

Being under threat: Deterrence, social systems, and the evolution of social consciousness. Chapter 1 Introduction

August 6, 2026

Contents

1	Introduction	2
1.1	Present aims	2
1.2	The field of inquiry	5
1.2.1	Diverse minds/agents	6
1.2.2	Experiential consequences	6
1.2.3	The architecture of social control	6
1.3	The structure of this book	7

NOTE: this is a working draft, not a finished product. Feel free to contact me with constructive feedback, but please do not quote or distribute this draft, although you may link to this page, which will be updated.

Here's a [talk](#) I gave a couple of years ago early in the process of thinking about this stuff.

1 Introduction

1.1 Present aims

This work is about *coercion* (also known as *deterrence*), which is using a threat to force someone else to do something they don't want to do, or to refrain from doing something they want to do. This phenomenon is something that every reader will be familiar with, in many forms. A dog may growl if you try to take away the bone it's chewing on. Parents may threaten their children with timeout, spanking, or removal of privileges to get them to behave compliantly. Churches threaten hell and damnation. Governments threaten their citizens with police violence and imprisonment, and threaten one another with warfare and nuclear apocalypse.

Deterrence is one of those things that is so familiar that it might seem to be simple, well-understood, or undeserving of analysis. The fact that all objects just fall down to the ground might seem that way too, if you've never thought about it too much. Deterrence is all around us, like gravity or other natural forces. As I will argue, it is something that naturally arises through the interaction, not of massive bodies per se as in the case of gravity, but of people and other *agents* who are capable of perceiving each others' agency. The potential for one agent to coerce another arises as soon as one perceives that the other tries to avoid being harmed; coercion is the use of that tendency toward harm-avoidance as a way to control the other agent's behavior. Coercion phenomena are staggeringly diverse: the simplest forms of coercion behavior, probably anti-predator deterrence, likely emerged early in animal evolution, simultaneously with the abilities to perceive and predict the behavior of other animals. More elaborated forms of coercion embedded in complex social structures have emerged many times across the tree of animal life. Studying these natural experiments can reveal a great deal about how the situational logic of threat shapes the lives of animals, and how coercion can interact with other social forces, such as conflict, cooperation, and positive affiliation or bonding. The second part of this book will explore the diversity of coercion/deterrence phenomena found in nonhuman animals, and consider questions about the evolutionary relationships between coercion and fundamental aspects of the mind, such as memory and prediction, motivational systems, social cognition, and consciousness. I will also consider the implications of the theory of coercion developed here for animal welfare and the design of humane systems for interacting with domesticated and wild animals.

Tremendous variation on the theme of coercion has been explored extensively by humans, and many of these experiments have resulted in vast amounts of suffering and can be regarded as great tragedies of history: Slavery of all types, repressive moral and criminal codes enforced by

torture and death, various forms of exploitative class, caste, and other apartheid systems based on race, gender, religion, or any factor, all typically exist as structures of coercive violence. Such systems of mass coercion have shaped human history for thousands of years. In the early 21st century, nearly the entire human population of the Earth constitutes a single contiguous social system, wherein the world's human populations are partitioned under a small number of control hierarchies (imperial governments) whose power is grounded in their technological capacities for surveillance and violence. Most human lives are constantly being monitored for violations of rules, which have at least some chance of being punished with swift, impersonal, professional violence. Individuals are subject to local and regional governments which are subject to coercive governance by higher political entities, all of which exist in an 'international community' of nation-states, the balance of power among these entities largely being a matter of their ability to inflict military (especially nuclear), and economic violence upon one another. Coercion can be said to be the dominant strategy in behavioral control in use by many of the world's institutions at a variety of scales (Sidman, Schelling, Foucault). The third section of this book explores the diversity of forms of coercive social control in human societies, and considers various theoretical perspectives from which these have been approached.

Seeing coercion as an emergent natural force in social life may at first suggest a pessimistic outlook regarding human society. I am not a defeatist, and to the contrary I see this project as driven by the fundamental optimism that understanding the world can help us to improve our situation. The existence of gravity does not mean that we are doomed to fall into a black hole, or even that we are unable to stand on our feet. Indeed, understanding gravity has allowed us to design bridges and airplanes. We can use gravitational lensing to sense blackholes thousands of lightyears away, and we can use gravitational slingshotting to project our technological presence into the outer solar system. We can use our understanding of gravity and its relation to other forces to change our relationship to gravity and in a sense free ourselves from what previously seemed like its unshakeable constraints. Knowledge is power.

It is time to update longstanding assumptions—misconceptions, as I shall argue—about human nature, consciousness, and the situational logic of risk, reward and uncertainty. These entrenched misunderstandings have resulted in the design, justification, and implementation of coercive institutions and policies at many scales that are far from optimal, if judged in terms of avoiding unnecessary suffering. In many cases, such institutions have, either openly or covertly, been designed precisely to produce a dynamic of coercive control. Sometimes this is supposedly from good intentions combined with mistaken beliefs about psychology. Often, this is done in bad faith, and other functions of the institution have served mainly as cover for coercive exploitation.

Many people blame human nature or fall to some other form of defeatism and inevitability about the situation. However, blaming human nature for this is like blaming gravity for a bridge failure. Poor design is not human nature; learning from error is the only way that knowledge can accumulate in this universe. The proper response to the tragedies of history is to learn what we can from them and do better.

We can do better, and understanding the ways our institutions function and exert their influence

on us and the other human beings that compose them, is essential. Many of these institutions are fundamentally coercive in that they rely on a structure of punishments contingent on non-compliance in order to incentivize the participation actually makes them up. Coercion is, in a sense, the life-blood or structural scaffold (the bones and tendons? The muscle by which one part exerts influence on another?) of many social institutions. Not all social institutions, which could in theory rely on pure reward inducement or cooperation or something else(Sidman)

My hope is that an improved theoretical understanding of coercion/deterrence can also be used to improve the design and function of existing institutions and to design new institutional models that rely less on coercion and more on other forms of agentic interaction—such as reward inducement and cooperation—and as a result, can be more effective, efficient, and reliable, as well as much more humane.

This is not pure idealistic fantasy. For example, in much of the world, over the last 50 years, a culture shift has taken place regarding the use of physical punishment in child rearing, with campaigns in many countries dedicated to educating and encouraging parents to use nonviolent means to shape their children's behavior. (Refs...) Similarly, the use of punishment to train animals has been largely shown to be ineffective, as well as inhumane, and in many cases practice has yielded to science. It may be in the future that, for a teacher to beat a student (or a trainer beat a dog) in order to help them learn, will be seen everywhere in the world as foolish and counterproductive, whatever the moral evaluation of the situation, much as pouring water on a fire to help it burn would be seen as foolish and counterproductive. This will not people stop being cruel to each other everywhere in the world, and in some places, certain educational systems are (arguably) designed as much *to* traumatize people and deter them from attaining higher education and the subsequent types of empowerment it can facilitate, as they are for any other purpose; educational institutions lie at a complex intersection of the political and biological, and are often have aims other than education. However, where they have succeeded and continue to evolve, educational reforms that break the toxic association between *teachers* and *violence*, have had far-reaching consequences (refs). While overt bodily violence has been removed from many educational settings (particularly in those countries such as the United States, Britain, and France, which were actually actually the colonial powers that exported these techniques all over the world), more subtle forms of coercive control, however, are still endemic to educational institutions, even in well-resourced populations, and still exert demonstrably harmful effects (Kohn refs...).

In the long run, I believe an improved understanding of incentive logic and the way it relates to human psychology and the experiential dimension of agency, can be a key component in a revolution in the design of social systems, from the level of the family and local communities to the international community of global powers and the design of global policies, such that social relations at all of these levels can be based less on the threat of force (or the use of extrinsic negative reinforcers like money which are misleadingly framed as rewards), and more on the intrinsic motivation of cooperation and the positive affiliation resulting from the pursuit of shared goals. It can also help humans relate to other sentient beings with more compassion and the ability to create better outcomes for all parties. As I shall argue in what follows, my

position is not a leap of faith or an expression of naive humanistic optimism, but a logical extrapolation from available scientific data. In short, as I will argue in what follows, a wide array of empirical data across domains shows that coercion is both inefficient and unreliable either as a technique for creating stable social structures, or for controlling group or individual behavior.

The many mistakes of history do not show that humanity has an evil nature, whatever you mean by ‘evil’ means. History shows that situations can arise that evoke evil behavior from humans. But history also shows that situations can arise that evoke stable arrangements characterized by peaceful, cooperative behavior.

Humans, like all agents, adapt our behavior to the incentive logic of our situation. Unlike most creatures on earth, we humans can use our collective knowledge and power to shape our situations of life in collectively intentional ways, even at global scales. Knowledge can only be gained by learning from our individual and collective experience, and a quarter of the way into the 21st century, we have unprecedented access to collective human experience.

Humanity does not need to be saved from itself, either by an enforcer god or an enforcer state imbued with godlike powers to surveil and punish. Our failures are no reason to think we cannot succeed; how else could we have learned? But if we want to experience a paradise of peace and harmony, we must work to cultivate it ourselves, it will not be given as a reward for submission to coercive violence.

I believe that we are now in a position to learn to engineer our own systems to allow us to live more cooperatively and freely than we ever have in human history. To do that, we must have the intellectual courage to understand, deconstruct, and challenge the systems that, as humans, keep us enslaved to one another. To do that, we must really understand the difference between cooperation and freedom on the one hand, and coercion—living under the threat of violence—at social, economical, psychological, biological, and phenomenological levels. That is the purpose of this book.

1.2 The field of inquiry

My approach is broadly interdisciplinary and integrative, in the sense of attempting to bring together innovations from distinct theoretical paradigms developed in different contexts. I see this as inevitably therefore a ‘science studies’ project, at least in part; it is a process of studying the production and application of knowledge, what works and what doesn’t within, among various paradigms, considering where lessons learned from one can be applied to another, and attempting to integrate or unify the theoretical insights achieved in the different domains into a single theory that can cover all the cases, hopefully offering additional power to the study of each case.

1.2.1 Diverse minds/agents

Coercion and deterrence have been of interest to a broad array of sciences, focusing on systems including the behavior of individual nonhuman animals (e.g. crabs, rodents, monkeys) and the evolution of complex social structures in certain nonhuman animals such as ants and wasps, as well as individual and collective social behavior in humans, from a variety of perspectives including family and relationship psychology, educational psychology and institutional design, the design of systems of laws and punishments to enforce them at national levels, and military/diplomatic strategic relations at international levels (Table 1).

1.2.2 Experiential consequences

I also see this as a project in applied consciousness studies; specifically, I think what has been largely deficient in studies of coercion/deterrence in many domains has been adequate understanding of agency and consciousness. A theme running through all literatures on deterrence I've reviewed is the tension between economic, rational, situational models and subjective, psychological models. Most theorists have fallen hard into one camp or another, and nobody has resolved the relation between these aspects, much less integrated them adequately.

Consciousness studies aims to model the diversity and dynamics of conscious experience, and to understand how the subjective experience of consciousness maps to conditions in 'objective'* reality. What kinds of conscious states are we capable of experiencing, and how does this relate to the experience of other humans and other creatures, and to the broader space of possible experience? In which animals and other systems do different types of experience occur, in which individuals, and which situations?

These issues, in all their detail, matter for human and nonhuman welfare.

1.2.3 The architecture of social control

Humans spend a lot of energy, time, and resources at least supposedly attempting reducing and minimize suffering, through things like designing humane social and technological systems, policies and programs, biomedical technologies like drugs and other therapy modalities. Many such interventions explicitly appeal to improving life or reducing suffering (directly or indirectly) in their stated purpose. This is true for coercive institutions too ('liberal' ones especially). Critiquing those and recommending alternative possibilities falls into the scope of applied consciousness studies, as does offering theoretical and methodological analysis and modeling to support those critiques and recommendations.

Table 1: Some major studies on coercion/deterrence from different perspectives.

Target system	Disciplinary paradigm	Examples
Animal behavior	Behavioral ecology, animal behavior, biology	Parker 1974 “Assessment Strategy and the Evolution of Fighting Behavior”
Composition of animal societies	Behavioral ecology, animal behavior, biology	Tibbetts EA, Pardo-Sanchez J, Weise C. 2022 “The establishment and maintenance of dominance hierarchies.”
Human psychology	Behavior Analysis, learning theory, educational psychology	Sidman, “Coercion and its Fallout”; Alfie Kohn, “Punished by Rewards”
Penal institutions, crime and law enforcement	Criminology, political theory, philosophy	Cesare Beccaria, “On Crimes and Punishments”; Bentham, “Panopticon, or The Inspection House”; Foucault, “Discipline and Punish”
Warfare, politics, and diplomacy	Military theory, game theory, economics, political science	Schelling, “Arms and Influence”
Interplanetary contact	Messaging Extra-Terrestrial Intelligence (METI), Space Policy, Risk management Theory	Santana 2021 “We Come in Peace? A Rational Approach to METI”

1.3 The structure of this book

the idea is to ease in with some preliminaries

case study review first, referring forward to the detailed analyses and arguments in the glossaries and appendix, allowing readers to deep dive into interesting details without getting bogged down in the formal definitions and heavier analysis

Part 1: The framework and the angle of approach

- Coercion in relation to other social modalities
 - Force
 - Reward
 - Communication
 - Cooperation
 - Charisma
- Key Concepts
 - Coercion

- Agency
- Consciousness
- Perspective
- Suffering

Part 2: Case review

- Forms of Coercion in Nonhuman Animals
- Distinctively Human Forms of Coercion

Part 3: Analysis

- Structure and dynamics of coercive systems: perspective, prediction, and paradox
- Focal case studies (?)

Part 4: Appendices

- Components of Wellbeing
 - Pursuing Goals
 - Managing Uncertainty
 - Safety and rest
 - Play and exploration
 - Social relationships
 - Relationships to places and things
- Appendix: history
- Appendix: The evolution of Bodies and Minds ???
 - Cognitive Prerequisites to controlling a complex body
 - The structure of consciousness
 - Consciousness as a scaffold for embodied action
- Appendix: Key authors