), 137–140.
- Feyerabend, P. (1993). *Against method*. Verso.

- 
- Ficke, A. D. (1916). *Spectra: A Book of Poetic Experiments*. Mitchell Kennerley.
- Fischer, C. (1999). Read this paper even later: Procrastination with time-inconsistent preferences. *Discussion Paper 99-20*, Resources for the Future.
- Flew, A. (1954). The justification of punishment. *Philosophy*, 29(111), 291–307.
- Foucault, M. (2002). *The order of things: An archaeology of the human sciences*. Psychology Press.
- Fowler, J. H. (2005). Second-order free-riding problem solved? *Nature*, 437(7058), E8.
- Frederick, S., Loewenstein, G., and O'Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature*, 40(2), 351–401.
- Fudenberg, D. and Levine, D. K. (1983). Subgame-perfect equilibria of finite- and infinite-horizon games. *Journal of Economic Theory*, 31(2), 251–268.
- Fudenberg, D., and Levine, D. K. (2006). A dual-self model of impulse control. *The American Economic Review*, 96(5), 1449–1476.
- Fudenberg, D., and Levine, D. K. (2012). Timing and self-control. *Econometrica*, 80(1), 1–42.
- Gale, D. (1967). On optimal development in a multisector economy. *The Review of Economic Studies*, 34(1), 1–18.
- Gächter, S. and Herrmann, B. (2009). Reciprocity, culture and human cooperation: Previous insights and a new cross-cultural experiment. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1518), 791–806.
- Gächter, S., Renner, E., and Sefton, M. (2008). The long-run benefits of punishment. *Science*, 322(5907), 1510.
- Gilpatric, S. M. (2008). Present-biased preferences, self-awareness and shirking. *Journal of Economic Behavior & Organization*, 67(3), 735–754.
- Gleick, J. (2012) *The information: A history, a theory, a flood*. Vintage Publishing.
- Gross, N., and Simmons, S. (2007). The social and political views of American professors. Working Paper, Department of Sociology, Harvard University.

- Gruber, J., and Kőszegi, B. (2001). Is addiction ‘rational’? Theory and evidence. *The Quarterly Journal of Economics*, 116(4), 1261–1303.
- Guala, F. F. (2012). Reciprocity: weak or strong? What punishment experiments do (and do not) demonstrate. *Behavioral and Brain Sciences*, 35(1), 1–15.
- Gul, F., and Pesendorfer, W. (2001). Temptation and self-control. *Econometrica*, 69(6), 1403–1435.
- Güth, W., Schmittberger, R., and Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization*, 3(4), 367–388.
- Hammond, P. J. (1976). Changing tastes and coherent dynamic choice. *The Review of Economic Studies*, 43(1), 159–173.
- Harris, C., and Laibson, D. I. (2001). Dynamic choices of hyperbolic consumers. *Econometrica*, 69(4), 935–957.
- Heidhues, P., and Kőszegi, B. (2009). Futile attempts at self-control. *Journal of the European Economic Association*, 7(2-3), 423–434.
- Herings, P. J. J., and Rohde, K. I. (2006). Time-inconsistent preferences in a general equilibrium model. *Economic Theory*, 29(3), 591–619.
- Herings, P. J. J., and Rohde, K. I. (2008). On the completeness of complete markets. *Economic Theory*, 37(2), 171–201.
- Herrmann, B., Thöni, C., and Gächter, S. (2008). Antisocial punishment across societies. *Science*, 319(5868), 1362–1367.
- Janssen, M. A. and Bushman, C. (2008). Evolution of cooperation and altruistic punishment when retaliation is possible. *Journal of Theoretical Biology*, 254(3), 541–545.
- Janssen, M. A., Holahan, R., Lee, A., and Ostrom, E. (2010). Lab experiments for the study of social-ecological systems. *Science*, 328(5978), 613–617.
- Jehiel, P., and Lilico, A. (2010). Smoking today and stopping tomorrow: A limited foresight perspective. *CESifo Economic Studies*, 56(2), 141–164.

- 
- Kahneman, D., Wakker, P. P., and Sarin, R. (1997). Back to Bentham? Explorations of experienced utility. *The Quarterly Journal of Economics*, 112(2), 375–406.
- Kiyonari, T. and Barclay, P. (2008). Cooperation in social dilemmas: Free riding may be thwarted by second-order reward rather than by punishment. *Journal of Personality and Social Psychology*, 95(4), 826–842.
- Kocherlakota, N. R. (1996). Reconsideration-proofness: A refinement for infinite horizon time inconsistency. *Games and Economic Behavior*, 15(1), 33–54.
- Kosfeld, M., Okada, A., and Riedl, A. (2009). Institution formation in public goods games. *American Economic Review*, 99(4), 1335–1355.
- Laibson, D. I. (1994). *Hyperbolic discounting and consumption*. Doctoral dissertation, Massachusetts Institute of Technology.
- Laibson, D. I. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*, 112(2), 443–478.
- Laibson, D. I., Repetto, A., Tobacman, J., Hall, R. E., Gale, W. G., and Akerlof, G. A. (1998). Self-control and saving for retirement. *Brookings Papers on Economic Activity*, 1998(1), 91–196.
- Lapied, A., and Renault, O. (2012). A subjective discounted utility model. *Economics Bulletin*, 32(2), 1171–1179.
- Loewenstein, G. (1996). Out of control: Visceral influences on behavior. *Organizational Behavior and Human Decision Processes*, 65(3), 272–292.
- Loewenstein, G. (2005). Hot-cold empathy gaps and medical decision making. *Health Psychology*, 24(4S), S49–S56.
- Liotard, J. F. (1984). *The postmodern condition: A report on knowledge*. (Vol. 10). University of Minnesota Press.
- MacKenzie, D. (2004). *Mechanizing proof: computing, risk, and trust*. MIT Press.
- Mahoney, M. J. (1977). Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive therapy and research*, 1(2), 161–175.

- Masclot, D., Noussair, C., Tucker, S., and Villeval, M. C. (2003). Monetary and nonmonetary punishment in the voluntary contributions mechanism. *American Economic Review*, 93(1), 366–380.
- Maskin, E., and Tirole, J. (2001). Markov perfect equilibrium: I. Observable actions. *Journal of Economic Theory*, 100(2), 191–219.
- McClennen, E. F. (1990). *Rationality and dynamic choice*. Cambridge University Press.
- Méder, Z. Z. (2005). The ‘proper’ words and things: The analytical language of John Wilkins. [in Hungarian] *Kellék*, 27–28, 293–314.
- Meltzer, T. (2014). Robot doctors, online lawyers and automated architects: the future of the professions? *The Guardian*, 2014 June 15. Retrieved from <http://www.theguardian.com/technology/2014/jun/15/robot-doctors-online-lawyers-automated-architects-future-professions-jobs-technology>.
- Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics*, 92(1–2), 91–112.
- Nikiforakis, N. and Normann, H.-T. (2008). A comparative statics analysis of punishment in public-good experiments. *Experimental Economics*, 11(4), 358–369.
- Nowak, M. A. and Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, 437(7063), 1291–1298.
- O’Donoghue, T., and Rabin, M. (1999a). Doing it now or later. *American Economic Review*, 89(1), 103–124.
- O’Donoghue, T., and Rabin, M. (1999b). Incentives for procrastinators. *The Quarterly Journal of Economics*, 114(3), 769–816.
- O’Donoghue, T., and Rabin, M. (2001). Choice and procrastination. *The Quarterly Journal of Economics*, 116(1), 121–160.
- Ohtsuki, H., Iwasa, Y., and Nowak, M. A. (2009). Indirect reciprocity provides only a narrow margin of efficiency for costly punishment. *Nature*, 457(7225), 79–82.

- 
- Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press, Cambridge.
- Ostrom, E., Walker, J., and Gardner, R. (1992). Covenants with and without a sword: Self-governance is possible. *The American Political Science Review*, 86(2), 404–417.
- Panchanathan, K. and Boyd, R. (2004). Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature*, 432(7016), 499–502.
- Peleg, B., and Yaari, M. E. (1973). On the existence of a consistent course of action when tastes are changing. *The Review of Economic Studies*, 40(3), 391–401.
- Perc, M. (2012). Sustainable institutionalized punishment requires elimination of second-order free-riders. *Scientific reports*, 2, 344–344.
- Peters, D. P., and Ceci, S. J. (1982). Peer-review practices of psychological journals: The fate of published articles, submitted again. *Behavioral and Brain Sciences*, 5(02), 187–195.
- Phelps, E. S., and Pollak, R. A. (1968). On second-best national saving and game-equilibrium growth. *The Review of Economic Studies*, 35(2), 185–199.
- Rand, D. G., Armao IV, J. J., Nakamaru, M., and Ohtsuki, H. (2010). Anti-social punishment can prevent the co-evolution of punishment and cooperation. *Journal of Theoretical Biology*, 265(4), 624–632.
- Rand, D. G. and Nowak, M. A. (2011). The evolution of antisocial punishment in optional public goods games. *Nature Communications*, 2, 1–7.
- Ramsey, F. P. (1928). A mathematical theory of savings. *Economic Journal*, 38(152), 543–559.
- Read, D., Frederick, S., Orsel, B., and Rahman, J. (2005). Four score and seven years from now: The date/delay effect in temporal discounting. *Management Science*, 51(9), 1326–1335.
- Robbins, B. (1996). On Being Hoaxed. Draft of article for *Tikkun* (circa August 1996). Retrieved from [http://www.physics.nyu.edu/sokal/robbins\\_prelim.html](http://www.physics.nyu.edu/sokal/robbins_prelim.html).

- Robbins, B. and Ross, A. (1996). Response by Social Text Editors Bruce Robbins and Andrew Ross. *Lingua Franca*, July/August 1996.
- Rockenbach, B., Guererck, O., and Irlenbusch, B. (2006). The competitive advantage of sanctioning institutions. *Science*, 312(5770), 108–111.
- Rosenthal, D. L. (1973). On being sane in insane places. *Science*, 179(4070), 250–258.
- Rubinstein, A. (1979). Equilibrium in supergames with the overtaking criterion. *Journal of Economic Theory*, 21(1), 1–9.
- Sáez-Martí, M., and Weibull, J. W. (2005). Discounting and altruism to future decision-makers. *Journal of Economic Theory*, 122(2), 254–266.
- Sarafidis, Y. (2004). Inter-temporal price discrimination with time inconsistent consumers. In: *Econometric Society 2004 North American Summer Meetings (No. 479)*, Econometric Society.
- Schelling, T. C. (1960). *The strategy of conflict*. Harvard University Press.
- Sefton, M., Shupp, R., and Walker, J. M. (2007). The effect of rewards and sanctions in provision of public goods. *Economic Inquiry*, 45(4), 671–690.
- Sigmund, K., De Silva, H., Traulsen, A., and Hauert, C. (2010). Social learning promotes institutions for governing the commons. *Nature*, 466(7308), 861–863.
- Sokal, A. D. (1996a). Transgressing the boundaries: Toward a transformative hermeneutics of quantum gravity. *Social Text* 46/47, 217–252.
- Sokal, A. D. (1996b). A physicist experiments with cultural studies. *Lingua Franca*, 6(4), 62–64.
- Sokal, A. D. (1996c). Sokal's Reply to Social Text Editorial. *Dissent*, Winter 1997.
- Sokal, A., and Bricmont, J. (1998). *Fashionable nonsense: Postmodern intellectuals' abuse of science*. Macmillan.
- Steel, P. (2007). The nature of procrastination: A meta-analytic and theoretical review of quintessential self-regulatory failure. *Psychological Bulletin*, 133(1), 65–94.

- 
- Strotz, R. H. (1956). Myopia and inconsistency in dynamic utility maximization. *The Review of Economic Studies*, 23(3), 165–180.
- Thaler, R. (1981). Some empirical evidence on dynamic inconsistency. *Economics Letters*, 8(3), 201–207.
- Taylor, M., Wedel, M., and Naish, D. (2013). Anti-tutorial: how to design and execute a really bad study. *Sauropod Vertebra Picture of the Week*. Retrieved from <http://svpow.com/2013/10/07/anti-tutorial-how-to-design-and-execute-a-really-bad-study/>.
- Thaler, R. H., and Shefrin, H. M. (1981). An economic theory of self-control. *The Journal of Political Economy*, 89(2), 392–406.
- Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, 46(1), 35–57.
- Vieille, N., and Weibull, J. W. (2009). Multiple solutions under quasi-exponential discounting. *Economic Theory*, 39(3), 513–526.
- Walker, J. M. and Halloran, M. A. (2004). Rewards and sanctions and the provision of public goods in one-shot settings. *Experimental Economics*, 7(3), 235–247.
- von Weizsäcker, C. C. (1965). Existence of optimal programs of accumulation for an infinite time horizon. *The Review of Economic Studies*, 32(2), 85–104.
- Yaari, M. E. (1978). Endogenous changes in tastes: A philosophical discussion. in: Eberlein, G. and Leinfellner, W.[eds.]: *Decision theory and social ethics*. Springer Netherlands, 59–98.
- Yamagishi, T. (1986). The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology*, 51(1), 110–116.
- Yamagishi, T. (1988). Seriousness of social dilemmas and the provision of a sanctioning system. *Social Psychology Quarterly*, 51(1), 32–42.

## **Brief curriculum vitae**

Zsombor Z. Méder was born in Oradea, Romania, July 2<sup>nd</sup>, 1984. He graduated from high school in 2002, and started studying for a degree in Philosophy. In 2003, he became a fellow at Eötvös József College, Budapest. He received his Master's degree in Economics from the Corvinus University of Budapest in 2009. Shortly thereafter, he started working towards a doctoral at the School of Business and Economics, Maastricht University. In 2013-14 he was a visiting researcher at Waseda University, Tokyo.