



UNIVERSITÀ
DI PAVIA

**DEPARTMENT OF ECONOMICS AND
MANAGEMENT**

MASTER PROGRAMME IN INTERNATIONAL
BUSINESS AND ENTREPRENEURSHIP

CAUGHT IN THE GAP
Cognitive Biases in Entrepreneurial
Decision-Making and the Regulatory Lag of
Technology Innovation

Supervisor:

Dr. Fabio Marazzi

Master thesis
by Constantin Botnaru

Matr. n.505696

Academic Year 2025 -2026

LIST OF CONTENTS

Abstract in italian.....	4
Chapter 1 — Introduction	5
1.1 The Nature of our Minds.....	5
1.2 Two Actors, Two Kinds of Bias	6
1.3 What Has Happened, and What Could	9
1.4 Research Question and the Argument in Outline.....	10
1.5 Structure of the Thesis	12
1.6 A Final Premise	14
Chapter 2 — Cognitive Bias in Entrepreneurial Decision-Making	15
2.1 The Entrepreneurial Mind Under Uncertainty	15
2.2 The Cognitive Foundations of Entrepreneurial Judgment	16
2.3 Optimism Bias and Overconfidence	18
2.4 The Planning Fallacy	20
2.5 Confirmation Bias and Motivated Reasoning.....	22
2.6 The Affect Heuristic and Narrative Thinking	23
2.7 Is Bias Always Dysfunctional?	24
2.8 From Individual Bias to Systemic Effect.....	27
Chapter 3 — Cognitive Bias in Regulatory and Legislative Decision- Making.....	29
3.1 The Regulatory Mind Under Complexity	29
3.2 The Pacing Problem and the Collingridge Dilemma	30
3.3 Status Quo Bias and Institutional Inertia	32
3.4 The Illusion of Understanding	34
3.5 The Availability Heuristic and the Regulation of the Last Crisis.....	36
3.6 Anchoring and the Limits of Analogy.....	38
3.7 Artificial Intelligence as the Central Case.....	40
3.8 Why Expertise Is Not a Solution.....	43
3.9 From Individual Bias to Institutional Pattern.....	45

Chapter 4 — The Cognitive Gap Model.....	48
4.1 Toward an Integrated Framework	48
4.2 The Structure of the Gap	50
4.3 The Interaction Effect	53
4.4 The Model at Work: The Case of Social Media.....	56
4.5 The Model in Relation to Adjacent Theories	58
4.6 The Gap Is Cognitive, Not Just Structural	61
4.7 The Falsifiability of the Model	63
4.8 Systemic Consequences	64
Chapter 5 — Implications and Conclusions	67
5.1 From Diagnosis to Response	67
5.2 Implications for Entrepreneurs.....	69
5.3 Implications for Policymakers	72
5.4 The Third Actor: Capital and the Amplification of Optimism.....	76
5.5 Applying the Framework: What Better Governance Would Require	79
5.6 Limitations	82
5.7 Conclusions.....	85
Bibliography —	89
Web contents —.....	93

ABSTRACT IN ITALIAN

La tesi analizza il divario persistente tra innovazione tecnologica e regolamentazione giuridica, sostenendo che tale divario non sia meramente strutturale ma soprattutto cognitivo. L'analisi esamina due attori centrali. Gli imprenditori, il cui bias ottimistico accelera lo sviluppo e la diffusione tecnologica oltre la capacità della società di valutarne i rischi. I legislatori, le cui limitazioni cognitive — recency bias, illusione di comprensione, status quo bias e ancoraggio ad analogie ereditate — producono una regolamentazione reattiva, incoerente e disallineata rispetto alle realtà dell'innovazione. Il contributo originale consiste nel Cognitive Gap Model, che integra le due dimensioni e individua il meccanismo della loro interazione: il circuito di retroazione che manterrebbe allineati innovazione e governance risulta disabilitato a entrambe le estremità, poiché il segnale di rischio viene soppresso sul versante imprenditoriale prima di essere trasmesso e frainteso su quello regolatorio al momento della ricezione. Il divario che ne deriva è moltiplicativo anziché additivo, e tende ad ampliarsi anziché stabilizzarsi. Da questa diagnosi discende una conclusione costruttiva: se il divario ha origine cognitiva, esso è almeno in parte affrontabile — non esortando gli attori a pensare meglio, ma progettando istituzioni che compensino i limiti cognitivi di chi le abita.

Chapter 1 — Introduction

1.1 The Nature of our Minds

Every significant decision a human being makes is, to some degree, the product of a mind operating under constraints it cannot fully see. These constraints are not failures of intelligence or education. They are structural features of human cognition — patterns of judgment that have been documented, measured, and theorised across decades of research in psychology and behavioural economics. They go by many names: overconfidence, the planning fallacy, confirmation bias, recency bias, the illusion of understanding. Collectively, they describe a mind that is less rational than it believes itself to be, and more predictable in its irrationality than it would be comfortable to admit.

This observation, which might seem abstract in isolation, takes on a very concrete significance when applied to the actors who shape the world we live in. Entrepreneurs build the companies, products, and systems that define the economic landscape of each generation. Legislators and regulators establish the rules within which those companies operate, those products circulate, and those systems are held accountable. Both groups make consequential decisions under conditions of uncertainty, complexity, and time pressure. And both groups, by virtue of being human, do so through a cognitive apparatus that is prone to systematic error in precisely those conditions.

The consequences of this are not evenly distributed. When an individual makes a biased judgment about a personal decision, the effects are localised. When an entrepreneur makes a biased judgment about the risks of a technology they are bringing to market, the effects may extend to thousands or millions of people who had no part in that decision. When legislators, shaped by their own cognitive limitations, respond to technological change in ways that are reactive, incoherent, or anchored in obsolete categories, the effects extend to entire economies and legal

systems. Cognitive bias, in other words, is not merely a psychological curiosity. At scale, it is a force that shapes institutions, markets, and societies.

The pages that follow are concerned with precisely this force. They trace how cognitive bias operates in the minds of those who create new realities and those who are tasked with governing them — and what happens, at a societal level, when both sets of minds are subject to distortions that pull in opposite directions. It is a study not of individual error, but of the structural conditions under which error becomes systemic, recurrent, and consequential far beyond the individuals who commit it.

1.2 Two Actors, Two Kinds of Bias

The entrepreneur is, in many respects, a figure defined by the conditions under which intuition flourishes and deliberation struggles. Operating under radical uncertainty, competitive pressure, and the chronic resource constraints of a new venture, the entrepreneur makes decisions that are necessarily rapid, often intuitive, and rarely amenable to the kind of slow, exhaustive analysis that sound judgment would theoretically require. In this environment, optimistic bias — the systematic tendency to overestimate the probability of success and underestimate the probability of failure — is not a peripheral distortion. It is a central feature of how entrepreneurs perceive their ventures, their markets, and their competitors. It shapes what gets built, how risks are communicated, and how setbacks are interpreted. And because entrepreneurs are not isolated individuals but actors whose decisions aggregate into market dynamics, industry trajectories, and societal transformations, the effects of this bias extend well beyond the individual venture.

The history of innovation is, in part, a history of this bias at work. Entrepreneurs who have built transformative companies have done so by maintaining conviction in the face of evidence that would have discouraged more cautious actors. This is

not incidental to their success — it is often constitutive of it. And yet the same cognitive disposition that drives the founding of a company also drives the underestimation of its risks, the dismissal of early warning signs, and the resistance to critical feedback. The planning fallacy has derailed ventures that were technically sound but managerially blind to the distance between a vision and its execution. Confirmation bias has led founders to build products for markets they imagined rather than markets that existed, filtering out the signals that would have corrected their course. These are not stories of incompetence. They are stories of cognition operating exactly as it tends to operate under pressure: selectively, optimistically, and with far greater confidence than the evidence warrants.

At a broader level, when optimistic bias is shared across an entrepreneurial ecosystem — when an entire sector is animated by the collective conviction that a given technology will transform the world — the consequences scale accordingly. Capital flows toward ventures whose risks have been systematically underweighted. Competitive pressure accelerates the pace of deployment beyond what safety or preparedness would justify. The social effects of the technology are addressed later, if at all, because the cognitive posture of the ecosystem was oriented toward possibility rather than consequence. The speculative booms that recur throughout the history of technological innovation follow a recognisable pattern: collective optimism, accelerated diffusion, and a societal confrontation with consequences that the original actors had not seriously contemplated.

The legislator operates under a different set of pressures, but arrives at equally systematic distortions. The task of governance requires forming judgments about phenomena that are technically complex, rapidly evolving, and politically contested. Recency bias leads attention to concentrate on the most recently visible events rather than on the most structurally significant trends. The illusion of understanding — the tendency to believe one comprehends a system more deeply than one actually does — produces regulatory frameworks that are confident in their framing but misaligned with the reality they purport to address. Status quo

bias generates institutional resistance to adaptive approaches that would require departing from established legal categories. And anchoring leads regulators, confronted with something genuinely new, to reach for the nearest familiar analogy and to govern the unprecedented as though it were a variation on something already known. The result, taken together, is a regulatory posture that is structurally reactive: it responds to what has already happened, classifies it through frameworks designed for what already existed, and moves at a pace determined not by the speed of the phenomenon being governed but by the inertia of the institutions doing the governing.

The practical consequences of this are visible across decades of technological governance. Financial derivatives were regulated, to the extent they were regulated at all, through frameworks designed for simpler instruments — until the crisis of 2008 made the inadequacy of those frameworks impossible to ignore. Online platforms were treated, for years, as neutral conduits of information, a classification with significant legal consequences that persisted long after the evidence suggested it was inaccurate. In each case, the regulatory response came after the harm had materialised, was shaped by the specific form that harm had taken rather than by its underlying causes, and was anchored in categories that the phenomenon itself had already rendered obsolete.

What is particularly significant — and what this thesis seeks to demonstrate — is that these two sets of biases do not operate independently. They interact. The optimistic bias of entrepreneurs accelerates the pace at which new technologies and systems enter the world, compressing the time available for society to evaluate their implications. The cognitive constraints of legislators slow the pace at which governance responds, extending the period during which those technologies operate in a regulatory vacuum. The result is a gap that is not merely additive but multiplicative: each side's biases amplify the effect of the other's, producing a systemic misalignment that is greater than the sum of its parts.

1.3 What Has Happened, and What Could

The dynamics described above are not hypothetical. They have already produced measurable consequences, and the pattern they follow is sufficiently consistent across contexts to warrant serious analytical attention.

Consider the trajectory of a technology sector in its early stages. A small number of entrepreneurs, animated by a genuine belief in the transformative potential of what they are building, attract capital, talent, and public attention. Their optimism is infectious — and in many respects, it needs to be, because the alternative is paralysis in the face of genuine uncertainty. But as the technology matures and diffuses, the gap between the vision that drove its creation and the reality it has produced begins to emerge. Side effects appear that were not anticipated: on markets, on privacy, on social behaviour, on labour. The entrepreneurs who created the technology were not necessarily reckless or dishonest. They were, in many cases, simply operating under a cognitive architecture that made those side effects difficult to foresee and easy to discount.

At this point, the task of governance becomes urgent. But the institutions responsible for it are already operating with a delay — because regulation, by its nature, responds to what exists rather than to what is emerging. And the cognitive biases of those institutions compound this structural delay. Recency bias means that regulatory attention focuses on the most dramatic recent manifestation of a problem rather than on its systemic roots. The illusion of understanding means that legislators reach for the nearest available analogy rather than developing new conceptual tools for a genuinely novel situation. Status quo bias means that departing from established legal categories requires a level of political will and intellectual effort that institutions rarely sustain in the absence of acute crisis.

The result is a recurring pattern in the history of technological governance: a period of rapid, largely ungoverned innovation, followed by a crisis that makes the governance gap impossible to ignore, followed by a regulatory response that

addresses the specific form the crisis took without necessarily resolving the underlying conditions that produced it. This pattern has repeated across sectors and decades, and there is little in the current institutional landscape to suggest it will not repeat again.

What could happen — and what this thesis argues is already beginning to happen in the domain of advanced technology — is a version of this pattern at a scale and speed that makes past episodes look manageable by comparison. The technologies now being developed and deployed are more powerful, more pervasive, and more opaque in their consequences than most of their predecessors. The entrepreneurial ecosystems around them are larger, better capitalised, and more globally distributed. The regulatory systems tasked with governing them are more fragmented, more technically underprepared, and more subject to the cognitive limitations described above than the scale of the challenge requires.

None of this is inevitable. The point of analysing cognitive bias is not to suggest that human judgment is irredeemably flawed, but to identify the specific conditions under which it is most likely to fail — and, by doing so, to create the conditions for doing better. An entrepreneur who understands the mechanisms of optimistic bias is better positioned to seek out the disconfirming evidence that bias tends to suppress. A legislator who understands the mechanisms of recency bias and the illusion of understanding is better positioned to resist the pull of the familiar and to engage seriously with what is genuinely new. This is the practical wager that underlies the theoretical work of this thesis, and the concluding chapter returns to it in detail.

1.4 Research Question and the Argument in Outline

The central research question of this thesis is the following:

To what extent do cognitive biases in entrepreneurial and regulatory actors explain the persistent gap between technological innovation and its governance?

The question is framed in terms of extent because the thesis does not claim that cognitive bias is the only explanation for this gap. Structural, political, and institutional factors play important roles, and this analysis does not seek to displace them. What it seeks to establish is that the cognitive dimension is a necessary part of any adequate account — and that its omission from existing analyses leaves those analyses incomplete in ways that matter for both theory and practice.

The argument advanced in answer to this question can be stated in outline here, so that the reader knows from the outset where the analysis leads. It proceeds in three steps. The first two are diagnostic and treat each actor separately: entrepreneurial cognition, biased toward optimism and insulated from correction, systematically accelerates the diffusion of new technologies; regulatory cognition, biased toward the familiar and the recently vivid, systematically delays and misdirects the governance response. The third step is synthetic, and it is where the thesis makes its original contribution. It proposes an integrated framework, termed here the Cognitive Gap Model, which holds that these two sets of biases do not merely coexist but interact in a specific and damaging way. Between innovation and governance there runs, in principle, a corrective feedback relationship: acceleration generates risk signals, risk signals prompt governance responses, and governance responses restrain acceleration. The model's central claim is that cognitive bias disables this relationship at both ends — suppressing the risk signal on the entrepreneurial side before it can be transmitted, and misperceiving it on the regulatory side when it arrives. A corrective loop broken at both ends cannot restore alignment. The gap it fails to close therefore widens rather than stabilising, and it does so in a self-reinforcing rather than a self-limiting manner.

From this diagnosis the thesis derives a constructive conclusion. If the gap is cognitive in origin rather than the product of an immovable structural necessity, then it is, at least in part, addressable — not by exhorting entrepreneurs to be less optimistic or legislators to be more perceptive, which the evidence suggests is largely futile, but by designing institutions whose procedures compensate for the cognitive limitations of the people who staff them. Certain existing mechanisms, examined in the final chapter, prove on inspection to be almost purpose-built for this task, and the model explains why they work.

Four objectives organise the work. The first is to establish a rigorous account of the cognitive biases most consequential for entrepreneurial decision-making. The second is to examine how cognitive limitations at the individual and institutional level shape regulatory responses to innovation, compounding the structural constraints that already produce governance lag. The third is to construct the theoretical framework that integrates these two dimensions and analyses their interaction. The fourth is to derive from that framework implications for how entrepreneurs and policymakers might think and act differently.

1.5 Structure of the Thesis

The thesis proceeds in five chapters. The present chapter has framed the problem, introduced the two central actors and the cognitive dynamics that define their roles, and set out the research question together with the argument advanced in answer to it.

Chapter 2 develops the first pillar of the analysis. It examines cognitive bias in entrepreneurial decision-making, beginning with the conditions of radical uncertainty under which entrepreneurs operate and proceeding through the biases most consequential in that setting: optimism and overconfidence, the planning fallacy, confirmation bias and motivated reasoning, and the affective and narrative

modes of judgment that reinforce them. The chapter closes by confronting a genuine tension in the literature — the possibility that some degree of bias is functionally necessary for entrepreneurial action — and by showing how individual-level biases aggregate into the systemic acceleration that the model requires.

Chapter 3 develops the second pillar. It turns to regulatory and legislative cognition, first establishing the structural terrain through the pacing problem and the Collingridge dilemma, and then examining four biases that shape governance judgment: status quo bias, the illusion of understanding, the availability heuristic, and anchoring to inherited analogies. It applies this analysis to the governance of artificial intelligence as its central case, sets the European and American approaches side by side to show that the same cognitive patterns produce different symptoms under different institutional conditions, and confronts the objection that technical expertise might dissolve the problem — showing instead that expertise is entangled in it. The chapter concludes by demonstrating that institutions amplify rather than cancel the biases of the individuals who compose them.

Chapter 4 is the theoretical core of the thesis. It proposes the Cognitive Gap Model, situates it in relation to the existing literature on responsible innovation and anticipatory governance, and specifies the mechanism by which the two sides of the gap interact. It then puts the model to work on the documented trajectory of social media, distinguishes it from three adjacent theories with which it might be confused, defends it against the objection that the gap is purely structural, and — unusually for a theoretical construction — specifies the observations that would falsify it.

Chapter 5 turns from diagnosis to response. It examines what the analysis implies for entrepreneurs, resolving the tension between the functional value of optimism and the need for calibrated forecasting, and for policymakers, identifying institutional mechanisms that counteract specific documented biases. It introduces

a third actor largely implicit in the preceding analysis — the investor, whose capital allocation amplifies rather than filters entrepreneurial optimism — and applies the framework concretely to the governance of artificial intelligence. It closes with an honest account of the thesis's limitations and with its conclusions.

1.6 A Final Premise

There is a particular irony in writing a thesis about cognitive bias. The act of analysis is itself subject to the limitations it seeks to describe. The author brings to this work a perspective shaped by particular readings, particular intellectual commitments, and particular blind spots. The behavioural literature on which the analysis draws, for all its explanatory power, is not the only way to understand human judgment, and the synthesis proposed here is one among several that could be constructed from the available evidence. A model that coheres well is not thereby true; the narrative satisfaction of an argument is no guarantee of its correspondence to reality, as the analysis of entrepreneurial storytelling in Chapter 2 will itself observe.

This is not a disclaimer intended to diminish the analysis. It is, rather, an invitation to read it in the spirit in which it is offered: as a contribution to a conversation, not a conclusion to one, and as a provisional account subject to the same demand for disconfirming evidence that it will urge upon entrepreneurs and legislators alike. The gap between how institutions respond to change and how fast change actually moves is real, consequential, and poorly understood. If this thesis sharpens the terms in which that gap can be discussed and analysed, it will have achieved what it set out to do.

Chapter 2 — Cognitive Bias in Entrepreneurial Decision Making

2.1 The Entrepreneurial Mind Under Uncertainty

To understand why cognitive bias occupies such a central place in the study of entrepreneurial decision-making, one must begin with the conditions under which that decision-making takes place. The entrepreneur does not operate in the controlled environment of the laboratory or the stable context of an established corporation. The entrepreneur operates in a world defined by uncertainty — not merely risk, in the technical sense of outcomes whose probabilities can be calculated, but genuine uncertainty, in which the range of possible outcomes and their likelihoods are themselves unknown. This distinction, first articulated by Frank Knight in 1921, is fundamental. Risk can be managed through the tools of probability and expected value. Uncertainty resists those tools entirely, because the information required to apply them does not exist.

It is precisely in this environment that the human mind is most vulnerable to systematic error. When information is incomplete, when time is short, and when the stakes are high and personal, the deliberate, analytical mode of reasoning that sound judgment would theoretically require becomes difficult to sustain. In its place, the mind reaches for faster, more intuitive shortcuts — heuristics that produce rapid judgments at the cost of systematic distortion. These distortions are not random. They follow predictable patterns, and it is these patterns that the literature on cognitive bias has spent decades documenting.

The entrepreneurial context intensifies this vulnerability in several ways. First, the entrepreneur typically operates under acute resource constraints, lacking the time, personnel, and capital that would permit exhaustive analysis of every decision. Second, the entrepreneur is rarely a neutral observer of the situation: the venture is often a personal project, an extension of identity and ambition, which introduces

powerful motivational pressures into what is supposed to be a dispassionate assessment of prospects. Third, the very act of founding a venture requires a willingness to act in the face of uncertainty that more cautious individuals would find paralysing. The entrepreneur, in other words, is self-selected for a disposition toward action — and that disposition, as this chapter will argue, is bound up with the cognitive biases that shape entrepreneurial judgment.

This creates a paradox that runs through the entire chapter, and indeed through the entire thesis. The cognitive dispositions that make entrepreneurship possible may be the very same dispositions that make entrepreneurial judgment systematically unreliable. To found a company in the face of overwhelming odds requires a degree of optimism that the evidence does not justify. To persist through the inevitable setbacks requires a confidence that borders on, and often crosses into, overconfidence. The entrepreneur who saw the world with perfect accuracy might never start at all. This tension — between the functional necessity of bias and its systemic costs — is not a flaw in the analysis to be resolved, but a central feature of the phenomenon to be understood.

2.2 The Cognitive Foundations of Entrepreneurial Judgment

The study of cognitive bias belongs to a broad tradition in psychology and behavioural economics that, over the second half of the twentieth century, progressively dismantled the assumption of human rationality on which classical economic theory had been built. The central finding of this tradition is both simple and profound: human beings do not reason like ideal statisticians. They rely instead on heuristics — mental shortcuts that are efficient and often effective, but that produce systematic and predictable errors under identifiable conditions. Crucially, these errors are not the product of stupidity or carelessness. They are shared by experts and novices alike, they persist even when the individual is fully aware of

them, and they are, in this sense, a structural feature of the human mind rather than an incidental malfunction.

Applied to entrepreneurship, this insight has generated a rich body of research. A comprehensive review of the field identified more than fifty studies dedicated to cognitive bias in entrepreneurial contexts, structured around a typology of distinct biases, and found that the great majority of scholarly attention has concentrated on two in particular: overoptimism and overconfidence (Zhang & Cueto, 2017). Other biases — the illusion of control, the belief in the law of small numbers, escalation of commitment, the planning fallacy, status quo bias, hindsight bias, and the self-serving bias — have received comparatively less attention, though each has been shown to operate in entrepreneurial settings. This chapter concentrates on the biases with the strongest empirical grounding and the greatest relevance to the argument of the thesis, while situating them within this broader taxonomy.

An important qualification must be introduced at the outset. To describe entrepreneurial judgment as biased is to measure it against a particular standard of rationality — the standard of accurate prediction and probabilistic calculation. But there is a serious argument, developed most influentially by Saras Sarasvathy, that this standard is itself inappropriate to the entrepreneurial situation. In her theory of effectuation, based on the study of expert entrepreneurs, Sarasvathy (2001) argues that experienced entrepreneurs do not attempt to predict the future at all. Rather than beginning with a predetermined goal and calculating the optimal means to achieve it — the logic she calls causation — they begin with the means available to them and allow goals to emerge through action. They focus on what they can afford to lose rather than on expected returns, they treat unexpected contingencies as opportunities rather than obstacles, and they seek to control the future through action rather than to predict it through analysis.

This matters for the present analysis because it complicates the very notion of bias. If the entrepreneur is not trying to predict, then the failure to predict accurately is

not obviously an error. A judgment that appears biased when measured against the standard of prediction may be entirely appropriate when measured against the standard of effectual action under uncertainty. This chapter does not resolve this tension immediately — it returns to it in full in section 2.7 — but it flags it here because it shapes how every subsequent claim about bias must be read. To say that entrepreneurs are overconfident or overoptimistic is not, by itself, to say that they are behaving irrationally. It is to say that their judgments systematically depart from what accurate calculation would produce — a departure whose consequences may be, depending on the context, either functional or catastrophic.

2.3 Optimism Bias and Overconfidence

If a single finding could be said to define the study of entrepreneurial cognition, it would be the discovery of the sheer magnitude of entrepreneurial optimism. In a landmark study, Cooper, Woo & Dunkelberg (1988) surveyed 2,994 entrepreneurs about their perceived chances of success. The results were striking. Eighty-one per cent of the respondents believed their odds of success were seven in ten or better, and a remarkable thirty-three per cent believed their odds were a perfect ten out of ten — that is, that their success was certain. When asked to assess the prospects of other businesses like their own, the same entrepreneurs offered significantly lower estimates, revealing a clear asymmetry between how they judged themselves and how they judged others in identical situations.

These figures acquire their full weight only when placed against the empirical reality. Depending on the measure and the time horizon, only a small minority of new ventures survive and prosper; a large proportion fail within the first few years. The gap between an entrepreneurial population that overwhelmingly rates its own chances at seven in ten or better and a reality in which a majority will fail is not a small discrepancy at the margins. It is a systematic and massive distortion. Perhaps

the most telling detail of the Cooper study is that those entrepreneurs who were, by objective measures, poorly prepared for their ventures were just as optimistic as those who were well prepared. Optimism, in other words, was not tracking the actual quality of the venture. It was a disposition brought to the situation rather than a conclusion drawn from it.

Optimism and overconfidence are related but distinct. Optimism, in this literature, refers to the tendency to expect positive outcomes even when such expectations are not rationally justified. Overconfidence refers more specifically to an excessive certainty in the accuracy of one's own judgments and knowledge. The seminal demonstration of entrepreneurial overconfidence came from Busenitz & Barney (1997), who compared 124 entrepreneurs in the startup phase with a sample of managers in large organisations. They found that entrepreneurs were significantly more overconfident than managers — specifically in the form of overprecision, an excessive confidence in the accuracy of their own information — and more prone to reasoning from small or unrepresentative samples. This study has become one of the most cited in the field, and its central finding, that entrepreneurs think differently from managers in ways that expose them to characteristic errors, has been replicated and extended across numerous contexts.

The most illuminating evidence on overconfidence, however, comes not from surveys but from controlled experiment. Camerer & Lovallo (1999) designed a laboratory study to test whether overconfidence could produce a specific and economically significant behaviour: excess entry into competitive markets. Participants decided whether to enter a market in which payoffs depended on their rank relative to other entrants. The result was clear. Overconfidence led participants systematically to overestimate their own relative ability and, consequently, to enter markets more frequently than a rational assessment of the competition would warrant. Most strikingly, this effect intensified when participants had self-selected into the task knowing that success would depend on skill. Those who chose to compete precisely because they believed themselves

skilled were the most prone to the error. This finding offers a compelling model of real entrepreneurial markets, in which the entrants are not a random sample of the population but a self-selected group of individuals who believe, disproportionately, in their own abilities.

The consequences of this bias are double-edged, and the literature is careful to note both sides. On the one hand, overconfidence is associated with self-esteem, ambition, and the drive to act — qualities without which few ventures would ever be launched. On the other, it is associated with characteristic and costly errors: the underestimation of risk, the overinvestment in production capacity for products whose markets have not been adequately analysed, and the tendency to persist in failing ventures long past the point at which withdrawal would be rational. The tension between these two faces of overconfidence — its role as both an engine of entrepreneurial action and a source of entrepreneurial failure — is a theme to which this chapter returns.

2.4 The Planning Fallacy

Closely related to optimism bias, but distinct in its mechanism, is the planning fallacy: the systematic tendency to underestimate the time, cost, and difficulty of a future task while simultaneously overestimating its benefits. First identified by Kahneman and Tversky in the late 1970s and subjected to rigorous experimental study by Buehler, Griffin & Ross (1994), the planning fallacy is remarkable for its persistence. People underestimate how long their own projects will take even when they are explicitly reminded that their past projects took longer than expected. The bias survives direct confrontation with disconfirming evidence, which places it firmly within the category of systematic rather than incidental error.

The mechanism of the planning fallacy is instructive. According to Lovallo &

Kahneman (2003), it arises because planners adopt what they call the inside view: they focus on the specific components of the project before them — its particular tasks, its particular resources, its particular plan — and construct a forecast by assembling these details into a scenario of successful completion. What this inside view systematically excludes is the outside view: the actual historical record of how similar projects have unfolded. Since that record almost always includes delays, cost overruns, and failures that the current planners believe will not befall them, the inside view produces forecasts that are consistently and predictably too optimistic.

The business consequences of this fallacy are, in the assessment of Lovallo & Kahneman (2003), severe. Writing in the *Harvard Business Review*, they observe that most major business initiatives — mergers and acquisitions, capital investments, market entries — fail to deliver the returns their proponents forecast. The conventional economic explanation attributes this to rational risk-taking, in which the gains from occasional successes outweigh the losses from frequent failures. Lovallo and Kahneman offer a different account: the forecasts themselves are systematically distorted by a combination of cognitive biases, including anchoring, competitor neglect, and the inside view, which lead managers to exaggerate benefits and discount pitfalls. The initiatives are not rational bets with unlucky outcomes; they are the product of predictably flawed judgment.

The proposed remedy follows directly from the diagnosis. If the error stems from the inside view, the correction lies in adopting the outside view — a technique that has come to be known as reference-class forecasting. Rather than estimating a project's cost and duration from its internal details, the forecaster identifies a reference class of similar completed projects, establishes the actual distribution of outcomes within that class, and locates the current project within that distribution. The work of Bent Flyvbjerg on large infrastructure projects has operationalised this technique, developing an empirical 'optimism bias uplift' by which forecasts can be adjusted to account for the systematic underestimation that the historical

record reveals (Flyvbjerg, 2008). For the entrepreneur, the lesson is direct: the surest guide to how a venture will unfold is not the compelling specificity of one's own plan, but the record of how comparable ventures have actually fared.

2.5 Confirmation Bias and Motivated Reasoning

The biases considered so far concern the formation of expectations. Confirmation bias concerns their maintenance. It describes the tendency to seek out, attend to, and give disproportionate weight to information that confirms existing beliefs, while neglecting, discounting, or reinterpreting information that contradicts them. In the entrepreneurial context, this bias is particularly consequential, because it governs how the entrepreneur processes the continuous stream of feedback — from customers, from markets, from competitors — that ought, in principle, to correct an initial misjudgment before it becomes fatal.

The evidence suggests that this corrective feedback is often filtered out before it can do its work. Entrepreneurs with high levels of psychological capital — the very confidence and self-belief that sustain entrepreneurial effort — have been shown to exhibit a marked tendency to emphasise information consistent with their existing plans while largely ignoring information that contradicts them (Cossette, 2015; Gibson & Sanbonmatsu, 2004). The same confidence that enables the entrepreneur to act becomes the mechanism by which the entrepreneur screens out the signals that action is misdirected. Overconfidence, in this way, does not merely distort the initial decision; it insulates that decision from subsequent correction, leading entrepreneurs to expand their ventures despite equivocal or negative market feedback (Shepherd, Williams & Patzelt, 2015).

The most striking evidence on this point concerns the failure to learn from failure itself. One might expect that the experience of failure would puncture optimism and sharpen judgment. The evidence suggests otherwise. In an experimental study

of 205 entrepreneurs, Amore, Garofalo & Martin-Sanchez (2021) found that dispositional optimism was negatively associated with the extent to which entrepreneurs updated their beliefs in response to negative feedback, and that this reduced the effectiveness of their firms' innovation. The optimistic entrepreneur, confronted with evidence of error, is less likely to revise. Consistent with this, research on serial entrepreneurs has found that even those who have experienced business failure tend to remain optimistic about the prospects of their subsequent ventures (Ucbasaran, Westhead, Wright & Flores, 2010). Failure, which ought to be the great teacher, is frequently not learned from, because the cognitive apparatus that would process its lessons is precisely the apparatus that optimism and overconfidence have compromised.

2.6 The Affect Heuristic and Narrative Thinking

The biases discussed to this point are often presented as failures of cognition, as if the problem were purely a matter of how information is processed. But entrepreneurial judgment is not a purely cognitive affair. It is suffused with emotion — with the passion that drives venture creation, the excitement of possibility, the fear of loss, and the tender attachment that founders develop to their ideas. The affect heuristic describes the process by which these emotional states, rather than dispassionate analysis, come to guide judgment and decision. When a decision is made on the basis of how one feels about the options rather than on a systematic evaluation of their properties, the affect heuristic is at work.

In the entrepreneurial context, positive affect has been shown to have a specific and consequential effect on information processing. Baron, Hmieleski & Henry (2012) demonstrate that entrepreneurs experiencing positive affect tend to reduce their attention to, and processing of, negative information — and particularly information that contradicts their existing beliefs and expectations. The emotional

state that accompanies entrepreneurial enthusiasm thus reinforces the confirmation bias described above, operating through a different channel to produce a similar effect: the systematic exclusion of inconvenient evidence. The entrepreneur who feels good about a venture is, by that very fact, less likely to register signals that the venture is in trouble.

Closely connected to the affect heuristic is the role of narrative in entrepreneurial cognition. Entrepreneurs do not typically experience their ventures as probability distributions; they experience them as stories — narratives of struggle, vision, and eventual glorious triumphs. This narrative mode of thinking is a powerful tool for mobilising resources, persuading investors, and sustaining the commitment of teams. But it is also a source of distortion. A compelling narrative derives its persuasive force from its coherence, not from its correspondence to the base rates of reality. The more vivid and internally consistent the story of a venture's success, the more convincing it becomes — and the more it displaces the sober, statistical thinking that would situate the venture within the unglamorous distribution of actual outcomes. Narrative coherence, in this sense, is a substitute for probabilistic accuracy, and the substitution is rarely noticed by those who make it.

2.7 Is Bias Always Dysfunctional?

It would be tempting to conclude from the preceding sections that cognitive bias is simply a defect — a set of malfunctions that lead entrepreneurs into error and ventures into failure, and that the ideal entrepreneur would be one who had somehow been cured of them. This conclusion would be too hasty, and the literature itself resists it. The relationship between bias and entrepreneurial outcomes is genuinely contested, and the contest is instructive. It is here that the analysis must become most careful, because it is here that the phenomenon reveals its true complexity.

On one side of the debate stands a body of research suggesting that optimism and overconfidence are not merely tolerable but functional — even necessary — for entrepreneurial action. Optimism has been shown to trigger the innovation, persistence, and sustained effort that venture creation demands (Baron, 2008; Cardon, Wincent, Singh & Drnovsek, 2009). A degree of overconfidence has been argued to generate value in the early phases of a business, when the willingness to commit resources in the absence of confirming evidence is precisely what separates the entrepreneur who acts from the analyst who hesitates (Koellinger, Minniti & Schade, 2007). From this perspective, the biases documented above are not bugs in entrepreneurial cognition but features of it — the psychological preconditions for taking action under conditions where perfect rationality would counsel inaction.

On the other side stands an equally substantial body of research documenting the costs of these same dispositions. Hmieleski & Baron (2009), in a study whose findings are particularly pointed, demonstrated a negative relationship between entrepreneurs' optimism and the subsequent performance of their ventures, measured by revenue and employment growth. More striking still, they found that this negative relationship was stronger, not weaker, for experienced entrepreneurs, and stronger in dynamic environments than in stable ones. Optimism, in other words, was most damaging precisely where one might have expected experience to provide a corrective and precisely where the environment demanded adaptive responsiveness. This finding directly contradicts the intuition that seasoned entrepreneurs learn to temper their optimism into something more calibrated.

How are these opposing bodies of evidence to be reconciled? The most persuasive synthesis holds that the effects of cognitive bias are contingent rather than uniform — that whether a given bias helps or harms depends on the context in which it operates and the stage of the venture at which it is expressed. Optimism that fuels the initial decision to act, and that sustains effort through early adversity, may be genuinely functional in the founding phase. The same optimism, carried into a later

phase that demands the sober assessment of market feedback and the flexible revision of a failing strategy, may become actively destructive. Similarly, optimism may serve the entrepreneur well in a stable environment where persistence is rewarded, while damaging the entrepreneur in a dynamic environment where the capacity to update and adapt is paramount. The contradictory findings in the literature, on this reading, are not evidence of confusion but evidence of contingency: the studies disagree because the phenomenon itself is conditional (Hmieleski & Baron, 2009).

This contingency is precisely where the notion of bias must be handled with care, and it returns us to the qualification introduced in section 2.2. If the standard against which we measure entrepreneurial judgment is accurate prediction, then any systematic departure from accuracy is a bias, and bias is, by definition, a failing. But if, following Sarasvathy's account of effectuation, the entrepreneur under genuine uncertainty is not engaged in prediction at all — if the entrepreneurial task is to act upon available means, to control an unpredictable future rather than to forecast it — then the relationship between bias and rationality becomes far less straightforward. A disposition that produces inaccurate predictions may nonetheless produce effective action. The overconfident entrepreneur who enters a market that a rational calculation would have counselled avoiding may, through the very act of committed entry, help to create the conditions of their own success. Under this view, what the prediction-based standard condemns as bias, the action-based standard may vindicate as the necessary engine of enterprise. The point is not that one standard is correct and the other mistaken, but that the evaluation of entrepreneurial cognition cannot be separated from a prior choice about what entrepreneurial rationality consists of.

2.8 From Individual Bias to Systemic Effect

The analysis so far has operated primarily at the level of the individual entrepreneur and the individual venture. But the significance of entrepreneurial cognitive bias does not remain confined to that level. Entrepreneurs are not isolated decision-makers whose errors affect only themselves. They are actors whose decisions aggregate — into markets, into industries, into the broad direction of technological and economic development. To understand the full stakes of entrepreneurial bias, the analysis must scale up from the individual to the systemic. This move is essential to the argument of the thesis, because it is at the systemic level that the entrepreneurial half of the governance gap takes shape.

Consider first what happens when optimistic bias is not the property of a single entrepreneur but the shared disposition of an entire population of them. The experimental finding of Camerer & Lovallo (1999) already points in this direction: because markets are entered by a self-selected group of individuals who disproportionately believe in their own abilities, the aggregate rate of market entry systematically exceeds what the objective prospects would justify. The result is excess entry — more ventures competing for a given opportunity than the opportunity can sustain, and correspondingly high rates of failure. The individual-level bias, aggregated across a population, produces a market-level pattern of overinvestment and shakeout that no individual entrepreneur intended or foresaw.

This aggregation effect acquires particular importance in the context of new and rapidly developing technologies. When a technology captures the collective imagination of an entrepreneurial ecosystem — when it becomes the object of a shared conviction that it will transform the world — the biases documented in this chapter operate not in isolation but in concert, and at scale. Capital flows toward the sector on the strength of narratives whose coherence outruns their probabilistic grounding. Optimistic forecasts, insulated from correction by confirmation bias

and positive affect, drive the pace of development and deployment. Competitive pressure, amplified by the excess entry that overconfidence produces, accelerates the diffusion of the technology beyond the pace at which its consequences can be assessed. The speculative booms that have recurred throughout the history of technological innovation — from railway manias to the dot-com bubble — can be read, in significant part, as the systemic expression of individual-level cognitive biases operating in aggregate.

It is this systemic acceleration that connects the argument of the present chapter to the larger thesis. The entrepreneur, driven by optimism and overconfidence, insulated from corrective feedback, and mobilised by compelling narratives, does not merely risk personal failure. Acting in concert with a population of similarly disposed individuals, the entrepreneur becomes an engine of technological acceleration — a force that pushes new capabilities into the world faster than the surrounding society, and in particular its institutions of governance, can evaluate and absorb them. The optimistic bias that this chapter has examined at the level of the individual venture thus reappears, at the systemic level, as one of the two forces that generate the gap between innovation and its governance. The other force, located in the cognition of those charged with governing, is the subject of the chapter that follows.

Chapter 3 — Cognitive Bias in Regulatory and Legislative

Decision-Making

3.1 The Regulatory Mind Under Complexity

The previous chapter examined one half of the gap between innovation and its governance: the entrepreneur, driven by optimism and overconfidence, who accelerates the creation and diffusion of new technologies beyond the pace at which their consequences can be assessed. This chapter turns to the other half. If entrepreneurial cognition explains why technology arrives so quickly, regulatory cognition explains why governance arrives so slowly. The two are not symmetrical opposites — the legislator is not simply a pessimist to the entrepreneur's optimist — but they are complementary, and the gap between innovation and governance is the product of their interaction.

It has long been observed that law lags behind technology. This lag is frequently attributed to structural causes, and with good reason. Legislative processes are deliberately slow, built around consultation, committee review, and successive readings designed to prevent hasty lawmaking. Regulatory authority is fragmented across agencies and jurisdictions, each with partial competence over technologies that respect no such boundaries. The burden of proof typically falls on those who would restrict an activity rather than on those who would pursue it, so that regulation tends to follow demonstrated harm rather than anticipated risk. These structural features are real, and any honest account of regulatory lag must acknowledge them. But they are not the whole story, and the argument of this chapter is that they are not even the most interesting part of it.

Beneath the structural explanation lies a cognitive one. Legislators and regulators are human beings, and they form judgments about the technologies they govern through the same cognitive apparatus that governs all human judgment under

uncertainty. The conditions of regulatory work — technical complexity that exceeds the expertise of most lawmakers, political pressure that rewards visible responsiveness over quiet foresight, and the sheer novelty of the phenomena to be governed — are precisely the conditions under which cognitive bias flourishes. Where the entrepreneur's biases push toward action, the regulator's biases push toward a particular kind of inaction: not the considered restraint of a cautious mind, but the systematic misperception of what is happening, how urgent it is, and what conceptual tools are adequate to address it.

This chapter proceeds in several movements. It begins by establishing the structural frame within which regulatory cognition operates — the pacing problem and the Collingridge dilemma — because these define the terrain on which cognitive bias does its work. It then examines four specific biases that shape regulatory judgment: status quo bias, the illusion of understanding, the availability heuristic together with its associated recency effects, and anchoring. It applies this analysis to the governance of artificial intelligence, the case in which the cognitive dimension of regulatory lag becomes most visible and most consequential, before confronting the objection that technical expertise might dissolve the whole problem. Throughout, the aim is not to excuse regulatory failure as an inevitable product of human cognition, but to understand it precisely enough that it might be mitigated — a question to which the chapter returns at its close.

3.2 The Pacing Problem and the Collingridge Dilemma

Before turning to the cognitive biases that distort regulatory judgment, it is necessary to establish the structural condition within which those biases operate. That condition has been given a name: the pacing problem. In its simplest formulation, offered by Downes (2009), the pacing problem observes that

technology changes exponentially while social, economic, and legal systems change incrementally. The two operate on fundamentally different timescales.

Computing power advances at a rate that compounds; legislation advances at the rate of deliberation, committee hearings, and electoral cycles. The gap between these two rates is not a temporary misalignment that better management could close. It is a structural feature of the relationship between technological development and institutional response, and it widens as the pace of technology accelerates.

The pacing problem is compounded by a deeper dilemma, first articulated by David Collingridge in 1980 and since known as the Collingridge dilemma, or the dilemma of control. Collingridge observed that the attempt to govern a new technology confronts a double bind. In the early stages of a technology's development, when it would be easiest to shape and control, its eventual impacts and risks are largely unknown — there is no basis on which to construct sensible regulation, because the harms to be regulated have not yet revealed themselves. By the time those impacts and risks have become clear, however, the technology has typically become so deeply embedded in economic and social life that controlling it has grown enormously costly and difficult. Regulation is thus perpetually caught between two unfavourable moments: too early, when it lacks the information to act wisely, and too late, when it lacks the power to act effectively.

The significance of the Collingridge dilemma for the present argument is considerable, because it appears, at first glance, to offer a purely structural explanation for regulatory lag — one that owes nothing to cognitive bias. If the dilemma is real, then even a perfectly rational regulator, entirely free of the biases this chapter will examine, would face the same trap. The information required to regulate well simply does not exist at the moment when regulation would be most effective. This is an important objection, and it must be confronted directly rather than evaded. It represents the strongest version of the claim that regulatory lag is

structural rather than cognitive, and the argument of this chapter cannot proceed without answering it.

The answer is that the Collingridge dilemma establishes the terrain but does not determine the outcome. It is true that a regulator cannot know in advance precisely how a technology will develop. But the response to irreducible uncertainty is itself a matter of judgment, and it is in that judgment that cognitive bias operates. Confronted with the dilemma, a regulator might respond by building adaptive, revisable frameworks designed to evolve as information accumulates; by investing in the foresight capacity to anticipate a range of possible developments; or by accepting a degree of provisional, experimental regulation that can be corrected over time. That regulators so rarely respond in these ways — that they instead cling to existing categories, react to the most recent visible crisis, and overestimate their grasp of what they are governing — is not dictated by the Collingridge dilemma. It is the contribution of cognitive bias, layered on top of the structural constraint. The dilemma explains why regulation is hard. It does not explain why regulation so consistently fails in the particular, patterned ways that it does. For that, one must look to cognition.

3.3 Status Quo Bias and Institutional Inertia

The first and perhaps most pervasive of the biases that shape regulatory judgment is the status quo bias: the systematic tendency to prefer the current state of affairs and to resist change, even when the alternatives on offer are demonstrably superior. The definitive demonstration of this bias came from Samuelson & Zeckhauser (1988), who established through a series of controlled experiments that individuals disproportionately favour the existing option simply because it is the existing option. Across a wide range of decision scenarios, and with a substantial body of participants, the mere designation of one alternative as the status quo increased the

likelihood that it would be chosen — independent of its actual merits. The bias was not confined to trivial choices; the authors found it substantial in consequential real-world decisions such as the selection of health plans and retirement programmes.

The sources of this bias are several, and each has a clear analogue in the regulatory context. Loss aversion leads decision-makers to weigh the potential losses from change more heavily than the potential gains, so that the risks of a new regulatory approach loom larger than the risks of maintaining an inadequate existing one. The costs of transition — the effort of drafting new frameworks, the political capital required to enact them, the uncertainty of untested arrangements — are concrete and immediate, while the benefits are diffuse and deferred. And there is a powerful psychological attachment to prior commitments: institutions, having invested in a particular way of doing things, are reluctant to concede that the investment should be abandoned. In combination, these forces produce a strong institutional gravitational pull toward doing nothing, or toward doing the minimum necessary to preserve the appearance of existing arrangements.

Applied to the governance of technology, status quo bias manifests as a persistent preference for fitting new phenomena into old categories rather than constructing new ones. When a novel technology emerges, the path of least resistance — and the path that status quo bias makes most attractive — is to ask which existing legal category it most resembles and to govern it accordingly. This is often less a reasoned judgment than a cognitive default: the existing framework is familiar, it is already in place, and adopting it requires no admission that the situation is genuinely unprecedented. The difficulty is that genuinely novel technologies frequently do not fit the existing categories, and the attempt to force them in produces regulation that is systematically misaligned with the reality it governs. The category, once selected, then becomes its own source of inertia, resistant to revision long after its inadequacy has become apparent.

It is worth pausing to note that status quo bias is not simply a defect to be lamented. As with entrepreneurial optimism in the previous chapter, there is a case to be made in its favour. A regulatory system that changed its fundamental categories in response to every new development would offer no stability, no predictability, and no basis for long-term planning by the actors it governs. Some degree of institutional conservatism is a precondition for the rule of law itself, which depends on rules being stable enough to be known and relied upon. The problem is not that regulatory institutions exhibit status quo bias, but that they exhibit it indiscriminately — applying the same conservative default to genuinely unprecedented situations that demand new thinking as to routine matters where stability is appropriate. The bias becomes pathological not through its existence but through its failure to distinguish the cases in which continuity serves from the cases in which it fails.

3.4 The Illusion of Understanding

If status quo bias explains why regulators cling to existing categories, the illusion of understanding explains why they do so with such confidence. First identified and named by Rozenblit & Keil (2002), the illusion of explanatory depth is the finding that people believe they understand complex phenomena with far greater precision, coherence, and depth than they actually do. In a series of studies, participants confidently rated their understanding of how everyday objects and processes worked — zippers, flush toilets, the formation of rainbows — only to discover, when asked to produce an actual step-by-step explanation, that their understanding was shallow and full of gaps. The confident feeling of comprehension, it turns out, is largely dissociated from the actual possession of comprehension.

Two features of this illusion make it particularly relevant to regulatory judgment. The first is that it is specific to explanatory knowledge — knowledge of how things work — rather than to knowledge of facts, procedures, or narratives. One does not tend to overestimate one's knowledge of a capital city or a historical date; the illusion attaches precisely to the kind of mechanistic, systemic understanding that governing a complex technology requires. The second is that the illusion is strongest, as Rozenblit and Keil found, where the environment supports the feeling of understanding with visible components and plausible surface descriptions. A technology that can be described in accessible terms — that presents a comprehensible face to the observer — invites precisely the overconfidence that its underlying complexity does not warrant.

The bridge from this cognitive finding to the domain of policy was built explicitly by Fernbach, Rogers, Fox & Sloman (2013), in a study whose title states its conclusion directly: political extremism is supported by an illusion of understanding. The researchers found that when people were asked to explain, in mechanistic detail, how a proposed policy would actually produce its intended effects, two things happened. Their stated understanding of the policy declined, as the illusion was punctured by the demand for a genuine explanation; and, strikingly, the extremity of their position on the policy moderated. The confidence that had sustained their strong position turned out to rest on an understanding they did not in fact possess. When the illusion collapsed, so did the certainty that had rested upon it. This finding is of direct relevance to the legislator, whose confidence in a regulatory framework may rest less on genuine comprehension of the technology than on the illusion of such comprehension.

A vivid illustration of the same mechanism in the specific context of technology policy comes from Fernbach and colleagues' later work on genetically modified foods, which found that the most extreme opponents of the technology tended to know the least about the underlying science while believing that they knew the most (Fernbach, Light, Scott, Inbar & Rozin, 2019). The inversion is instructive:

confidence and actual knowledge were not merely uncorrelated but negatively related. Those most certain of their judgment were those least equipped to make it. Transposed to the regulatory arena, this suggests a troubling possibility — that the confidence with which a legislative body classifies and governs a novel technology may bear little relationship to, and may even run counter to, its actual understanding of what that technology is and does.

The consequences for regulation are direct. A regulator in the grip of the illusion of understanding does not experience the uncertainty that would counsel humility, adaptive frameworks, or deference to technical expertise. On the contrary, the illusion produces a confident conviction that the technology has been correctly understood and appropriately classified, and this confidence forecloses precisely the reconsideration that novel technologies require. The regulator does not know what he does not know, and — more dangerously — does not know that he does not know it. The illusion of understanding thus reinforces status quo bias: not only is the existing category the path of least resistance, but the confident sense that it is the correct category removes the doubt that might otherwise prompt a search for something better.

3.5 The Availability Heuristic and the Regulation of the Last Crisis

The third bias central to regulatory cognition is the availability heuristic: the tendency to judge the likelihood or importance of an event by the ease with which instances of it come to mind. First documented by Tversky & Kahneman (1973), the availability heuristic is efficient in many everyday contexts — events that are genuinely common are often genuinely easy to recall — but it produces systematic distortion whenever the ease of recall diverges from actual frequency. Vivid, recent, emotionally charged, or heavily publicised events are disproportionately

available to memory, and are therefore judged disproportionately likely and important, regardless of their actual statistical significance.

The application of this heuristic to the domain of law and regulation was developed in what is perhaps the single most important source for this chapter: Kuran & Sunstein's (1999) analysis of availability cascades and risk regulation, published in the *Stanford Law Review*. Kuran and Sunstein describe an availability cascade as a self-reinforcing process of collective belief formation, in which an expressed perception gains plausibility through its rising availability in public discourse. A vivid event or claim, once introduced, is amplified through repetition; the more it is discussed, the more available it becomes; the more available it becomes, the more plausible and important it seems; and this heightened sense of importance generates further discussion, in a spiral that can carry a perception far beyond what the underlying evidence would support. Crucially, the authors identify the role of what they call availability entrepreneurs — activists and interested parties who deliberately manipulate public discourse to trigger cascades that advance their agendas.

The regulatory consequence of these cascades, as Kuran and Sunstein document, is a pattern of law-making driven not by systematic analysis of risk but by the salience of recent events. Their canonical example is the Alar scare of the late 1980s, in which media coverage of a chemical used on apples snowballed into a national panic and a disproportionate regulatory response, despite limited empirical evidence of meaningful harm. The mechanism operates directly on legislators, agencies, and courts, whose attention and priorities are shaped by the same availability dynamics that shape public perception. The result is what might be called the regulation of the last crisis: a regulatory apparatus that directs its energy toward whatever risk has most recently become *available* through a dramatic public event, rather than toward the risks that a dispassionate assessment would identify as most significant.

This dynamic interacts with the pacing problem in a particularly damaging way. Because regulatory attention is captured by vivid recent events, and because such events by definition concern harms that have already materialised, the availability heuristic systematically directs regulation toward the past rather than the future. Lawmakers regulate the crisis that just occurred, using the conceptual tools that the crisis made salient, while the next development — less vivid because it has not yet produced a dramatic failure — commands no comparable attention. In a domain of slow-moving technology, this reactive posture might be tolerable, since the next development would arrive slowly enough for a subsequent reaction. In a domain of exponential technological change, it is a recipe for perpetual lag: by the time a given development has become vivid enough to command regulatory attention, the technology has already moved on, and the regulatory response, when it finally arrives, addresses a situation that no longer exists.

3.6 Anchoring and the Limits of Analogy

The three biases examined so far — status quo bias, the illusion of understanding, and the availability heuristic — are joined by a fourth that deserves separate treatment, because it governs the specific mechanism by which regulators make sense of the unfamiliar. Anchoring is the tendency to rely disproportionately on the first piece of information encountered, or on an initially salient reference point, when forming a judgment. Subsequent reasoning adjusts away from the anchor, but insufficiently, so that the final judgment remains distorted in the direction of the starting point. The anchor does not need to be relevant to be effective: in the original demonstration, estimates of the proportion of African nations in the United Nations were shifted simply by exposing participants beforehand to a number generated by a visibly random wheel of fortune (Tversky & Kahneman, 1974)

What makes anchoring particularly significant for the regulation of technology is the robustness of the effect and its indifference to expertise. Anchoring persists when the anchor is wildly implausible, when participants are offered financial incentives for accuracy, and when they are explicitly warned about the bias and instructed to resist it. Most importantly for the present argument, it persists among experts operating within their own domains. A now-classic study exposed a group of experienced judges to a sentencing recommendation that they were explicitly told had been generated at random, and found that those exposed to a high figure imposed sentences roughly fifty per cent longer than those exposed to a low one — and that the more experienced judges, far from being immune, were merely more confident in the judgments the anchor had distorted (Englich, Mussweiler & Strack, 2006). This finding disposes of the comfortable assumption that professional training and experience protect against the bias. They do not; they simply make its victims more certain.

In the regulation of emerging technology, anchoring operates through analogy. Confronted with a genuinely novel phenomenon, the regulator does not reason from first principles about what the phenomenon is and what governance it requires. Instead, the regulator reaches for the nearest familiar category — the most available analogy — and anchors the entire subsequent analysis to it. The novel technology is understood, classified, and governed as though it were a variation on something already known. The question that organises regulatory debate becomes not what is this, and what does it require, but which existing thing it resembles to the most. Is a ride-hailing service a taxi company or a technology platform? Is a cryptocurrency token a security, a commodity, or a currency? Is a social media platform a publisher, a common carrier, or something without precedent? In each case, the regulatory contest reduces to a choice among inherited categories, and the category selected as the anchor determines the entire framework of obligations that follows — often before anyone has seriously asked whether any of the available categories fits at all.

The interaction between anchoring and the other biases is tight and mutually reinforcing. The availability heuristic determines which analogy first comes to mind, and thus which category becomes the anchor: the most salient precedent, not necessarily the most apposite one. Status quo bias then ensures that the anchoring category, once adopted, is retained against challenge. And the illusion of understanding supplies the confidence that the chosen analogy is not merely convenient but most importantly correct. Anchoring is, in this sense, the connective tissue of regulatory cognition — the mechanism through which a novel reality is assimilated to a familiar and often inadequate template, and through which the other biases translate into concrete regulatory outcomes. It is also the bias that most directly produces the misalignment at the heart of this thesis, for a framework anchored to the wrong analogy is misaligned not by degree but by kind, governing the technology as though it were something it fundamentally is not.

3.7 Artificial Intelligence as the Central Case

It is time to make it specific, and to do so by turning deliberately toward the most salient technology of the present moment: artificial intelligence. There is a certain irony in this choice, and it should be acknowledged rather than concealed. In selecting AI as the central case, this analysis knowingly submits to the very availability heuristic it has just finished criticising — it reaches for the example that is most vivid, most discussed, most present in public consciousness. But the availability heuristic misleads only when salience and significance diverge, when the vividness of an example outruns its actual importance. The wager of this chapter is that artificial intelligence represents the rarer case in which the two coincide: it is at once the most available technology and, plausibly, the most consequential. To choose it is therefore not to be captured by salience but to exploit a moment in which salience happens to track significance — while remaining alert, precisely because the bias has been named, to the possibility that it does not. Artificial intelligence represents the case in which the cognitive dimension of regulatory lag becomes most acute, for reasons that follow directly from the distinctive features

of the technology itself. AI develops at a pace that strains even the accelerated timescales of contemporary technology; its consequences are unusually opaque, often diffuse and probabilistic rather than concrete and localised; and it operates across virtually every sector of economic and social life simultaneously, defying the sector-by-sector organisation of existing regulatory systems. Each of these features amplifies the biases examined above.

The most ambitious attempt to date to construct a comprehensive governance framework for artificial intelligence is the European Union's Artificial Intelligence Act, Regulation (EU) 2024/1689, published in July 2024 and widely described as the world's first comprehensive horizontal legal framework for the technology. The Act adopts a risk-based approach, sorting AI systems into tiers — from minimal risk through limited and high risk to a category of unacceptable risk subject to outright prohibition — and calibrating regulatory obligations to the assessed level of risk. As an attempt to govern AI proactively rather than merely reactively, the Act is a serious and in many respects admirable undertaking, and it would be unfair to present it as a simple failure. It represents, on the contrary, a genuine effort to escape the reactive posture that the biases of this chapter tend to produce.

It is instructive to set the European approach beside that of the United States, because the contrast reveals that the same cognitive traps produce different symptoms under different institutional conditions. Where the European Union pursued a single comprehensive statute, the United States has, to date, produced no equivalent general legislation, governing artificial intelligence instead through a patchwork of existing sectoral authorities, agency guidance, and executive action of uncertain durability. The two approaches appear opposite — the one comprehensive and codified, the other fragmented and improvised — but they are shaped by the same underlying biases, merely expressed through different institutional structures. The European approach exhibits the illusion of understanding in its confident enumeration of a complete risk taxonomy; the American approach exhibits status quo bias and anchoring in its reflexive

assimilation of AI to the jurisdictions of existing agencies, each of which addresses the technology through the narrow aperture of its established mandate. Neither approach escapes the cognitive pattern. The comprehensive framework risks obsolescence the moment the technology outruns its enumeration; the fragmented framework risks incoherence, as a dozen agencies anchor the same technology to a dozen different and partial analogies. The comparison suggests that the cognitive dimension of regulatory lag is not an artefact of any particular legal tradition but a general feature of how institutions confront the unprecedented.

The pattern is clearer still in cases where regulation was not even attempted proactively. The paradigmatic example is Section 230 of the United States Communications Decency Act of 1996 — famously described as the twenty-six words that made the internet — which granted providers of interactive computer services immunity from liability for content created by their users. The provision was drafted for a nascent online environment of small forums and bulletin boards, in an era of considerable technological optimism. Nearly three decades later, that same provision, substantially unchanged, governs the trillion-dollar global platforms that now shape the informational environment of entire societies. The framework designed for the internet of 1996 was anchored so firmly in place that it persisted through the complete transformation of the phenomenon it governed, and it has proven extraordinarily resistant to revision despite a broad and bipartisan consensus that it no longer fits the reality it addresses. Section 230 is status quo bias and anchoring made law: a category selected in the technology's infancy, retained by institutional inertia long after the technology outgrew it.

A comparable dynamic can be observed in the regulation of financial derivatives prior to the crisis of 2008. Complex instruments were governed, to the extent that they were governed at all, through frameworks designed for simpler financial products, and the inadequacy of those frameworks became fully apparent only when the harm had already materialised on a systemic scale. Here the Collingridge dilemma reveals its full force: the risks were not clearly visible while the

instruments were novel and the frameworks might have been shaped to contain them, and by the time the risks were undeniable, the instruments were so deeply woven into the global financial system that control had become enormously costly. In each of these cases — AI, the internet, financial derivatives — the same cognitive pattern recurs, varying only in its particulars. Regulatory attention arrives late, captured by whatever crisis has become available; regulatory comprehension is overconfident, resting on an illusion of understanding; and the regulatory response is anchored in categories inherited from an earlier and simpler world.

3.8 Why Expertise Is Not a Solution

At this point an obvious objection presents itself, and it is sufficiently important that the argument cannot proceed without addressing it. If the problem is that legislators do not understand the technologies they govern — if the illusion of understanding, anchoring to inapt analogies, and the misperception of risk all stem from a deficit of genuine technical comprehension — then the remedy appears straightforward. Put experts in the room. Staff the regulatory agencies with engineers and scientists who actually understand the technology, and the cognitive distortions born of ignorance should dissolve. This objection is intuitive, widely held, and largely mistaken, and understanding why it fails deepens rather than undermines the argument of this chapter.

The first difficulty is that expertise does not confer immunity from cognitive bias. The evidence assembled throughout this chapter speaks directly to this point. The judges in the anchoring study were experts operating within their own domain, and they were not merely susceptible to the bias but more confident in the distorted judgments it produced (Englich, Mussweiler & Strack, 2006). The illusion of explanatory depth afflicts specialists as well as laypeople; indeed, expertise in one

aspect of a complex system can generate an unwarranted confidence about the system as a whole. Experts are, in many domains, superior decision-makers, but the literature is clear that they remain subject to schematic thinking, overconfidence, and the systematic errors that heuristics produce. The expert does not escape the biases of human cognition; the expert experiences them with greater conviction, armoured by credentials against the doubt that might otherwise prompt correction. Replacing an ignorant legislator with a confident expert may change the character of the error without reducing its magnitude.

The second difficulty is structural rather than cognitive, and it concerns where technical expertise actually resides. In a rapidly developing technology, the deepest expertise is concentrated overwhelmingly in the private firms that are developing the technology commercially. The regulator who seeks genuine understanding must therefore turn, in large part, to the very industry the regulation is meant to constrain. This is the classic setup for what the economist George Stigler (1971) termed regulatory capture: the process by which a regulatory agency comes, over time, to serve the interests of the industry it is meant to oversee rather than the public it was created to protect. Stigler's insight was that capture need not involve corruption; it follows from the ordinary structure of incentives. The industry has a concentrated interest in regulatory outcomes and the resources to pursue it, while the diffuse public has neither. The mechanism is accelerated by the revolving door, through which regulators are recruited from the industries they oversee — because that is where the expertise lies — and return to those industries afterwards, often at far higher salaries.

The consequence is a genuine dilemma that the call for expertise cannot resolve. The knowledge required to regulate a complex technology competently is held disproportionately by the parties with the strongest interest in shaping the regulation to their advantage. To acquire the expertise, the regulator must draw closer to the industry; but the closer the regulator draws, the greater the risk that the resulting framework will reflect the industry's preferences rather than the public

interest. In the specific case of artificial intelligence, this dilemma is acute, because the information asymmetry between the leading developers and any public body is vast and widening. The expertise to understand the frontier of the technology exists almost entirely within a small number of private laboratories, which creates precisely the conditions in which regulatory capture flourishes. Expertise, then, is not a clean solution standing outside the problem. It is entangled in the problem, subject to the same cognitive biases and exposed to a structural capture that its very pursuit intensifies. The point is not that expertise is worthless — competent regulation plainly requires understanding — but that expertise is not the external corrective the objection imagines it to be. The cognitive and structural obstacles to good technology governance reach even into the remedy most commonly proposed for them.

3.9 From Individual Bias to Institutional Pattern

A further objection must be addressed before this chapter concludes. The biases examined here — status quo bias, the illusion of understanding, the availability heuristic, anchoring — were established, for the most part, through experiments on individuals. Regulation, however, is not the act of an individual but the product of institutions: legislatures, agencies, courts, composed of many people operating through elaborate procedures. It might be objected that individual cognitive biases would wash out at the institutional level, cancelled by the diversity of participants or corrected by the deliberative procedures that institutions employ. If this were so, the cognitive analysis of regulatory lag would collapse, and the structural account would suffice after all.

The objection has force, but the evidence does not support the optimistic conclusion it points toward. Institutions do not reliably cancel individual biases; frequently, they amplify them. The availability cascade described by Kuran and

Sunstein is precisely a mechanism by which an individual-level bias — the availability heuristic — is magnified rather than corrected through social and institutional transmission. A vivid perception does not average out as it passes through a legislature; it gathers force, as legislators respond to the same salient events, the same public pressure, and the same media coverage that shape one another's attention. Deliberative procedures, far from correcting the illusion of understanding, may entrench it, as the collective confidence of a committee exceeds the confidence any individual member would hold alone. And status quo bias is, if anything, stronger at the institutional level than the individual, because institutional change requires overcoming not one person's inertia but the combined inertia of many, mediated through procedures deliberately designed to make change difficult.

There is, moreover, a distinctively institutional source of bias that has no individual analogue. Legislators and regulators operate under incentives that reward the appearance of responsiveness to visible problems and penalise the quiet, unglamorous work of anticipating problems that have not yet materialised. The political economy of regulation thus aligns with and reinforces the availability heuristic: there is credit to be claimed for responding to a vivid crisis, and little to be gained from guarding against a risk that remains abstract. The incentive structure of democratic institutions, in this respect, does not counteract the cognitive biases of the individuals who inhabit them but runs in the same direction, converting individual availability bias into institutional priority. The cognitive and the structural, in other words, are not competing explanations but reinforcing ones: the structure of regulatory institutions is such that individual cognitive biases are transmitted and amplified rather than filtered and corrected.

This completes the account of the regulatory half of the gap. Where Chapter 2 showed the entrepreneur, driven by optimism, accelerating technology into the world faster than its consequences could be assessed, this chapter has shown the legislator, subject to a different but equally systematic set of biases, responding to

that acceleration slowly, confidently, and with tools inherited from an earlier age. The entrepreneur regulates nothing and races forward; the legislator races nothing and regulates backward, toward the crisis that has already passed. The gap between them is not a failure of individual competence on either side. It is the structural product of two cognitive systems operating under conditions that reliably distort the judgment of each. How these two halves combine — how entrepreneurial acceleration and regulatory lag interact to produce a gap greater than the sum of its parts — is the subject of the chapter that follows, in which the two threads of the argument are drawn together into a single framework.

Chapter 4 — The Cognitive Gap Model

4.1 Toward an Integrated Framework

The two preceding chapters have each told half of a story. Chapter 2 examined the entrepreneur, whose optimism and overconfidence, insulated from correction and amplified across a population, drive the creation and diffusion of new technologies at a pace that outruns society's capacity to assess them. Chapter 3 examined the legislator, whose status quo bias, illusion of understanding, availability-driven attention, and reflexive anchoring to inherited analogies produce a governance response that arrives late, rests on overconfidence, and reaches for the wrong tools. Each analysis stands on its own. But to leave them standing separately would be to miss the entire point that gives this thesis its meaning. The gap between innovation and its governance is not the sum of two independent failures. It is the product of their interaction, and that interaction has a structure that neither chapter, taken alone, can reveal.

This chapter proposes a framework for understanding that structure. It is called, for the reasons that will become clear, the Cognitive Gap Model. Its purpose is to show how entrepreneurial and regulatory cognition combine — not merely coexist, but actively reinforce one another — to produce a gap that is wider, more persistent, and more resistant to conventional remedy than either cognitive failure would produce in isolation.

It is worth stating clearly, at the outset, what kind of contribution this is and what conversation it enters. The governance of emerging technology is not an unexamined field. On the contrary, it has produced a sophisticated body of scholarship, at the centre of which stands the framework of responsible innovation

articulated by Stilgoe, Owen & Macnaghten (2013). That framework identifies four dimensions along which the governance of innovation should operate — anticipation, reflexivity, inclusive deliberation, and responsiveness — and it has become the leading normative account of how societies ought to steward the technologies they create. Alongside it sits the related concept of anticipatory governance, developed by Guston (2014), which seeks to build the capacities of foresight, engagement, and integration as a response to the pathologies of innovation that are, in Guston's phrase, conventionally realised only with hindsight.

These frameworks are valuable, and this thesis does not seek to displace them. But they share a striking and consequential silence. For all their sophistication about what responsible governance should look like, they say remarkably little about why responsible governance is so rarely achieved — about the specific mechanisms that cause societies to fail, again and again, to anticipate, to reflect, to deliberate, and to respond in time. They describe the destination without adequately mapping the terrain that makes it so hard to reach. It is precisely this terrain that the preceding chapters have charted, and that the present chapter now proposes to formalise. The Cognitive Gap Model does not compete with the responsible innovation framework; it supplies the diagnostic account of cognitive failure that the framework's prescriptions presuppose but do not themselves provide. If responsible innovation tells us where to go, the model developed here explains why we so consistently fail to go there — and locates the explanation not in the absence of good intentions or adequate resources, but in the systematic architecture of human judgment.

There is, moreover, a point of direct contact between the existing literature and the argument of this thesis that is worth drawing out, because it offers independent corroboration of the analysis. Within the responsible innovation tradition, it can be seen that definitions in technology governance are typically made, at least initially, by means of analogy — analogies that serve to normalise a new technology by

assimilating it to something already familiar (von Schomberg). This is offered, in that literature, as a more or less neutral description of how governance proceeds. The analysis of Chapter 3 allows it to be seen for what it is: not a neutral method but the operation of *anchoring*, a documented cognitive bias, at the level of institutional practice. Where the existing literature notes the reliance on analogy as a fact about governance, the Cognitive Gap Model identifies it as a bias with predictable and systematic consequences. This is characteristic of what the model contributes: it takes phenomena that the governance literature has described and reframes them in cognitive terms, thereby explaining not merely that they occur but why they occur and why they prove so difficult to correct.

4.2 The Structure of the Gap

The Cognitive Gap Model rests on a simple observation with complex consequences. Innovation and governance operate as two coupled systems, each responding to the other, and each driven by the cognition of the human actors who compose it. On one side stands the entrepreneurial system: the population of founders, firms, and investors who create and deploy new technologies. On the other stands the regulatory system: the legislatures, agencies, and courts that govern them. Between the two runs a continuous exchange of signals. The entrepreneurial system produces technologies, which generate consequences, which the regulatory system perceives and responds to, and those responses in turn shape the conditions under which the entrepreneurial system continues to operate. This much is uncontroversial, and it is the common ground of the coevolutionary account of technology and institutions.

What the model adds is the recognition that each side of this exchange is governed by a characteristic cognitive bias, and that the two biases point in opposite directions. The entrepreneurial system is biased toward acceleration. As Chapter 2 established, optimism and overconfidence lead entrepreneurs to underestimate risk, overestimate their prospects, and act with a confidence the evidence does not

warrant — and when these dispositions are aggregated across a competitive population, the result is a systematic tendency to push technology into the world faster than a sober assessment would justify. The excess entry demonstrated by Camerer & Lovo (1999) is the signature of this acceleration: more ventures, more investment, more rapid deployment than the objective opportunity supports. The entrepreneurial system, driven by its characteristic biases, runs fast and forward, unstoppable.

The regulatory system, by contrast, is biased toward lag. As Chapter 3 established, status quo bias, the illusion of understanding, the availability heuristic, and anchoring combine to produce a response that is slow to arrive, misdirected toward the wrong problem, overconfident in its grasp of the technology, and anchored to obsolete categories. The regulatory system, driven by its characteristic biases, runs slow and backward — toward the crisis that has already occurred, governed through the analogy that first came to mind. The gap between innovation and governance is, in the first instance, simply the distance between a system biased toward acceleration and a system biased toward lag. One races ahead; the other trails behind; and the space that opens between them is the gap the thesis is named for. The overall structure is represented in Figure 4.1.

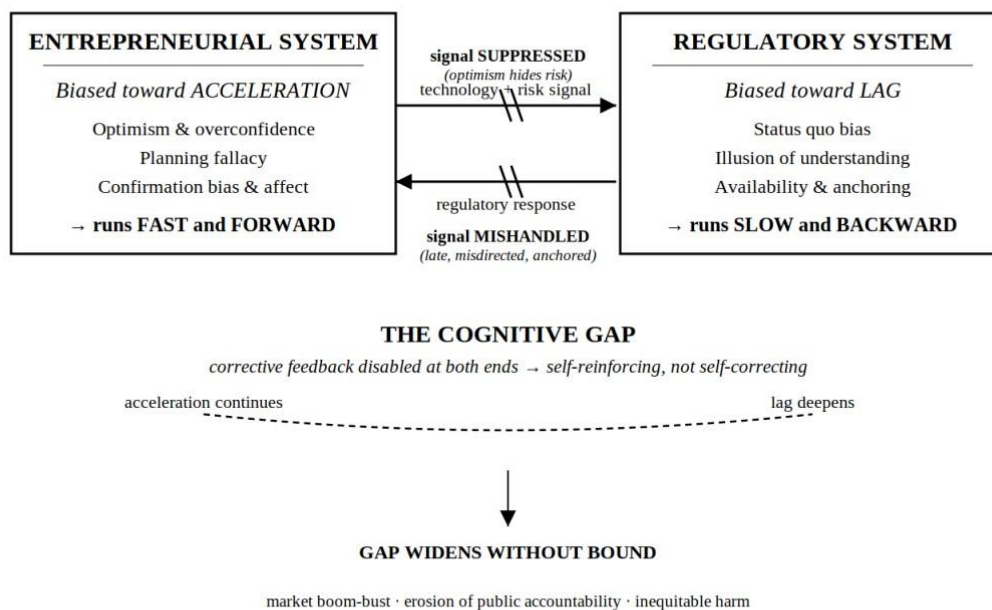


Figure 4.1 The Cognitive Gap Model. Two coupled systems, each driven by a characteristic and opposing cognitive bias, are joined by a feedback relationship that is disabled at both ends: the risk signal is suppressed on the entrepreneurial side before it can be sent, and mishandled on the regulatory side when it is received. The corrective loop that would keep the systems aligned is broken, and the gap between them widens without bound. Source: author's elaboration.

But this initial description, though accurate, is incomplete, because it treats the two biases as independent. It suggests a picture in which the entrepreneurial system would race ahead at its own pace regardless of what the regulator did, and the regulatory system would trail at its own pace regardless of what the entrepreneur did, so that the gap between them would be merely the arithmetic difference between two independently determined speeds. If this were the whole picture, the gap would be a serious problem, but a bounded one — the sum of two separate failures, each of which might be addressed on its own terms. The central claim of the Cognitive Gap Model is that this picture is wrong. The two biases are not independent. They interact. And their interaction transforms the gap from a bounded sum into an unbounded, self-reinforcing dynamic.

4.3 The Interaction Effect

To see why the two sides of the gap interact rather than merely coexist, it is necessary to consider the feedback relationship between them. Coupled systems that exchange signals do not simply run in parallel; they respond to one another, and the character of that response determines whether the coupling is stabilising or destabilising. A well-designed feedback relationship is self-correcting: when the two systems drift apart, the signals that pass between them generate responses that pull them back together. A poorly-designed feedback relationship is self-reinforcing: when the systems drift apart, the signals generate responses that drive them further apart still. The decisive question for the Cognitive Gap Model is which of these two characters the innovation–governance relationship possesses.

The answer the model gives is that cognitive bias converts what could be a self-correcting relationship into a self-reinforcing one.

Consider first how the relationship would behave in the absence of bias. An entrepreneurial system racing ahead would generate risks; a perceptive regulatory system would detect those risks and respond; the response would impose friction on the entrepreneurial system, slowing its acceleration to a pace the governance system could match; and the gap would stabilise. The risk signals generated by acceleration would function as the corrective feedback that keeps the two systems aligned. This is, in idealised form, what the responsible innovation framework envisions: a governance system with the capacity to anticipate and respond, coupled to an innovation system in a way that keeps the two in step. The corrective mechanism exists, in principle. The question is why it so rarely functions.

The Cognitive Gap Model's answer is that the biases of the two systems systematically disable the corrective mechanism at every point where it might operate. Begin on the entrepreneurial side. The corrective loop depends on risk

signals being generated and transmitted — on the entrepreneurial system registering and communicating the dangers its own acceleration produces. But the biases established in Chapter 2 attack precisely this function. Optimism leads entrepreneurs to underestimate the risks in the first place; confirmation bias and the affect heuristic lead them to filter out the evidence of those risks as it arrives; and the narrative mode of entrepreneurial cognition translates uncertain and dangerous ventures into compelling stories of assured success. The risk signal that the corrective loop requires is thus suppressed at its source. The entrepreneurial system does not merely fail to slow itself; it fails even to generate the accurate signal of danger that would prompt the regulator to slow it.

Now consider the regulatory side. Even the attenuated risk signal that does escape the entrepreneurial system's optimism must be received, interpreted, and acted upon by the regulatory system — and here the biases of Chapter 3 disable the loop a second time. The availability heuristic ensures that the regulatory system attends not to the most significant risks but to the most vivid and recent ones, so that the signal it acts upon is systematically the wrong signal. The illusion of understanding ensures that it interprets that signal with an overconfidence that forecloses reconsideration. Status quo bias and anchoring ensure that its response, when it comes, is assimilated to an existing category rather than tailored to the actual novelty of the risk. And the institutional dynamics examined at the close of Chapter 3 — the availability cascade, the political economy that rewards responding to visible crises — ensure that these individual biases are amplified rather than corrected as they pass through the machinery of governance. The signal is received late, misread, and mishandled.

The consequence is a feedback relationship in which the corrective mechanism is disabled twice over: once on the entrepreneurial side, where the risk signal is suppressed before it can be sent, and once on the regulatory side, where the diminished signal that survives is misperceived and mishandled on receipt. A loop that is broken at both ends cannot self-correct. And a coupled system whose

corrective mechanism is disabled does not merely fail to close the gap between its components; it allows that gap to widen without bound, because nothing in the system now operates to restrain the divergence. This is the core mechanism of the Cognitive Gap Model, and it is the reason the gap is properly described as multiplicative rather than additive. The entrepreneurial bias does not simply add its acceleration to the regulatory bias's lag. It disables the very mechanism by which the regulatory system might have detected and restrained that acceleration — so that each bias amplifies the effect of the other, and the gap they jointly produce exceeds, often greatly, the sum of the gaps that either would produce alone.

A further turn of the same dynamic deepens the effect. Because the regulatory system, when it finally does respond, so often responds badly — late, misdirected, anchored to an inapt category — its interventions frequently impose costs without delivering the intended control. This visible record of clumsy, ineffective regulation then feeds back into the entrepreneurial system as a rhetorical resource: it becomes evidence, deployed by those with an interest in acceleration, that regulation as such is incompetent and should be resisted or pre-empted. The regulator's cognitive failures thus supply the entrepreneurial system with a justification for evading governance altogether, further weakening the corrective loop. Each side's bias, in this way, does not merely coexist with the other's but actively nourishes it. The optimism that accelerates technology generates the ungoverned risks that eventually force a clumsy regulatory response; the clumsy regulatory response discredits governance and licenses further acceleration. The gap is not a static distance but a widening spiral, and it is driven by the interaction of cognitive biases that, individually documented in the preceding chapters, here combine into a single self-reinforcing system.

4.4 The Model at Work: The Case of Social Media

A model is only as valuable as its capacity to illuminate real cases. To demonstrate that the Cognitive Gap Model is an analytical instrument rather than a mere abstraction, it is worth tracing its operation through a single technology whose trajectory is now sufficiently complete to be examined in full. Social media offers an almost ideal case, both because its two-sided cognitive dynamic is unusually well documented and because its regulatory dimension connects directly to the Section 230 analysis of Chapter 3. Running the model through this case shows each of its components operating in sequence, and shows the disabled feedback loop producing precisely the widening gap the model predicts.

Begin with the entrepreneurial side. The founders of the major social media platforms did not, on the available evidence, set out to damage democratic discourse or adolescent mental health. They operated under a powerful and genuinely held optimistic narrative: that connecting the world was an unambiguous good, that giving everyone a voice would democratise information and empower the marginalised, and that whatever problems arose could be solved by the same technical ingenuity that had built the platforms in the first place. This is entrepreneurial optimism in its purest form — a compelling story of assured benefit that structured perception and filtered evidence. As the platforms scaled, warning signs accumulated: evidence of addictive design, of the amplification of outrage and misinformation, of harm to the well-being of young users. But these signals were precisely the kind that entrepreneurial cognition is disposed to suppress. They contradicted the founding narrative; they arrived as inconvenient interruptions to a story of world-historical benefit; and they were filtered, minimised, or reframed as problems of implementation rather than of the fundamental model. The risk signal, in the model's terms, was suppressed at its source. The platforms did not merely fail to slow their own expansion in response to mounting evidence of harm; the optimism that drove their creation impaired their capacity to register that evidence as the serious warning it was.

Turn now to the regulatory side, and the disabling of the loop repeats. For years, the dominant regulatory posture toward social media was governed by the anchoring analysis of Chapter 3: the platforms were assimilated to the category established by Section 230, understood as neutral conduits of third-party content rather than as active shapers of the information environment. This anchoring category, selected in the internet's infancy, persisted through the complete transformation of the phenomenon it governed, held in place by status quo bias and by the illusion that the nature of the technology was adequately understood. When regulatory attention did finally arrive, it arrived in the manner the availability heuristic predicts: in response to vivid, dramatic, highly publicised crises — a particular electoral controversy, a particular whistleblower's revelations, a particular tragedy — rather than in response to the slow, structural risks that had been accumulating for years. The regulatory response attended to whichever harm had most recently become available, addressed it through the aperture of inherited categories, and did so with a confidence in its understanding that the complexity of the technology did not warrant. Late, misdirected, anchored: the signal that survived the entrepreneurial system's optimism was mishandled on receipt exactly as the model describes.

The interaction effect completes the picture. The platforms' accelerated growth, unrestrained by an effective governance response, established facts on the ground — network effects, entrenched business models, a global user base — that made subsequent regulation progressively more costly and difficult, in a textbook realisation of the Collingridge dilemma. Meanwhile, the eventual regulatory responses, when they came, were frequently clumsy and technically ill-informed, and this clumsiness became itself an argument against regulation, deployed to suggest that legislators simply did not understand the technology well enough to govern it. The optimism accelerated the technology; the acceleration entrenched it; the entrenchment and the cognitive limitations of regulators produced a late and clumsy response; and the clumsiness of that response was turned into a justification

for resisting governance further. The gap did not stabilise. It widened, across more than two decades, in precisely the self-reinforcing spiral the model predicts. The case does not prove the model — a single case cannot — but it demonstrates that the model's components correspond to identifiable features of a real technological trajectory, and that its central prediction, the widening rather than the closing of the gap, is borne out.

4.5 The Model in Relation to Adjacent Theories

The Cognitive Gap Model does not arise in a theoretical vacuum, and its distinctiveness is best established by situating it precisely against the three bodies of theory to which it is most closely related: the Collingridge dilemma, the theory of regulatory capture, and the coevolutionary account of technology and institutions. In each case the model shares something with the existing theory, and in each case it departs from it in a way that constitutes its specific contribution. Drawing these relationships explicitly is necessary both to clarify what the model claims and to defend it against the charge that it merely restates what is already known.

Consider first the Collingridge dilemma, examined at length in Chapter 3. Collingridge identified a genuine informational trap: that a technology is easy to control before its risks are understood, and its risks are understood only after it has become hard to control. The Cognitive Gap Model shares Collingridge's concern with the temporal mismatch between technology and governance, and it incorporates his dilemma as a real structural constraint. But it departs from Collingridge in a decisive respect. The Collingridge dilemma is a theory about information — about what can be known, and when. It applies even to perfectly rational actors, and on its own it suggests a certain fatalism, since the informational trap appears inescapable. The Cognitive Gap Model locates the problem not only

in information but in cognition — in how actors process the information they do have. This shift has a crucial consequence: where Collingridge's informational trap is largely immovable, the cognitive dimension the model adds is, at least in principle, addressable. One cannot dissolve the informational asymmetry Collingridge identified, but one can design against the cognitive biases the model identifies. The model thus transforms Collingridge's fatalism into something more tractable, by locating a part of the problem in a place where intervention is possible.

Consider next the theory of regulatory capture, introduced in Chapter 3 through the work of Stigler (1971). Capture theory explains regulatory failure as the product of misaligned incentives: regulators come to serve the concentrated interests of the industries they oversee rather than the diffuse interests of the public, because the former are better organised and resourced to influence them. The Cognitive Gap Model shares capture theory's interest in why regulation systematically fails to serve the public interest, and it treats capture as one of the mechanisms by which the regulatory side of the gap is sustained. But it differs from capture theory in its fundamental explanatory logic. Capture theory is a theory of interest and incentive; it assumes rational actors responding to the structure of rewards they face. The Cognitive Gap Model is a theory of cognition; it explains failures that occur even where incentives are properly aligned and actors are sincerely attempting to serve the public. A captured regulator serves the wrong master; a cognitively biased regulator, even one sincerely serving the right master, misperceives the situation, attends to the wrong risks, and reaches for the wrong tools. The two theories are complementary rather than competing — capture and cognition can operate simultaneously, and indeed reinforce one another — but the model captures a category of failure that interest-based theories, by their nature, cannot: the failure of the honest, uncaptured regulator who is nonetheless systematically wrong.

Consider finally the coevolutionary account of technology and institutions, on which the model's very structure is built. That account, well established across science and technology studies and institutional economics, holds that

technologies and the institutions that govern them develop together, each shaping the other through continuous mutual adjustment. The Cognitive Gap Model adopts this coevolutionary structure wholesale; its two coupled systems, exchanging signals and adjusting to one another, are a coevolutionary picture. Where the model departs from the standard coevolutionary account is in its analysis of why the coevolution so often goes wrong. The coevolutionary literature describes the mutual adjustment of technology and institutions, but it tends to treat that adjustment as a broadly functional process, capable of reaching workable accommodations over time. The Cognitive Gap Model identifies the specific mechanism — the cognitively disabled feedback loop — by which the coevolutionary process fails to self-correct, and by which mutual adjustment gives way to mutual divergence. It is, in this sense, a theory of pathological coevolution: an account of the specific conditions under which the ordinarily adaptive coupling of technology and institutions becomes maladaptive. The contribution is to explain not that technology and governance coevolve, which is well known, but why their coevolution so reliably produces a widening gap rather than a stable accommodation.

Across all three comparisons, a consistent pattern emerges in what the model contributes. It takes an existing theory that describes a real phenomenon — an informational trap, a structure of captured incentives, a coevolutionary coupling — and it adds the cognitive dimension that the existing theory omits. In each case, the addition does explanatory work the original could not: it renders Collingridge's trap partly addressable, it captures the failures of the honest regulator that capture theory misses, and it explains the pathological coevolution that the coevolutionary account leaves unexplained. The model is not a replacement for these theories but an integration and extension of them, unified by the recognition that runs through the entire thesis — that the gap between innovation and its governance cannot be fully understood without attending to the systematic architecture of the human judgment that produces it.

4.6 The Gap Is Cognitive, Not Just Structural

The Cognitive Gap Model makes a strong claim, and intellectual honesty requires that it be tested against its most serious competitor. That competitor is the structural account of regulatory lag, which holds that the gap between innovation and governance is adequately explained by structural factors — the deliberate slowness of legislative procedure, the fragmentation of regulatory authority, the resource asymmetry between industry and state, the Collingridge dilemma's genuine informational trap — without any need to invoke cognitive bias at all. If the structural account is sufficient, then the cognitive model is superfluous: an elaborate way of redescribing a phenomenon that simpler and more familiar factors already explain. This objection must be met directly, for it is the strongest case against the thesis as a whole.

The response begins by conceding what should be conceded. The structural factors are real. This thesis has affirmed their reality at every stage, and it does not claim that cognitive bias is the sole cause of the gap. Legislative procedure is genuinely slow; regulatory authority is genuinely fragmented; the informational trap that Collingridge identified is genuine. A complete account of the gap must include these structural factors, and any analysis that ignored them would be incomplete. The question is not whether structural factors matter — plainly they do — but whether they are sufficient, or whether an adequate explanation requires the cognitive dimension as well. The claim of the model is the latter: that the structural account, however necessary, is not sufficient, and that what it leaves unexplained is precisely what the cognitive account supplies.

The decisive consideration is this. If the gap were purely structural, then improvements in the structural conditions should reliably narrow it. Provide the regulatory system with more resources, more technical expertise, faster procedures, and better coordination, and a purely structural gap should shrink accordingly. Yet the evidence assembled in this thesis suggests that it does not —

or at least, that it does not shrink as the structural account predicts. The biases documented in Chapter 3 operate independently of resources and procedure. A well-funded agency staffed by capable people, working through efficient procedures, remains subject to the availability heuristic, the illusion of understanding, status quo bias, and anchoring. Indeed, Chapter 3 showed that the most obvious structural remedy — the injection of technical expertise — does not dissolve the cognitive problem but relocates it, since experts are themselves subject to bias and their recruitment opens the door to regulatory capture. The persistence of the gap even where structural conditions improve is the empirical signature of a cause that is not structural. That cause is cognitive.

The point can be sharpened by observing what the two accounts predict about the character of regulatory failure. The structural account predicts that regulation should fail slowly and neutrally — that it should simply arrive late, in proportion to the procedural delays and resource constraints it faces, without any systematic directional distortion. The cognitive account predicts something more specific: that regulation should fail not merely slowly but in patterned ways — attending to the wrong risks (availability), with unwarranted confidence (the illusion of understanding), through inapt inherited categories (anchoring and status quo bias). It is this second, more specific pattern that the evidence displays. The Section 230 case, the EU AI Act's vulnerability to its own enumeration, the recurring regulation of the last crisis — these are not the neutral slowness that structural delay alone would produce. They are directional distortions of exactly the kind the cognitive account predicts and the structural account does not. The signature of the failure reveals its cause. A gap produced by mere structural delay would look different from the gap we actually observe; the gap we observe bears the specific fingerprints of cognition.

This is the sense in which the gap is cognitive and not just structural. The structural factors set the stage — they establish that governance will be difficult, slow, and imperfect. But the particular way in which governance fails, the specific and

repeated pattern of its failures, and above all the persistence of those failures in the face of structural improvement, cannot be explained by structure alone. They require the cognitive dimension. The Cognitive Gap Model does not deny the structural account; it completes it, supplying the explanation for the residue that structure leaves unexplained. And that residue is not a minor remainder. It is the systematic, directional, self-reinforcing character of the gap — which is to say, it is the most important thing about it.

4.7 The Falsifiability of the Model

A theoretical model that explained everything would explain nothing; a model earns its standing partly by specifying what would count as evidence against it. It is therefore necessary to ask directly what would falsify the Cognitive Gap Model — what observations, if they were made, would show it to be wrong. This question is uncomfortable for any theorist to pose about his own construction, but evading it would leave the model weaker, not stronger, and would forfeit precisely the intellectual seriousness that the model aspires to.

Several observations would count against the model. If regulatory responses to emerging technologies were found to be directed toward the most statistically significant risks rather than the most vivid and available ones, the availability mechanism at the heart of the model would be disconfirmed. If regulators confronting novel technologies characteristically expressed calibrated uncertainty about their understanding rather than the overconfidence the model predicts, the role of the illusion of understanding would be called into question. If gaps between innovation and governance reliably narrowed when structural resources were increased, holding cognition constant, the model's central claim that the gap is cognitive rather than merely structural would be undermined. And if the

entrepreneurial and regulatory biases were found to operate independently — if the presence of one had no amplifying effect on the consequences of the other — then the interaction effect that distinguishes the model from a simple sum of two separate failures would be refuted. Each of these is a genuine empirical possibility, and each would, if observed, damage or destroy the model.

That these disconfirming observations have not, on the evidence assembled here, been made is what gives the model its provisional standing. But the word provisional must be taken seriously. This thesis is a theoretical and conceptual work; it constructs the model and argues for its plausibility on the basis of existing evidence, but it does not itself conduct the systematic empirical testing that would move the model from plausible to established. That testing is possible — the model's predictions are specific enough to be operationalised, its mechanisms concrete enough to be measured — but it lies beyond the scope of the present work. What this chapter offers is a model that is coherent, that integrates the evidence of the preceding chapters, that improves upon the purely structural account, and that is falsifiable in principle. Whether it survives the testing it invites is a question that this thesis poses rather than answers, and to which the concluding chapter returns.

4.8 Systemic Consequences

If the Cognitive Gap Model is correct, its consequences extend well beyond the academic question of how to explain regulatory lag. A gap that is self-reinforcing rather than self-correcting, and that widens rather than stabilises as technology accelerates, has implications for markets, for democratic accountability, and for the distribution of the benefits and harms of technological change. These implications are worth drawing out, because they establish why the model matters and set the agenda for the concluding chapter's turn toward what might be done.

For markets, the model predicts recurrent cycles of overinvestment and correction. The excess entry that entrepreneurial optimism produces, unrestrained by a governance system whose corrective function has been disabled, generates booms in which capital floods toward technologies whose risks have been systematically underweighted, followed by corrections when those risks belatedly materialise. The absence of effective early governance does not prevent these cycles; it merely ensures that their costs are borne after the fact, socialised across the public rather than internalised by the actors whose optimism produced them. The history of technological bubbles, read through the model, is not a series of unrelated accidents but the repeated expression of a single structural dynamic.

For democratic accountability, the model predicts a progressive transfer of effective decision-making power from public institutions to private ones. Each time the regulatory system responds late, misdirected, and anchored to obsolete categories, it cedes ground: the technology becomes entrenched, the private actors who deployed it establish the facts on the ground, and the space for meaningful public governance narrows. Over time, the accumulation of these individual failures amounts to a systemic shift in where the decisions that shape society are actually made — away from the accountable institutions of the state and toward the unaccountable private actors who move faster than governance can follow. The gap, in this sense, is not merely an inconvenience but a slow erosion of the public's capacity to govern the technologies that increasingly govern it.

For the distribution of benefits and harms, the model predicts systematic inequity. The benefits of accelerated innovation accrue disproportionately to those who create and own the technologies; the harms, being diffuse, probabilistic, and deferred, fall disproportionately on those least able to anticipate or avoid them. A governance system functioning as intended would work to correct this distribution, internalising harms and protecting the vulnerable. A governance system whose corrective function has been cognitively disabled cannot perform this role, and the distributional consequences of technological change are left to fall where the

market and the biases of its most powerful actors direct them. The cognitive gap is thus not distributionally neutral. It systematically favours the fast over the slow, the technically sophisticated over the lay, and the concentrated interests of innovators over the diffuse interests of the public.

These consequences are sobering, and they might seem to counsel pessimism. But the logic of the model points in a more constructive direction, and it is with this that the chapter closes. If the gap is cognitive in origin — if it arises not from immutable structural necessity but from identifiable and specific patterns of human judgment — then it is, at least in principle, addressable. Cognitive biases, though they cannot be eliminated, can be anticipated, counteracted, and designed around. The very specificity of the model's account of failure is what makes it a foundation for improvement: a gap whose mechanisms are understood in detail is a gap against which targeted countermeasures can be devised. The concluding chapter takes up this possibility, asking what entrepreneurs and policymakers might do differently once the cognitive architecture of the gap is understood. The diagnosis developed across the preceding chapters, and formalised in the model presented here, is not an end in itself. It is the precondition for a cure.

Chapter 5 — Implications and Conclusions

5.1 From Diagnosis to Response

The preceding chapters have constructed an argument. Chapter 2 established that entrepreneurial judgment is systematically distorted by optimism, overconfidence, the planning fallacy, and the mechanisms of confirmation and affect that insulate those distortions from correction. Chapter 3 established that regulatory judgment is systematically distorted by status quo bias, the illusion of understanding, the availability heuristic, and anchoring to inherited analogies. Chapter 4 brought these two analyses together into the Cognitive Gap Model, arguing that the biases of the two systems interact to disable the feedback mechanism that would otherwise keep innovation and governance aligned, producing a gap that widens rather than stabilises. The argument, to this point, has been entirely diagnostic. This chapter asks what follows from it.

The move from diagnosis to response is not automatic, and it deserves justification. A diagnosis identifies the mechanism of a failure; it does not, by itself, establish that the failure can be remedied. Some conditions are well understood and nonetheless untreatable. The question this chapter must therefore answer first is whether the cognitive account of the gap is of the treatable or the untreatable kind — and the answer, importantly, is neither a confident yes nor a resigned no, but something more qualified that requires careful statement.

The case for treatability rests on a structural feature of the argument developed here. Recall the objection confronted in Chapter 3: that the Collingridge dilemma establishes an informational trap so fundamental that regulatory lag is inevitable regardless of how anyone thinks. If that objection had succeeded, this chapter could offer nothing but resignation, because an informational impossibility cannot be reasoned away. The response given there, and elaborated in Chapter 4, was that the Collingridge dilemma sets the terrain without determining the outcome — that the

response to irreducible uncertainty is a matter of judgment, and that judgment is where cognition operates. This response has a direct consequence for the present chapter. If part of the gap is cognitive rather than informational, then that part is, in principle, addressable, because cognition can be worked upon in ways that the fundamental unavailability of information cannot. The very move that saved the thesis from the structural objection is the move that makes this chapter possible.

But the case for treatability must be stated with precision, because the literature on debiasing counsels caution. It is one thing to say that cognitive failures are in principle addressable; it is quite another to say that we know how to address them reliably. This chapter proceeds under a discipline that the evidence imposes: it will describe interventions that have empirical support, it will state plainly where that support is thin, and it will resist the temptation — powerful in any concluding chapter — to convert a rigorous diagnosis into a confident prescription. A thesis that has spent four chapters documenting the human tendency toward unwarranted confidence would be poorly served by concluding with an unwarranted confidence of its own.

The chapter proceeds by addressing two audiences in turn. For entrepreneurs, it examines what the analysis of Chapter 2 implies for how ventures might be led, and confronts a genuine tension: if optimism is functionally necessary for entrepreneurial action, as section 2.7 argued it may be, then advising entrepreneurs to debias themselves may be advising them to disable the disposition that makes them entrepreneurs. For policymakers, it examines institutional mechanisms that appear, on inspection, to be almost purpose-built to counteract the specific biases documented in Chapter 3. The chapter then turns to the limitations of the analysis as a whole, before offering its conclusions.

5.2 Implications for Entrepreneurs

The entrepreneurial implications of this analysis begin with a difficulty rather than a recommendation. Chapter 2 concluded that the relationship between cognitive bias and entrepreneurial outcomes is genuinely contested — that optimism and overconfidence, while producing documented errors, may also be the psychological preconditions for action under conditions where sober calculation would counsel inaction. If that conclusion holds, then the obvious recommendation — that entrepreneurs should reduce their optimism and calibrate their confidence — is not obviously good advice. It may be advice to become a better forecaster at the cost of becoming a worse entrepreneur. Any honest treatment of the entrepreneurial implications must begin by taking this tension seriously rather than quietly setting it aside.

The resolution proposed here rests on a distinction between two things that the word optimism is used to describe. The first is a disposition: a general orientation toward possibility, a willingness to act under uncertainty, a resilience in the face of setback. This is what sustains a founder through the years of difficulty that venture creation involves, and there is little reason to think it should be suppressed — nor much prospect of suppressing it, since it is largely what selected the individual into entrepreneurship in the first place. The second is a set of specific judgments: this product will take six months to build; this market is worth this much; these risks are manageable; this negative signal is noise rather than evidence. These specific judgments are forecasts, and forecasts can be wrong in measurable ways with measurable consequences. The distinction matters because the interventions that follow target the second without requiring the suppression of the first. One can maintain the drive while calibrating the forecast. The recommendation is not that entrepreneurs should become pessimists, but that they should build processes that subject their specific predictions to a discipline their general disposition will not supply.

The most direct of these processes is the pre-mortem, a technique developed by Gary Klein (2007) as a form of prospective hindsight. Before a significant decision is executed, the team is asked to imagine that a year has passed and the project has failed comprehensively, and then to explain why. The technique's power lies in what it does to the cognitive framing of the question. Asking a committed team what might go wrong invites the optimistic filtering that Chapter 2 documented; the risks are acknowledged and then minimised, because the prevailing narrative of success remains intact. Asking the same team to explain a failure that is stipulated to have already occurred removes that filter, because the narrative of success has been suspended by the premise of the exercise. The failure is no longer a possibility to be argued against but a fact to be explained, and the mind that would have defended the plan is redirected toward diagnosing its collapse. The pre-mortem thus operates not by asking entrepreneurs to be less optimistic but by constructing a situation in which optimism has no purchase.

A second intervention addresses the planning fallacy specifically. Chapter 2 established, following Lovallo & Kahneman (2003), that the fallacy arises from the inside view — from constructing forecasts out of the particular details of the project at hand, which systematically excludes the historical record of how comparable projects have actually unfolded. The corrective, reference-class forecasting, requires the entrepreneur to identify a class of similar prior ventures, establish the actual distribution of their outcomes, and locate the present venture within that distribution before adjusting for its specific features. This is harder than it sounds, and not merely for technical reasons. The exercise requires the founder to accept that the venture belongs to a class at all — that it is an instance of a type rather than the singular exception its narrative presents it as. Entrepreneurial cognition resists this categorisation, because the sense of exceptionality is part of what motivates the enterprise. But it is precisely this resistance that makes the technique valuable: the discomfort of situating one's venture among its statistical peers is the sensation of a bias being interrupted.

A third intervention targets confirmation bias and the affective filtering examined in section 2.6. The general principle, established in the debiasing literature under the heading of consider the opposite, is that judgments improve when the judge is required to construct the case against their own conclusion rather than merely to remain open to it. Openness, as the evidence suggests, is insufficient: the biased mind experiences itself as open while filtering systematically. What is required is an active and structured obligation to generate disconfirming considerations. In organisational practice this takes the form of the devil's advocate — a person formally assigned to critique the assumptions underlying a plan (Schwenk, 1984) — or, in the entrepreneurial method most widely adopted, of structured falsification: the practice of stating assumptions explicitly as hypotheses and designing tests intended to disconfirm rather than to validate them. The logic is Popperian (Popper, 1959), and the discipline it imposes is real. A test designed to confirm will almost always confirm; a test designed to disconfirm can actually fail, and it is the possibility of failure that makes the exercise informative.

The practical importance of this last point is difficult to overstate, because the failure mode it addresses is the most common in entrepreneurship. Founders driven by conviction routinely substitute encouraging but uninformative signals — polite enthusiasm from friendly interlocutors, expressions of interest that never convert, engagement metrics disconnected from willingness to pay — for the harder evidence of genuine demand. The consequence is visible in the pattern of venture failure: a substantial proportion of failed ventures collapse not because the product could not be built or the funding could not be raised, but because no real market need existed for what was built. This is not a failure of execution or of funding. It is the accumulated cost of the optimism-and-confirmation complex that Chapter 2 analysed, operating over the life of a venture until the evidence becomes undeniable in the only form the mechanism cannot filter: the absence of revenue.

There is, finally, an implication that follows less from the individual biases than from the systemic argument of Chapter 4, and it is the implication most likely to

be resisted. If the model is correct, then the aggregate effect of entrepreneurial optimism is to suppress the risk signals on which effective governance depends. The entrepreneur who minimises the risks of a technology in public communication is not merely making an individual judgment; they are degrading the informational environment on which the regulatory system relies. This suggests a responsibility that the standard framing of entrepreneurship does not acknowledge — a responsibility not to be pessimistic, but to be accurate in the specific communications through which risk information reaches the public and its institutions. Entrepreneurs who take this responsibility seriously would communicate uncertainty as uncertainty rather than converting it into narrative confidence, and would treat the accurate transmission of risk information as part of what they owe the societies in which they operate. Whether competitive dynamics permit such behaviour is a serious question, and the Cognitive Gap Model gives grounds for pessimism about it: in a market where optimistic communication attracts capital and attention, the entrepreneur who communicates calibrated uncertainty is at a disadvantage. This is one reason the individual-level implications of the analysis, however valid, are insufficient on their own — and why the institutional implications examined next carry the greater weight.

5.3 Implications for Policymakers

The institutional implications of the analysis are, in one respect, more promising than the individual ones. Where the entrepreneur must work against biases operating inside a single mind, with limited tools and strong competitive pressures pulling the other way, the policymaker has access to a resource the individual lacks: the capacity to design institutions. Institutions can embed corrections that individuals cannot reliably sustain. A rule that forces reconsideration will produce reconsideration whether or not the officials subject to it happen to feel uncertain; a procedure that requires the generation of alternatives will produce alternatives

whether or not anyone is disposed to look for them. The institutional level is where cognitive bias is most effectively addressed, not because officials are more rational than entrepreneurs, but because institutions can be built to compensate for the irrationality of the people who staff them. Institutions are barriers.

This insight aligns with the most important qualification in the debiasing literature. The evidence, reviewed systematically, suggests that because many biases arise from deeply entrenched cognitive mechanisms, attempting to defeat them through effort and awareness alone is an uphill struggle demanding sustained motivation. The more promising approach is to modify the environment in which decisions are made rather than the cognitive capacities of the decision-makers — to construct contexts in which the biased judgment is intercepted, structured, or circumvented rather than contexts in which it must be personally overcome. For the design of governance institutions, this is a directly actionable principle. The question is not how to make legislators think better, which is largely beyond anyone's control. The question is how to build institutions whose procedures compensate for how legislators actually think.

Two mechanisms deserve particular attention, because each maps with unusual precision onto a specific bias documented in Chapter 3. The first is the regulatory sandbox: a controlled environment in which innovators may test new applications under regulatory supervision, typically with temporary exemptions from rules whose applicability to the novel technology is uncertain. Sandboxes have been implemented across financial services, energy, health, and mobility, and provision for them appears in the European Union's AI Act itself. Their conventional justification is that they permit innovation to proceed without premature regulatory obstruction. But their deeper significance, for the argument of this thesis, lies elsewhere.

The sandbox is best understood as an institutional bridge across the Collingridge dilemma. Recall the structure of that dilemma: a technology is controllable before

its risks are understood, and understood only after it has become difficult to control. The trap is informational — the knowledge required for good early regulation does not exist at the moment it is needed. The sandbox attacks this trap directly by *manufacturing* the missing knowledge under controlled conditions.

Controlled regulatory experimentation generates the knowledge that makes governance possible, without requiring governance to wait for knowledge that would otherwise emerge only after full deployment (Ranchordás, 2015). This is a genuine escape from what appeared to be an inescapable bind, and it is worth pausing on its significance for the thesis as a whole. The Collingridge dilemma was the strongest structural objection to the cognitive account; the sandbox demonstrates that even that structural constraint is partially addressable through institutional design. The informational trap is not a law of nature. It is a default that better-designed institutions can escape.

The second mechanism is the sunset clause: a statutory provision terminating a rule, programme, or agency on a fixed date unless it is affirmatively reauthorised. Sunset provisions spread across American state legislatures in the 1970s explicitly as a device against regulatory obsolescence and weak oversight, forcing periodic reconsideration of rules under updated information. Their relevance to the present analysis is precise, and it concerns status quo bias directly. Chapter 3 established that status quo bias operates by making inaction the path of least resistance: to change a rule requires effort, political capital, and the acknowledgement that the existing arrangement has failed, while to leave it in place requires nothing at all. The sunset clause reverses this default. Under a sunset provision, inaction terminates the rule rather than preserving it, and the burden of inertia is shifted from those who would change the arrangement to those who would maintain it. This is not an attempt to make legislators less prone to status quo bias; it is a structural arrangement that makes their status quo bias produce revision rather than stagnation. It is, in this sense, an exemplary instance of the environmental strategy the debiasing literature recommends.

The empirical record on sunset provisions is instructive and appropriately mixed. There is evidence that they lower the political costs of legislating under uncertainty: bills containing sunset provisions have been found to show a substantially higher probability of enactment, on the order of sixty per cent, presumably because the reversibility of a temporary measure makes it easier for legislators to support something whose consequences they cannot fully foresee. This is a significant finding for technology governance, where the reluctance to legislate under uncertainty is a major source of delay. At the same time, sunset provisions carry real costs. Temporariness weakens legal certainty, which is itself a value that regulated actors legitimately rely upon, and the systematic evidence on whether sunsets actually deliver the promised improvements in regulatory quality remains thin. This thesis therefore recommends attention to the mechanism rather than confident advocacy of it: the sunset clause is a well-designed response to a real cognitive problem whose practical effects require more evidence than currently exists.

Beyond these two mechanisms, the analysis of Chapter 3 suggests a third line of institutional response, addressed to the availability heuristic and the cascades it generates. If regulatory attention is systematically captured by whatever risk has most recently become vivid, then institutions that determine regulatory priorities require some insulation from the dynamics of public salience. Kuran and Sunstein, whose analysis of availability cascades supplied the foundation for that section, propose precisely this: a degree of insulation for civil servants from mass demands for regulatory change, coupled with the construction of accessible scientific databases that allow regulatory judgment to be anchored in systematic evidence rather than in the ease with which particular hazards come to mind. Such proposals sit uncomfortably with democratic intuitions, and the discomfort should be acknowledged rather than dismissed: insulating officials from public pressure is a serious step, and the argument that they will use their insulation wisely is exactly the argument that the illusion of understanding should make us cautious about.

There is a genuine dilemma here between responsiveness and calibration, and this thesis does not claim to resolve it. What the analysis does establish is that the dilemma is real and that it cannot be evaded by pretending that unmediated responsiveness to public salience produces good regulatory priorities. It does not; the evidence on availability cascades is clear on this point.

A final institutional implication concerns the timing of governance rather than its content. The Cognitive Gap Model predicts that regulatory attention will arrive late, after harms have materialised and become vivid, because that is when the availability heuristic delivers them to the attention of legislators. An institution designed against this pattern would need a dedicated capacity for anticipation — a function whose specific purpose is to attend to risks that are not yet salient, insulated from the pressure to address whatever crisis currently dominates public discussion. This is the core insight of the anticipatory governance literature discussed in Chapter 4: that foresight is a capacity to be institutionalised rather than an attitude to be adopted, and that its whole point is to counteract the tendency to recognise the pathologies of innovation only in hindsight. The Cognitive Gap Model supplies the specific explanation for why such a capacity is necessary — not because officials lack goodwill, but because the availability heuristic will otherwise ensure that attention flows toward the vivid and away from the significant.

5.4 The Third Actor: Capital and the Amplification of Optimism

The Cognitive Gap Model as presented in Chapter 4 identified two actors: the entrepreneur who accelerates and the legislator who lags. This binary structure has analytical advantages, and it captures the essential dynamic. But it leaves implicit a third participant whose role, on closer examination, is considerable — and whose inclusion strengthens rather than complicates the model. Between the entrepreneur

and the regulator stands the investor: the venture capitalist, the institutional fund, the public market participant whose allocation decisions determine which technologies are built, how quickly they scale, and how much pressure they exert on the governance systems that must eventually address them. If entrepreneurial optimism is the engine of acceleration, capital is its fuel, and the cognitive dynamics of capital allocation deserve examination in their own right.

The immediate question is whether investors correct or amplify entrepreneurial bias. There is a plausible case for correction. Investors are, in principle, professional evaluators of risk. They see many ventures rather than one, which should give them access to precisely the outside view that section 2.4 identified as the corrective to the planning fallacy: an investor who has observed a hundred ventures in a sector possesses, in principle, the reference class that the individual founder lacks. They are financially motivated to be accurate rather than optimistic, since their returns depend on distinguishing genuine opportunities from compelling narratives. And they are structurally independent of the ventures they assess, free of the personal attachment that Chapter 2 identified as a source of distortion in founders. On this account, the capital allocation layer should function as a debiasing mechanism, filtering entrepreneurial optimism before it translates into deployed technology.

The evidence, and the logic of the model, suggest the opposite is closer to the truth. Investors are subject to the same cognitive biases as entrepreneurs, and in several respects their situation intensifies rather than mitigates them. They too must decide under uncertainty; they too rely on narrative to make sense of ventures whose prospects cannot be calculated; and they too are exposed to the availability dynamics that make recently successful categories appear more promising than a dispassionate assessment would support. More importantly, investors operate under competitive pressures that specifically reward participation in optimism rather than resistance to it. A fund that declines to invest in a sector that subsequently performs well suffers reputational and financial consequences; a fund

that invests alongside its peers in a sector that collapses suffers considerably less, since the failure is shared and attributable to conditions rather than judgment. This asymmetry — the professional safety of being wrong in company — creates a powerful incentive toward herding, and herding converts individual optimism into collective conviction.

The consequence for the Cognitive Gap Model is significant. Rather than filtering entrepreneurial optimism, the capital allocation layer amplifies it, and does so through a mechanism that closely resembles the availability cascade documented in Chapter 3. A technology attracts initial investment on the strength of a compelling narrative; that investment is itself taken as evidence of the technology's promise; the resulting attention draws further investment; and the perception of inevitability that emerges from this process bears an increasingly tenuous relationship to the underlying evidence. Capital, in other words, does not merely fund acceleration. It participates in the cognitive dynamic that produces acceleration, converting the optimism of individual founders into a sector-wide conviction that operates as a social fact regardless of its evidential basis.

This has a direct implication for the risk signal at the heart of the model. Chapter 4 argued that the corrective feedback loop is disabled on the entrepreneurial side because optimism suppresses the transmission of risk information. The inclusion of the capital layer shows that this suppression is not merely a matter of individual founders failing to perceive dangers. It is reinforced by an entire ecosystem whose participants have aligned incentives to sustain the optimistic narrative — founders who need capital, investors who need deal flow and returns, and the surrounding infrastructure of advisors, media, and analysts whose relevance depends on the sector's continued prominence. The risk signal is not simply lost in transmission; it is actively crowded out by a chorus of interested optimism. When regulators eventually turn their attention to a technology, they encounter an informational environment that has been shaped, over years, by actors with a systematic interest in minimising precisely the risks that governance would need to address.

It follows that the implications developed in this chapter should extend to investors as a distinct audience. Some of the same interventions apply directly: referenceclass forecasting is, if anything, more feasible for an investor with portfolio-wide data than for a founder with a single venture, and the pre-mortem translates naturally to investment committee decisions. But the more consequential implications are structural, and they concern the incentives that produce herding. Arrangements that reward differentiated judgment rather than consensus participation, evaluation horizons long enough to distinguish skill from momentum, and disclosure practices that surface the assumptions underlying valuations rather than only their conclusions would each work against the amplification dynamic described here. Whether such arrangements can survive competitive pressure is uncertain, and the same doubt expressed in section 5.2 about entrepreneurial self-restraint applies with equal force. But the identification of capital as a third cognitive actor at least clarifies where the leverage lies, and it corrects a simplification in the model as originally stated. The gap is produced not by two actors but by three, and the third has been the least examined.

5.5 Applying the Framework: What Better Governance Would Require

It is one thing to identify mechanisms in the abstract and another to specify what their application would mean in a concrete case. Since artificial intelligence has served throughout this thesis as the central instance of the governance problem, it is appropriate to close the constructive portion of the argument by asking what the Cognitive Gap Model implies for the governance of AI specifically. This exercise is not offered as a policy proposal — the limitations acknowledged in the next section preclude that — but as a demonstration that the model generates concrete implications rather than only general observations.

Consider first what the model implies about the diagnosis. Chapter 3 examined the European Union's AI Act and found that, despite its genuine ambition, it exhibits the cognitive signature the model predicts: a risk taxonomy that enumerates presently understood categories, encoding current knowledge into legal structure with a confidence that the pace of the technology does not warrant, and creating gaps wherever novel applications escape the enumeration. The model's diagnosis is therefore not that the Act is too strict or too permissive — the axis on which most public debate about it proceeds — but that it is constructed on a cognitive assumption that its own subject matter falsifies. A framework that enumerates risks assumes that the relevant risks are known. The illusion of understanding is precisely the disposition that makes such an assumption feel safe.

What would a framework built on the opposite assumption look like? The model suggests three design principles, each corresponding to a bias identified in the analysis. The first responds to the illusion of understanding: a governance framework for a rapidly evolving technology should be constructed around processes for identifying risks rather than around lists of risks already identified. The difference is not cosmetic. An enumerated framework becomes obsolete when reality departs from the enumeration; a process-based framework, which specifies who must assess what, on what evidence, at what intervals, and with what consequences, continues to function when the specific risks change. The former encodes present understanding; the latter encodes a method for updating understanding. Only the second is robust to the discovery that present understanding was incomplete.

The second principle responds to status quo bias, and it is the sunset logic discussed in section 5.3 applied at the level of an entire regulatory framework rather than an individual rule. If the analysis of Chapter 3 is correct, the greatest long-term danger to AI governance is not that current frameworks are imperfect — all initial frameworks are — but that their imperfections will be preserved by institutional inertia long after they have become evident, as occurred with Section 230. A

framework designed against this danger would build reconsideration into its own structure: mandatory review at defined intervals, with the burden of justification falling on continuation rather than on change, and with the review conducted by a body that has no institutional investment in the framework's original design. The point is not that AI governance should be temporary, which would sacrifice the legal certainty that regulated actors legitimately require, but that its permanence should be a decision periodically taken rather than a default silently maintained.

The third principle responds to the availability heuristic and to the timing problem it creates. The model predicts that AI governance will be driven by whichever harms become vivid first, and that the risks which materialise slowly or diffusely — effects on labour markets, on epistemic environments, on the concentration of economic power — will receive attention only after the vivid ones have been addressed, if at all. An institution designed against this pattern requires a dedicated anticipatory function, insulated from the pressure to respond to current controversy, whose specific mandate is to attend to what is not yet salient. The sandbox mechanism discussed above serves part of this purpose by generating early information, but generating information is not the same as attending to it. The anticipatory capacity must have standing sufficient to place its findings on the regulatory agenda even when no crisis has made them urgent — which is, precisely, the hardest thing for such a body to achieve, since the political economy examined in Chapter 3 supplies no reward for acting on risks that have not yet frightened anyone.

Applying these principles to the entrepreneurial side yields a corresponding set of implications. If the capital and development ecosystem systematically suppresses risk information, as sections 5.2 and 5.4 have argued, then governance frameworks cannot rely on that ecosystem to supply the evidence on which regulation depends. This suggests that requirements for the disclosure of capabilities, evaluations, and known limitations should be understood not merely as transparency measures but as corrections to a documented cognitive failure: they compel the transmission of

a signal that the ordinary dynamics of the sector suppress. The justification for such requirements is stronger under the Cognitive Gap Model than under conventional accounts, because the model explains why voluntary disclosure will systematically under-supply the relevant information even when the actors involved are acting in good faith. The problem is not primarily dishonesty. It is that an ecosystem organised around optimistic conviction does not perceive the risks it would need to disclose.

Taken together, these implications describe a governance posture quite different from the one that currently prevails — process rather than enumeration, periodic reconsideration rather than durable settlement, institutionalised anticipation rather than reactive response, and compelled rather than voluntary transmission of risk information. Whether such a posture is politically achievable is a separate question on which this thesis takes no position. What the model establishes is that these design choices are not arbitrary preferences among governance styles. Each addresses a specific, documented cognitive failure, and the case for each rests on the analysis developed across the preceding chapters. This is what it means for a diagnostic model to have constructive implications: not that it dictates policy, but that it explains why particular institutional designs are likely to succeed where others predictably fail.

5.6 Limitations

An honest account of this thesis must address what it does not establish, and the limitations are substantial enough to warrant direct statement rather than the perfunctory acknowledgement that concluding chapters often provide. Four in particular deserve attention.

The first and most fundamental is methodological. This is a theoretical and conceptual work. It constructs a model, argues for its coherence, and demonstrates

its correspondence to documented cases, but it conducts no primary empirical research of its own. Every empirical claim it makes is borrowed from studies designed to answer other questions, and the synthesis it performs — connecting entrepreneurial cognition, regulatory cognition, and the dynamics of technology governance — has not itself been tested. The model is therefore plausible rather than established, and the distinction matters. A coherent model that fits the available evidence is a contribution, but it is a contribution of a specific and limited kind: it proposes a way of understanding a phenomenon and specifies how that understanding might be verified. It does not verify it. Readers should treat the Cognitive Gap Model as a hypothesis with argumentative support rather than as a finding.

The second limitation concerns the evidential base. The studies on which the analysis rests are drawn overwhelmingly from Western research traditions, conducted largely with Western participants, and concerned largely with Western institutions. This is a well-known constraint on behavioural research generally, and it bears on the present argument in a specific way: the biases documented in Chapters 2 and 3 may vary in strength or expression across cultural contexts, and the institutional dynamics through which they are amplified certainly do. The thesis has been careful to draw its regulatory examples from more than one jurisdiction, and the comparison between European and American approaches in Chapter 3 was designed partly to show that the cognitive pattern is not an artefact of a single legal tradition. But two jurisdictions from the same broad institutional family do not establish generality, and the claim that the model describes a general feature of how institutions confront the unprecedented remains, on the present evidence, an extrapolation.

The third limitation concerns the remedies discussed in this chapter, and it is the one this thesis is most concerned to state plainly. The debiasing literature is considerably less confident than the enthusiasm surrounding individual techniques would suggest. Reviews of the debiasing literature caution that the evidence does not

establish that any specific approach will prove effective in a specific future application, and that debiasing interventions may carry unintended consequences (Larrick, 2004). Techniques that reduce one bias may amplify another; interventions that work in laboratory conditions may fail in the field; and the heuristics that debiasing seeks to interrupt are, in many contexts, genuinely useful, so that suppressing them carries costs of its own. The recommendations offered in sections 5.2 through 5.5 should therefore be read as informed suggestions grounded in the best available evidence, not as validated prescriptions. It would be a peculiar failure for a thesis about overconfidence to conclude with an overconfident account of its own remedies, and the caution expressed here is not modesty but consistency.

The fourth limitation is conceptual, and it concerns the boundary between the cognitive and the structural that the thesis has worked to establish. Chapter 4 argued that the cognitive dimension explains a residue that structural factors leave unaccounted for, and offered as evidence the directional character of regulatory failure — the fact that regulation fails in patterned rather than merely slow ways. The argument is, in the author's judgement, sound. But the separation between cognitive and structural causes is cleaner in analysis than in reality. Institutions shape the cognition of those who inhabit them; cognition shapes the institutions that are built. The two are entangled in ways that a model distinguishing them must to some extent simplify. This thesis has treated that simplification as analytically productive, and it has argued in Chapter 3 that the cognitive and structural accounts are reinforcing rather than competing. But a more integrated treatment, one that theorised the mutual constitution of institutional structure and individual cognition rather than layering one upon the other, would be a genuine advance on what is offered here.

A fifth limitation deserves mention because it concerns the scope of the argument rather than its execution. This thesis has framed the innovation-governance gap as a problem, and its constructive chapter has proceeded on the assumption that narrowing the gap is desirable. That assumption is defensible but not self-evident.

A society whose governance kept perfect pace with its technology would be one in which no technology arrived before institutions were ready for it, which might also be a society in which fewer technologies arrived at all. The gap that this thesis has criticised is, on one reading, the space in which innovation happens — the interval of regulatory permission, however inadvertent, within which new things can be tried before anyone has decided whether to allow them. Section 2.7 raised this possibility at the level of the individual entrepreneur, in asking whether optimistic bias might be functionally necessary; the same question arises at the level of the system. It is possible that a certain amount of governance lag is the price of technological dynamism, and that a perfectly synchronised system would purchase its safety at a cost in innovation that few would accept if the trade were made explicit. This thesis has not resolved that question. Its position is the narrower one that the gap as it currently operates is not the product of a deliberate trade-off but of unexamined cognitive failure, and that a society which decided to accept some governance lag as the price of dynamism would be in a very different position from one that suffers it without ever having chosen it. Making the tradeoff visible is a precondition for making it well, and that is the more modest claim on which the constructive argument of this chapter ultimately rests.

5.7 Conclusions

This thesis began with a question: to what extent do cognitive biases in entrepreneurial and regulatory actors explain the persistent gap between technological innovation and its governance? It is time to answer it.

The answer this work has developed is that cognitive bias explains a substantial and, more importantly, a specific part of that gap — the part that structural accounts leave unexplained, and the part that matters most for understanding why the gap persists. Structural factors establish that governance of new technology will be

slow and difficult. They do not explain why it fails in the particular ways it does: why regulatory attention lands on the vivid rather than the significant; why legislators approach genuinely novel technologies with a confidence their understanding does not warrant; why the categories of an earlier era are retained long past their obsolescence; and why entrepreneurial acceleration meets so little effective resistance from the institutions charged with governing it. These are not the neutral symptoms of delay. They are directional distortions, and their direction is precisely what the cognitive account predicts.

The thesis has argued further that these two sets of distortions are not independent. The Cognitive Gap Model holds that entrepreneurial and regulatory biases interact, disabling at both ends the feedback loop through which a healthier relationship between innovation and governance would correct itself. Entrepreneurial optimism suppresses the risk signal before it can be transmitted; regulatory bias misperceives and mishandles what survives. A corrective loop broken at both ends cannot restore alignment, and the gap it fails to close therefore widens rather than stabilising. This is why the gap is properly described as multiplicative rather than additive, and why addressing either side alone is insufficient.

Three consequences follow, and they are worth stating as the substantive claims on which this work rests. The first is that the governance gap is not an accident. It is the predictable output of a system whose components are biased in opposing and mutually reinforcing directions, and it will recur, with variations, wherever those conditions obtain. The recurrence documented across financial derivatives, social media, and artificial intelligence is not a series of unrelated failures but the repeated expression of a single dynamic. The second is that the gap is not, for that reason, inevitable. A cognitive account, unlike a purely structural one, identifies mechanisms that can be interrupted; the sandbox and the sunset clause demonstrate that even the Collingridge dilemma, the most fundamental of the structural constraints, admits of partial institutional escape. The third is that the most promising interventions are institutional rather than individual. Biases are difficult

to defeat inside a single mind and easier to compensate for through the design of the environment in which decisions are made — which places the primary responsibility not on entrepreneurs to be less optimistic or legislators to be more perceptive, but on institutional designers to build arrangements that function despite the cognitive limitations of the people who staff them.

It remains to say what this analysis does not claim. It does not claim that cognitive bias is the sole cause of the governance gap; structural and political factors are real and have been acknowledged throughout. It does not claim that the model has been empirically established; it has been argued for, and the terms of its testing have been specified. And it does not claim that the interventions discussed in this chapter will reliably work; the evidence on debiasing counsels a caution that this thesis has tried to observe rather than to override.

What the analysis does claim is that the cognitive dimension of the innovation–governance gap is real, specific, and largely absent from the frameworks that currently dominate the field. The leading accounts of responsible innovation and anticipatory governance describe with sophistication what good governance of technology would look like, while saying comparatively little about the mechanisms that make it so consistently unachievable. This thesis has attempted to supply part of that missing account, and to show that the missing part is cognitive.

There is a final observation that the argument invites, and it returns to the reflexive point acknowledged at the outset. A thesis about the systematic distortion of human judgment is written by a human being whose judgment is subject to the same distortions. The model developed here is a construction of the author's own cognition, assembled from evidence selected by that cognition, and organised around a framework whose appeal may owe something to its coherence rather than its truth. The narrative satisfaction of an argument is, as Chapter 2 observed of entrepreneurial storytelling, no guarantee of its correspondence to reality. This is

not a reason to discount the analysis, but it is a reason to hold it in the manner the analysis itself recommends: as a *provisional* account, offered for testing rather than for acceptance, and subject to the same demand for disconfirming evidence that it has urged upon entrepreneurs and legislators alike. The gap between how fast technology moves and how well institutions understand it is real, consequential, and widening. If this thesis has sharpened the terms in which that gap can be described and analysed, and if it has made the cognitive dimension of the problem harder to overlook, it will have achieved what it set out to do.

Bibliography

- Amore M. D., Garofalo O. e Martin-Sanchez V. (2021). Failing to learn from failure: How optimism impedes entrepreneurial innovation. *Organization Science*, 32(4), 940-964.
- Baron R. A. (2008). The role of affect in the entrepreneurial process. *Academy of Management Review*, 33(2), 328-340.
- Baron R. A., Hmieleski K. M. e Henry R. A. (2012). Entrepreneurs' dispositional positive affect: The potential benefits - and potential costs - of being “up”. *Journal of Business Venturing*, 27(3), 310-324.
- Buehler R., Griffin D. e Ross M. (1994). Exploring the “planning fallacy”: Why people underestimate their task completion times. *Journal of Personality and Social Psychology*, 67(3), 366-381.
- Busenitz L. W. e Barney J. B. (1997). Differences between entrepreneurs and managers in large organizations: Biases and heuristics in strategic decisionmaking. *Journal of Business Venturing*, 12(1), 9-30.
- Camerer C. e Lovallo D. (1999). Overconfidence and excess entry: An experimental approach. *American Economic Review*, 89(1), 306-318.
- Cardon M. S., Wincent J., Singh J. e Drnovsek M. (2009). The nature and experience of entrepreneurial passion. *Academy of Management Review*, 34(3), 511-532.
- Collingridge D. (1980). *The Social Control of Technology*. St. Martin's Press, New York.
- Cooper A. C., Woo C. Y. e Dunkelberg W. C. (1988). Entrepreneurs' perceived chances for success. *Journal of Business Venturing*, 3(2), 97-108.

- Cossette P. (2015). Heuristics and cognitive biases in entrepreneurs: A review of the research. *Journal of Small Business and Entrepreneurship*, 27(5), 471-496.
- Downes L. (2009). *The Laws of Disruption: Harnessing the New Forces that Govern Life and Business in the Digital Age*. Basic Books, New York.
- Englich B., Mussweiler T. e Strack F. (2006). Playing dice with criminal sentences: The influence of irrelevant anchors on experts' judicial decision making. *Personality and Social Psychology Bulletin*, 32(2), 188-200.
- Fernbach P. M., Light N., Scott S. E., Inbar Y. e Rozin P. (2019). Extreme opponents of genetically modified foods know the least but think they know the most. *Nature Human Behaviour*, 3(3), 251-256.
- Fernbach P. M., Rogers T., Fox C. R. e Sloman S. A. (2013). Political extremism is supported by an illusion of understanding. *Psychological Science*, 24(6), 939-946.
- Flyvbjerg B. (2008). Curbing optimism bias and strategic misrepresentation in planning: Reference class forecasting in practice. *European Planning Studies*, 16(1), 3-21.
- Gibson B. e Sanbonmatsu D. M. (2004). Optimism, pessimism, and gambling: The downside of optimism. *Personality and Social Psychology Bulletin*, 30(2), 149-160.
- Guston D. H. (2014). Understanding “anticipatory governance”. *Social Studies of Science*, 44(2), 218-242.
- Hmieleski K. M. e Baron R. A. (2009). Entrepreneurs' optimism and new venture performance: A social cognitive perspective. *Academy of Management Journal*, 52(3), 473-488.
- Kahneman D. e Tversky A. (1979). Intuitive prediction: Biases and corrective procedures. *TIMS Studies in Management Science*, 12, 313-327.

- Klein G. (2007). Performing a project premortem. *Harvard Business Review*, 85(9), 18-19.
- Knight F. H. (1921). *Risk, Uncertainty and Profit*. Houghton Mifflin, Boston.
- Koellinger P., Minniti M. e Schade C. (2007). "I think I can, I think I can": Overconfidence and entrepreneurial behavior. *Journal of Economic Psychology*, 28(4), 502-527.
- Kuran T. e Sunstein C. R. (1999). Availability cascades and risk regulation. *Stanford Law Review*, 51(4), 683-768.
- Larrick R. P. (2004). Debiasing. In Koehler D. J. e Harvey N. (eds.), *Blackwell Handbook of Judgment and Decision Making*, 316-337. Blackwell, Oxford.
- Lovallo D. e Kahneman D. (2003). Delusions of success: How optimism undermines executives' decisions. *Harvard Business Review*, 81(7), 56-63.
- Popper K. R. (1959). *The Logic of Scientific Discovery*. Hutchinson, London.
- Ranchordás S. (2015). Sunset clauses and experimental regulations: Blessing or curse for legal certainty? *Statute Law Review*, 36(1), 28-45.
- Rozenblit L. e Keil F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science*, 26(5), 521-562.
- Samuelson W. e Zeckhauser R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty*, 1(1), 7-59.
- Sarasvathy S. D. (2001). Causation and effectuation: Toward a theoretical shift from economic inevitability to entrepreneurial contingency. *Academy of Management Review*, 26(2), 243-263.
- Schwenk C. R. (1984). Devil's advocacy in managerial decision-making. *Journal of Management Studies*, 21(2), 153-168.

- Shepherd D. A., Williams T. A. e Patzelt H. (2015). Thinking about entrepreneurial decision making: Review and research agenda. *Journal of Management*, 41(1), 11-46.
- Stigler G. J. (1971). The theory of economic regulation. *The Bell Journal of Economics and Management Science*, 2(1), 3-21.
- Stilgoe J., Owen R. e Macnaghten P. (2013). Developing a framework for responsible innovation. *Research Policy*, 42(9), 1568-1580.
- Tversky A. e Kahneman D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5(2), 207-232.
- Tversky A. e Kahneman D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124-1131.
- Ucbasaran D., Westhead P., Wright M. e Flores M. (2010). The nature of entrepreneurial experience, business failure and comparative optimism. *Journal of Business Venturing*, 25(6), 541-555.
- Zhang S. X. e Cueto J. (2017). The study of bias in entrepreneurship. *Entrepreneurship Theory and Practice*, 41(3), 419-454.
- European Parliament and Council of the European Union (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union*, L 2024/1689, 12 July 2024.
- United States Congress (1996). *Communications Decency Act of 1996*, 47 U.S.C. § 230.

Web Sources

Cambridge Forum on AI: Law and Governance. Experimentalism beyond ex ante regulation: A law and economics perspective on AI regulatory sandboxes, <https://www.cambridge.org/core/journals/cambridge-forum-on-ai-law-andgovernance>

European Journal of Risk Regulation. Getting Regulatory Sandboxes Right: Design and Governance Under the AI Act, <https://www.cambridge.org/core/journals/european-journal-of-riskregulation>

Journal of Regulatory Economics. An iridescent sunset: An empirical analysis of sunset legislation, <https://link.springer.com/article/10.1007/s11149-02509498-5>

Nature. Scientific Reports. Debiasing training reduces confirmation bias in national risk analysts, <https://www.nature.com/articles/s41598-025-28794-w>