

Cambridge Elements

Austrian Economics

Hayek for Our Time

Cass R. Sunstein

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

ISSN 2515-0693 (online)
ISSN 2515-0685 (print)

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

Cambridge Elements

Elements in Austrian Economics
edited by
Peter Boettke
George Mason University

HAYEK FOR OUR TIME

Cass R. Sunstein
Harvard University

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

Shaftesbury Road, Cambridge CB2 8EA, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre,
New Delhi – 110025, India

Cambridge University Press is part of Cambridge University Press & Assessment,
a department of the University of Cambridge.

We share the University's mission to contribute to society through the pursuit of
education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781047826150

DOI: 10.1017/9781047826167

© Cass R. Sunstein 2026

This publication is in copyright. Subject to statutory exception and to the provisions
of relevant collective licensing agreements, no reproduction of any part may take
place without the written permission of Cambridge University Press & Assessment.

When citing this work, please include a reference to the DOI 10.1017/9781047826167

First published 2026

A catalogue record for this publication is available from the British Library

*A Cataloging-in-Publication data record for this Element is available from
the Library of Congress*

ISBN 978-1-047-82615-0 Hardback
ISBN 978-1-047-82613-6 Paperback
ISSN 2399-651X (online)
ISSN 2514-3867 (print)

Cambridge University Press & Assessment has no responsibility for the persistence
or accuracy of URLs for external or third-party internet websites referred to in this
publication and does not guarantee that any content on such websites is, or will remain,
accurate or appropriate.

For EU product safety concerns, contact us at Calle de José Abascal, 56, 1°, 28003
Madrid, Spain, or email eugpsr@cambridge.org

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

Hayek for Our Time

Elements in Austrian Economics

DOI: 10.1017/9781047826167
First published online: November 2026

Cass R. Sunstein
Harvard University

Author for correspondence: Cass R. Sunstein, csunstei@law.harvard.edu

Abstract: Friedrich Hayek is one of the most important thinkers of the modern era. He is immensely relevant to the present period. In important ways, the twenty-first century ought to belong to him. Hayek was the Apostle of Ignorance. He offered enduring lessons about undesigned orders, traditions, psychology, markets, coercion, and the rule of law. He celebrated designs without designers. He emphasized how much individuals benefit from practices and institutions that no individual created and that individuals do not and cannot understand. He linked his claims about undesigned orders and dispersed knowledge with his arguments about freedom and the constitution of liberty. Hayek's claims about human ignorance cast new light on current issues and debates. He shows the possibility of Hayekian behavioral economics. He shows the limits of artificial intelligence – what it cannot do now, and what it cannot ever do. All over the world, nations can learn from Hayek.

Keywords: Austrian economics, knowledge problem, artificial intelligence, traditions, free markets, socialism, planning, freedom

© Cass R. Sunstein 2026

ISBNs: 9781047826150 (HB), 9781047826136 (PB), 9781047826167 (OC)
ISSNs: 2399-651X (online), 2514-3867 (print)

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

Contents

	Introduction	1
1	Ignorance	4
2	Mistakes	12
3	Hayekian Fuel Economy Standards	22
4	The AI Calculation Debate and the Socialist Calculation Debate Are the Same Thing	31
5	Hayek, Mill, and Tradition	47
	Epilogue	59
	References	61

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

> It is high time, however, that we take our ignorance more seriously.
>
> –Friedrich Hayek

> In no system that could be rationally defended would the state just do nothing. An effective competitive system needs an intelligently designed and continuously adjusted legal framework as much as any other. Even the most essential prerequisite of its proper functioning, the prevention of fraud and deception (including exploitation of ignorance), provides a great and by no means yet fully accomplished object of legislative activity.
>
> –Friedrich Hayek

Introduction

Friedrich Hayek is one of the most important thinkers of the modern era, and among the most intriguing. He is immensely relevant to the present period. In some ways, the twenty-first century ought to belong to him.

Hayek offered enduring lessons about undesigned orders, traditions, psychology, markets, coercion, and the rule of law. He celebrated designs without designers. He emphasized how much we benefit from practices and institutions that no individual created and that we do not and cannot understand. He linked his claims about undesigned orders and dispersed knowledge with his moral and economic arguments about freedom, the rule of law, and the constitution of liberty.

Hayek is also an immensely polarizing thinker. Some people demonize him. They think of him as a satanic figure, the inspiration for all sorts of horrible things. They think that he is the principal thinker behind "neoliberalism," which is, in their view, responsible for economic inequality, an inadequate social safety net, the excesses of the free market, and the plight of workers.

But some people worship Hayek; they consider him to be some kind of divinity. They think that if you are asked a hard question, you should ask this one instead: "What would Hayek think?" They believe that if you obtain the answer to that question, you know what *everyone* should think. There are Hayek loyalists, with an intense emotional commitment to Hayek. There is also a cult of Hayek, and it is not entirely different from the cults of Karl Marx, John Rawls, Michel Foucault, and Ayn Rand – and in some ways, with the cults of Bob Dylan, the Beatles, and Star Wars. If you read Hayek, you enter a whole world, one with a great deal of order and coherence and even majesty. It puts everything, or most things, in place. It is immensely repetitive, and not worse for that. (Familiar songs can sound best; they are comforting.) It is inspiring and in its way irresistible.

I understand the cult, I think, and have spent a little time in it. But while I greatly admire Hayek, I am not a member.

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

I received a glimpse of the demonization during an interview with The New Yorker, in which Isaac Chotiner, a journalist known for his "gotcha" interviews, noted my enthusiasm for Hayek and then asked me this: "Hayek talked about disliking 'Near Eastern' students because they were 'fundamentally dishonest.' He said that he had a 'profound dislike for the typical Indian students at the London School of Economics, which I admit are all one type – Bengali moneylender sons.'"

My answer:

> I didn't know about that. Wow. Hayek was a very complicated figure. He wrote a lot. Some of it's not so good. The Hayek that I admire is the Hayek who was an extremely persistent and repetitive and powerful critic of socialist planning on the ground that it couldn't be made to work because the socialist planners won't know enough, even if they're competent and well motivated. And also the Hayek who would occasionally say that liberty has an economic foundation, but also a moral foundation. And, in a time like ours, Hayek as Mr. Liberty – that's someone who we want to be in the same club with.

Then Chotiner's trap was sprung: "Hayek supported the 1973 coup in Chile. There are lots of things that don't make him seem like such a democrat or liberal, correct?"

Then me, lightly: "The criticisms are either accurate or, in some sense, fair. So when I say Hayek is a great figure or a great part of the liberal tradition, I don't mean to endorse everything he said. I tried to be measured, but I got a flood of hate mail."

I have learned an immense amount from Hayek specialists: a group of dedicated scholars who are, in a sense, keeping the flame alive. They have done exceedingly important work in helping people to understand Hayek's complicated, voluminous oeuvre, to keep up with his many twists and turns, and to see how things do or do not fit together. I am not trying to do anything like that here. This is not a summary of Hayek, or anything like an intellectual biography.

My aims are narrower and more specific. I want to show that Hayek is the Apostle of Ignorance. I want to show that Hayek's claims about our ignorance – I am tempted to say Our Ignorance – bear on current issues and debates, including uses of behavioral economics and debates about the potential of artificial intelligence and the evaluation of traditions. I want to suggest how behavioral economists and AI specialists might learn from Hayek and do their business better. Actually, I want to suggest how all of us might learn from Hayek and do our business better. With Hayek's help, I want to show the limits of prediction. I want to say a few things about traditions and those who do not like them. My goals, in the end, are to say what I think is true, not to insist that something is true because Hayek thought it.

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

Hayek was not right on everything. He was full of insights, but he could be dogmatic. Some of his ideas were, I fear, a bit like slogans. Abstractions did too much work. He was too sure of himself. We could do better than he did, not only because we know so much more, but also because we can be, and should be, more pragmatic and more empirical. Still, we have a ton to learn from him. He repays rereading. Some of his writings are like poems or novels, in which you discover something new or interesting, or different, every time.

Speaking of which: What would Hayek think about something on which he did not write and about which he knew little or nothing – say, contemporary behavioral economics or artificial intelligence? This is a complicated matter. There might be no fact of the matter. We cannot read Hayek and then announce the answer. We might try to do that, but sometimes we really cannot, because Hayek did not speak to the topic.

With respect to the views of dead thinkers, answers to many particular questions are inevitably interpretive in Ronald Dworkin's sense (Dworkin, 1985): they must attempt (1) to fit the materials to be interpreted and (2) to justify them, that is, to put them in the best constructive light. It is of course true that if an approach does not fit the materials, it cannot claim to be interpretive. But with respect to many specific questions of interpretation, more than one conclusion will do well, or well enough, along the dimension of fit, and the real question is one of justification. What looks like (1), or what purports to be (1), is often (2).

That is, when a follower of Kant urges that "Kant would say x," or that "Kantianism entails y," the enterprise may be to make the best constructive sense of Kant and Kantianism, not to follow some edict of Kant himself. The interpreter might even be said to be authorizing a new chapter in a kind of chain novel, whose initial chapters were written by Kant, but which Kant did not complete (Dworkin, 1985). Which understanding of Kant puts his views in the best constructive light? Because that is often the inevitable question, it should not be amazing to find that Jones' preferred interpretation of Kant fits with Jones' normative views, while Smith's interpretation fits with Smith's. The dispute between Jones and Smith might not be resolvable without speaking of justification – of what does, in fact, put Kant in the best constructive light. Inevitably, it is an interpretation, in which we attempt to make the best constructive sense of what he actually said and thought.

I need to say this by way of preface, because some of my discussion will invoke Hayek to cast light on issues that Hayek did not engage, and because some followers of Hayek will dispute, not unreasonably, what I have to say. To say the least, Hayek was not enthusiastic about government regulations, and I will be offering a Hayekian understanding and defense of fuel economy regulations. Modern Hayekians might reasonably cry: Foul!

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

So be it. The question is whether the defense works, given the truth in what Hayek said and wrote. Let's see.

1 Ignorance

Hayek is generally known as a fierce critic of socialism and a committed advocate for free markets. So he was. His most famous book is called *The Road to Serfdom* (1944), and he did not approve of that road. He despised socialism. He thought that planners were doomed to fail (and to destroy freedom). His objections to planning have immense relevance and resonance today, whether we are speaking of market-friendly nations that occasionally plan (such as the United States and Germany) or market-skeptical nations that prize planning (such as China and Russia).

Both sets of nations have a great deal to learn from Hayek (Boettke, Candela, & Truitt, 2024). If, for example, a nation in North America or Europe is taking steps to address climate change, it should have Hayek's work in firm view: *Planners beware*. I worked in the US government on climate change issues under President Barack Obama, and I did try to keep Hayek in mind, not least in issuing energy efficiency regulations, which are a form of planning.

Even so, Hayek's fundamental preoccupation, and what focused his attention from the 1930s to the 1990s, lay elsewhere, or more precisely, lay behind and above everything that he criticized and everything that he advocated. His preoccupation, in short, was ignorance (Caldwell, 2003). Hayek was the Apostle of Ignorance. In a sense, he was a celebrant of ignorance, because he associated it with freedom. He emphasized what human beings cannot know, and what they can never know. Time and again, he drew attention to information that we do not have, and that we will never have.

In this way, he went well beyond thinkers in the Austrian School of Economics, who also focused on knowledge and ignorance, who taught him so much, and where he encamped, in one or another respect, for most of his life (Caldwell & Klausinger, 2022). In particular, Hayek went beyond Ludwig von Mises, his brilliant, bold, iconoclastic, persistent, rigid, edgy, and cranky mentor, who first saw why socialism could not work, and who explained the impossibility of socialist planning with immense clarity. Hayek was deeper than Mises, gentler, more curious, more wide-ranging, less edgy, less harsh, more pragmatic, and somehow less ideological. He took the knowledge problem, as it is often called, from Mises (and others in the Austrian School), but he did more with it than anyone else did, and he saw it in all sorts of places. His intellectual and personal trajectories are endlessly fascinating (Caldwell & Klausinger, 2022).

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

Hayek's claims about ignorance organized everything, or almost everything, that he wrote. They undergirded his negative claims, that is, his objections to his assortment of bête noires: planning, designed orders, constructivism, rationalism, scientism, socialism, commands, abolition of private property, price-fixing, and hubris of all kinds. (Are all these in some sense the same? For Hayek, I think so.) They paved the way for his positive claims, that is, his claims for individualism, markets, traditions, customs, invisible hand mechanisms, undesigned orders, spontaneous orders, liberalism of a certain kind, Scottish enlightenment thinkers, the price signal, and the rule of law. They helped explain his frequent oppositions between what he liked and what he did not like. Indeed, that was a common strategy in his writing: *There are the constructivists, and there are the individualists, and here is what they both think, and this is why the individualists are right*.

Hayek admired what he saw as evolutionary thinkers, including those in the Scottish Enlightenment, who celebrated social learning, invisible hands, traditions, and undesigned orders (Ullmann-Margalit, 1978; Ullmann-Margalit, 1997). Adam Ferguson, David Hume, and Adam Smith pointed to the limits of reason and to designs without designers. Carl Menger, founder of the Austrian School and one of Hayek's inspirations, was an evolutionary thinker. Hayek loved Edmund Burke, who was no constructivist, and who embraced traditions and even prejudices, writing in his *Reflections on the Revolution in France* (Burke [1790] 2014): "You see, Sir, that in this enlightened age I am bold enough to confess that we are generally men of untaught feelings, that, instead of casting away all our old prejudices, we cherish them to a very considerable degree, and, to take more shame to ourselves, we cherish them because they are prejudices; and the longer they have lasted and the more generally they have prevailed, the more we cherish them." Burke's writing is unmistakably Burkean ("You see, Sir, that in this enlightened age I am bold enough to confess"). But the sentiment here is Hayekian, too ("the longer they have lasted and the more generally they have prevailed, the more we cherish them").

Hayek contrasted evolutionary thought with constructivism, exemplified by René Descartes, Voltaire, and (alas) John Stuart Mill, which, he thought, was based on a "fatal conceit," and which led to socialism. Constructivists (whom Hayek sometimes labeled as "rationalists") celebrate reason and what human beings can do with it. They think that anything can be redesigned, and that with reason, we can figure out whether it should be redesigned. Thus Voltaire's remarks, which Hayek saw as the height of folly: "if you want good laws, burn those you have and make yourselves new ones" (Hayek [1978] 2018, p. 4). Hayek hated that, and he thought that it captures a defining mistake.

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

Among other things, constructivists think that we can assess existing institutions and practices and improve on them, perhaps through planning. Constructivists assume "that all institutions which benefit humanity have in the past and ought in the future to be invented in clear awareness of the desirable effects that they produce" (Hayek, 1967, p. 85). What hubris, Hayek thought. (Cost-benefit analysts, be forewarned.) Constructivists adopt "the innocent sounding formula that, since man has himself created the institutions of society and civilisation, he must also be able to alter them at will so as to satisfy his desires or wishes" (Hayek [1978] 2018, p. 3). What foolishness, Hayek thought.

To put it mildly, Hayek did not approve of that formula. Who invented the English language, after all? Who invented German? Who invented French? He contended that, like language, "morals, law, skills and social institutions" are not a product of any "preconceived design." He drew attention to "the proper limits of reason in the arrangement of social affairs" (Hayek, 1967, p. 95). Hayek claimed that we must defer "to a moral order to which we owe our existence . . . to a tradition which we must revere and care for . . . to a system which we must accept as given . . . to a gradually evolved set of abstract rules of which human reason can avail itself to build better than it knows . . . to structures based on more information than any human agency can use" (Hayek, 1984). Note the word choice here. "Revere and care for" – there is intense feeling there, and Hayek hopes that it is contagious.

The fox, it is said, knows many things, but the hedgehog knows one big thing. Hayek knew many things, and so he was a fox; but he was mostly a hedgehog, because there was one thing that he really, really knew.

Take, for example, his Nobel Lecture, delivered on December 11, 1974. The title of the lecture: *The Pretense of Knowledge* (Hayek, 1974; the lecture can be found in many places, including Hayek, 2014b). (Pause over the title; there we go.) Hayek objected to the "scientistic attitude," modeled on the natural sciences, which is, in his view, exactly the wrong approach to the social sciences, which must focus on what we do not know. Prices and wages, for example, depend on "the effects of particular information possessed by every one of the participants in the market process – a sum of facts which in their totality cannot be known to the scientific observer, or to any other single brain." Human beings have not sufficiently appreciated that fact. But that is just an illustration of the much broader problem of "the scientistic prejudice." Consider, for example, "essentially complex phenomena," which are all around us, and which often make predictions impossible.

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

Hayek offers a simple, arresting example:

> Consider some ball game played by a few people of approximately equal skill. If we knew a few particular facts in addition to our general knowledge of the ability of the individual players, such as their state of attention, their perceptions and the state of their hearts, lungs, muscles etc. at each moment of the game, we could probably predict the outcome. Indeed, if we were familiar both with the game and the teams, we should probably have a fairly shrewd idea on what the outcome will depend. But we shall of course not be able to ascertain those facts, and in consequence the result of the game will be outside the range of the scientifically predictable, however well we may know what effects particular events would have on the result of the game.

It is not possible to predict the outcome of sports events involving players of approximately equal skill, because so much depends on what happens on the ground, and because we cannot know, in advance, what that will be. Man, Hayek urged, will have to learn that in "fields where essential complexity of an organized kind prevails, he cannot acquire the full knowledge which would make mastery of the events possible." There is a characteristic claim from the Apostle of Ignorance.

In his magnificent essay on complex phenomena, written in 1964, inevitable ignorance is also the central theme (Hayek, 1964, 2014b). (I will have a lot to say about that essay.) Focusing on natural selection, Hayek goes macro: Who could possibly have predicted that living species would take the form they have taken? The answer, of course, is no one, and the reason is similar to the failures of prediction in sports events. We cannot know what will happen on the ground. Hayek also goes micro: "almost any event in the course of a man's life may have some effect on almost any of his future actions," which "makes it impossible that we translate our theoretical knowledge into predictions of specific events." We will see plenty of examples.

For the outcomes of economic competition, Hayek urged, "the number of distinct elements involved" are "fairly large" (an understatement!): "the overall pattern that will form itself is determined by the significantly different behavior of the different individuals so that the obstacle of obtaining the relevant data cannot be overcome by treating them as members of a statistical collective." This is the ball game once more. (Again we will see plenty of examples.)

Hayek concluded that economic theory can describe "kinds of patterns which will appear if certain general conditions are satisfied, but can rarely if ever derive from this knowledge any predictions of specific phenomena" (Hayek, 1967, p. 35). If you shuffle a deck of cards, you can be confident that no card is going to appear twice. That is a *pattern prediction*. But you cannot know in what order the cards will fall. So too, you can predict that price controls will create

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

shortages, but you cannot specify their size, and you cannot offer a lot of details about who will be hurt most.

Return to evolution, one of Hayek's favorite examples (Hayek, 1964). You can be clear that because of natural selection, organisms will adapt (that is a pattern prediction). But you cannot know what specific organisms will emerge. In many domains, you can say something about the form or class of outcomes, without being able to make specific predictions with any reliability. You probably cannot predict, for example, what movies will be most successful in a given year, or foresee the outcome of a presidential election five years in advance. Too many variables are in play.

1.1 Liberty

Why respect liberty? Hayek's distinctive account was rooted in his claims about ignorance, which undergird his critique of socialism and centralized planning. He emphasized that, however well-motivated, planners have far less knowledge than participants in markets do. Many of his central insights grew out of what he saw as the fatal problems with centralized government judgments about prices and quantities. For him, those problems are defining, and they were above all epistemic. Hayek emphasized that the price system must be taken as an institution, and that on epistemic grounds, it has unique properties.

Some of Hayek's most important contributions to social thought are captured in his great (and short) 1945 essay, "The Use of Knowledge in Society" (Hayek, 1945).[1] Here again, the main emphasis is on what we cannot know, and how evolved, undesigned institutions help solve our problems. In that essay, Hayek claimed that the advantage of prices is that they aggregate both the information and the tastes of numerous people, incorporating far more material than could possibly be assembled by any other institution, such as a central planner, group, or board. Hayek emphasized the unshared nature of information – the "dispersed bits of incomplete and frequently contradictory knowledge which all the separate individuals possess." That knowledge certainly includes facts about products, but it also includes preferences and tastes, and all of these are taken into account by a well-functioning market. Hayek stressed above all the "very important but unorganized knowledge which cannot possibly be called scientific in the sense of general rules: the knowledge of the particular circumstances of time and place." Here, then, is the core of his later objection to scientism.

For Hayek, the key economic question involves a matter of institutional design: how to incorporate that unorganized and dispersed knowledge. No particular person or group can possibly solve that problem (Bergh, 2020;

[1] Superb treatments of Hayek's thought include Caldwell (2003) and Boettke (2018).

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

Boettke, Candela, & Truitt, 2024). However legitimate and appealing their goals may be, central planners, and the institutions they run, cannot have access to all of the knowledge held by diverse people. Taken as a whole, the knowledge held by those people is far greater than that held by even the wisest and most well-chosen group of experts. Hayek's central point is that the best solution comes from the price system. His claim is that in a system in which knowledge of relevant facts is dispersed among many people, prices act as an astonishingly concise and accurate coordinating, signaling, and discovery device. They incorporate dispersed knowledge and, in a sense, also publicize it, because the price itself operates as a signal to all.

Even better, the price system has a wonderfully automatic quality, particularly in its ability to respond quickly to changes. If new information shows that a product – a television, a car, a cell phone, a watch – does not always work, people's demand for it will rapidly fall, and so too the price. And when a commodity suddenly becomes scarcer, the market will respond to that fact. In Hayek's account, the price system works remarkably well as an institution, not because any participant can see all its features, but because the relevant information is communicated to everyone through prices.

Hence Hayek claims that it "[i]s more than a metaphor to describe the price system as a kind of machinery for registering changes, or a system of telecommunications which enables individual producers to watch merely the movement of a few pointers" (Hayek, 1945). Hayek describes this process as a "marvel," and adds that he has chosen that word on purpose so as "to shock the reader out of the complacency with which we often take the working of the mechanism for granted." On Hayek's account, the price system is an extraordinary device for capturing collective intelligence, in part because it collects what everyone knows, and in part because it imposes the right incentives.

1.2 Ethics and Utility

But there is a problem. It is not small, and it is one of my primary concerns here. Daniel Kahneman, Amos Tversky, Richard Thaler, George Loewenstein, Robert Shiller, George Akerlof, and legions of others have drawn attention to human error – to how human beings make systematic mistakes, even on simple questions of fact (Thaler & Sunstein, 2021). Contemporary behavioral science is full of findings about those mistakes. In view of modern behavioral findings of human error, it would be possible to object that, or at least to wonder whether, the price system is not always so marvelous, and to ask whether in principle, at least, other institutions might do better.

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

Suppose, for example, that consumers show limited attention, unrealistic optimism, or present bias, leading to foolish purchases. If so, the price signal might miss something important, and the market's system of telecommunications will give the wrong messages (Gabaix, 2017). It would also be possible to agree with Hayek's arguments about planning and prices, while also thinking that certain forms of regulation, run by public institutions, are not out of bounds. In fact, and this is often overlooked, Hayek himself did not abhor regulation as such, and Hayek did not by any means insist, or even think, that individuals are entirely rational.

Consider these words: "Probably nothing has done as much harm to the liberal cause as the wooden insistence of some liberals on some rough rules of thumb, above all the principle of laissez faire." Hayek ([1944], 2007, p. 71) did not choose his words carelessly, and it is worth pausing over that claim. Or these: "To prohibit the use of certain poisonous substances or to require special precautions in their use, to limit working hours or to require certain sanitary arrangements, is fully compatible with the preservation of competition. The only question here is whether in the particular instance the advantages gained are greater than the social costs that they impose" (Hayek [1944], 2007, p. 86). Hayek appears here to be endorsing cost-benefit analysis, undertaken by public officials. (Who else would do it?)

Of course it is fair to ask what Hayek, or a Hayekian, would include as costs and benefits, and also to wonder whether public officials, attempting to assess costs and benefits, would make serious mistakes. Hayek's concerns about constructivism and rationalism suggest that he would indeed worry about that. Still, he said what he said.

Perhaps a mandatory seatbelt law, a ban on trans fats, or regulation of exposure to certain carcinogens in the workplace would be unobjectionable. Do Hayek's arguments count against cigarette taxes, or taxes on sugar-sweetened beverages? Do they amount to a general, large-scale objection to paternalism from public institutions?

Maybe so. Hayek showed no enthusiasm for paternalism. Still, the answers to these questions are not entirely clear. In much of his work, Hayek was concerned with the largest issues – the nature of liberty, the impossibility of planning, the best conception of the rule of law, the proper role of government, the role of tradition. On many specific issues, involving which institutions should do exactly what, he did not offer particular prescriptions. Here is a revealing passage, written at a characteristically high level of generality, and worth quoting at length (Hayek [1944], 2007, p. 87):

> There are, finally, undoubted fields where no legal arrangements can create the main condition on which the usefulness of the system of competition and

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

> private property depends: namely, that the owner benefits from all the useful services rendered by his property and suffers for all the damages caused to others by its use. Where, for example, it is impracticable to make the enjoyment of certain services dependent on the payment of a price, competition will not produce the services; and the price system becomes similarly ineffective when the damage caused to others by certain uses of property cannot be effectively charged to the owner of that property.
>
> In all these instances there is a divergence between the items which enter into private calculation and those which affect social welfare; and, whenever this divergence becomes important, some method other than competition may have to be found to supply the services in question. Thus neither the provision of signposts on the road nor, in most circumstances, those of the roads themselves can be paid for by each individual user. Nor can certain harmful effects of deforestation, of some methods of farming, or of the smoke and noise of factories, be confined to the owner of the property in question or to those who are willing to submit to the damage for an agreed compensation. In such instances we must find some substitute for the regulation by the price mechanism. But the fact that we have to resort to the substitution of direct regulation by authority where the conditions for the proper working of competition cannot be created, does not prove that we should suppress competition where it can be made to function.

From this passage, we see that Hayek was keenly alert to both the public goods problem and the problem of externalities, and that he favored "some substitute for regulation by the price mechanism." But he did not spend much time specifying his preferred "substitute," or the institution that would design it.

Hayek did not much like coercion, especially, it seems, if it was paternalistic: "Coercion is evil precisely because it thus eliminates an individual as a thinking and valuing person and makes him a bare tool in the achievement of the ends of another" (Hayek [1960], 2011, p. 71). This is a strikingly Kantian formulation, not speaking of welfare at all. Notice the use of the word "evil" and the objection to treating people as means rather than ends. And indeed, Hayek seemed to embrace something like a Kantian, nonwelfarist foundation for freedom. In his introduction to *The Constitution of Liberty*, he wrote, "Some readers will perhaps be disturbed by the impression that I do not take the value of individual liberty as an indisputable ethical presupposition and that, in trying to demonstrate its value, I am possibly making the argument in its support a matter of expediency. This would be a misunderstanding" (Hayek [1960], 2011, p. 52). In that same book, he wrote that coercion "is bad because it prevents a person from using his mental powers to the full and consequently making the greatest contribution that he is capable of to the community" (pp. 200–201). That seems to combine Kantian and welfarist arguments.

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

At pivotal points, Hayek extended his epistemic argument in favor of markets to make a very general institutional claim, rooted in human ignorance, and not at all in "an indisputable ethical presupposition," on behalf of liberty and against coercion. Consider his suggestion that "the awareness of our irremediable ignorance of most of what is known to somebody is *the chief basis of the argument for liberty*. This is especially true in the economic field. If it appears that the market mechanism leads to the effective utilization of more knowledge than any directing agency can possess, this is the chief foundation of the case for economic freedom . . ." (Hayek, 2014b, p. 384). That passage is in a long-unpublished lecture. He put the same point more concisely in *The Constitution of Liberty*: "the case for individual freedom rests chiefly on the recognition of the inevitable ignorance of all of us concerning a great many of the factors on which the achievement of our ends and welfare depends" (Hayek [1960], 2011, p. 80).

For Hayek, the key point involves "irremediable ignorance," a problem that besets outsider institutions and planners of all kinds. This can be taken to be a much broader argument than his claim that the price system is "a marvel." Consider in the same vein this suggestion: "If there were omniscient men, if we could know not only all that affects the attainment of our present wishes but also our future wants and desires, there would be little case for liberty" (Hayek [1960], 2011, p. 81). That is a provocative statement, and it is worth pausing over it. Hayek roots his claim for liberty, his most cherished ideal, in the absence of "omniscient men." If there were such men, we would be able to offer "little case for liberty." I disagree with that view, for multiple reasons, and I doubt that Hayek really believed it, either; but let us not let that point detain us here.

2 Mistakes

Hayek was keenly interested in psychology. In some ways, that topic might have been his deepest passion. He wrote an ambitious and lengthy book on the subject, in which he emphasized that human beings do not have unmediated access to physical reality and instead see it through categories of their own making (Hayek [1952], 2014a). The book presaged a great deal in the way of modern understandings, and much of Hayek's argument holds up today. Still, nothing in his work in psychology foresaw modern behavioral science, and it would be implausible to say that Hayek was a behavioral economist in the contemporary sense. Nothing in his work anticipates behavioral findings;[2] he

[2] Frantz and Leeson (2013) and Frantz (2020) offer many instructive essays on Hayek, his emphasis on individual ignorance, and his conception of "behavior," but the essays do not claim that behavioral economics, in the form originated and practiced by Robert Shiller, Richard Thaler,

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

did not explore individual biases, such as present bias, optimistic bias, or availability bias. He did not have much to say about the potential relevance of behavioral findings to institutional design, to the role of government, or to the price system.

We do have to be careful here. Hayek did not embrace the "rational actor" model that dominates Chicago School economics, and indeed he argued explicitly against that model. As he wrote, "The antirationalistic approach" that he favored, and that he wanted to distinguish from constructivism, "regards man not as a highly rational and intelligent but as a very irrational and fallible being, whose individual errors are corrected only in the course of a social process" (Hayek, 1948, pp. 8–9). In a passage that should be more widely known, he lamented "the current misconceptions of the individualism of Adam Smith and his group in the common belief that they have invented the bogey of the 'economic man' and that their conclusions are vitiated by their assumption of a strictly rational behavior or generally by a false rationalistic psychology. They were," Hayek insisted, "very far from assuming anything of this kind" (Hayek, 1948, p. 11).

On the contrary, "in their view man was by nature lazy and indolent, improvident and wasteful," and "it was only by the force of circumstances that he could be made to behave economically or carefully to adjust his means to his ends. But even this would be unjust to the very complex and realistic view which these men took of human nature" (id.). Thus Hayek claimed that "we can learn more about the behavior of men" from Smith "than from most of the more pretentious modern treatises on 'social psychology'" (id.). So, perhaps, might Hayek say about contemporary behavioral economics.

We have seen that Hayek was much more than alert to the fact that individual choosers lack important information. That problem played a defining role in his thinking. He often emphasized the extent to which each of us lives amidst, and benefits from, a set of norms, cultural understandings, traditions, and institutions that were not designed by anyone, that have been built up over time, that serve essential functions, and that we do not and cannot understand (Hayek, 2014b, p. 342):

> What I want to show is that men are in their conduct *never* guided *exclusively* by their understanding of the causal connection between particular known

David Laibson, George Loewenstein, Matthew Rabin, and others, was anticipated by Hayek's work. Some work that criticizes modern behavioral economics might be described as Hayekian. *See, e.g.*, Rizzo & Whitman, 2019. But I would be cautious about the association. Hayek did not speak to the question of how to handle individual departures from rationality. He certainly did not argue that apparent departures were no such thing – in fact he argued just the opposite. We might be able to say, very loosely, that it is in a Hayekian spirit to see various behaviorally suspect choices as an outgrowth of evolved practices and norms, so that we should be cautious about interfering with them – but very loosely.

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

> means and certain desired ends, but always also by rules of conduct of which they are rarely aware, which they certainly have not consciously invented, and that to discern the function and significance of this is a difficult and only partially achieved task of scientific effort.

If there is a Hayekian theory of behavioral biases, it might begin there. Call it "Undesigned Rules Neglect." It is a serious bias, and to Hayek, the most serious of all. But if anything, Hayek is best taken to suggest not that people are knowledgeable and unbiased, but that despite their ignorance and their biases, their actual practices work well enough, or even magnificently well, as they act in accordance with rules that help us to cope with our inevitable ignorance (Caldwell, 2003; Vanberg, 2017).

Hayek's frequent emphasis on the relevance of our ignorance, and our departures from perfect rationality, to the question of institutional design is exceedingly important, and central to his thinking. Still, it should not be confused with a suggestion that choosers err when they make food choices, when they decide how to respond to safety risks, or when they select appliances or pension plans. And when Hayek spoke of the ignorance of planners, he was casting doubt on paternalism – not because choosers know best, but because planners know worst. One of his particular concerns was the process of learning and discovery that choice makes possible.

John Stuart Mill was of course the great liberal critic of paternalism, and the relationship between Hayek and Mill is both immensely complicated and fascinating (Caldwell, 2008); it will be worthwhile to turn to Mill in that light. Mill's central arguments did not involve autonomy as such, but knowledge of relevant facts, and who was most likely to have it. In *On Liberty*, Mill insisted:

> The only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others. His own good, either physical or moral, is not a sufficient warrant. He cannot rightfully be compelled to do or forbear because it will be better for him to do so, because it will make him happier, because, in the opinion of others, to do so would be wise, or even right.

Mill's argument for his famous Harm Principle is emphatically epistemic. Like Hayek, Mill believed that choosers are in the best position to know what is good for them. In Mill's view, the problem with outsiders, including but not limited to government officials, is that they lack the necessary information. Mill contended that the individual "is the person most interested in his own well-being," and the "ordinary man or woman has means of knowledge immeasurably surpassing those that can be possessed by any one else." When society seeks

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

to overrule the individual's judgment, it does so on the basis of "general presumptions," and these "may be altogether wrong, and even if right, are as likely as not to be misapplied to individual cases." If the goal is to ensure that people's lives go well, Mill concludes that the best solution is to allow people to find their own path. (As discussed in more detail in Section 5, Mill probably developed these ideas together with Harriet Taylor Mill; see Mill & Mill [1859], 2026)

Hayek agreed. But Hayek did not endorse the Harm Principle. He did say that "the most conspicuous attribute of liberalism that distinguishes it as much from conservatism as from socialism is the view that moral beliefs concerning matters of conduct which do not directly interfere with the protected sphere of other persons do not justify coercion" (Hayek [1960], 2011, p. 524). That sounds close to Mill, and it might even be seen as an effort to update Mill and be more precise; "interfere with the protected sphere of other persons" is more careful than "harm to others." But it goes by very quickly, and there is no sustained analysis, here or anywhere else, of Mill's account of liberty. The most specific statements can be found in *The Constitution of Liberty* (Hayek [1960], 2011, p. 212), where Hayek notes that Mill offered "a distinction between actions that affect only the acting person and those which also affect others." Hayek said that the distinction "has not proved very useful," because "there is hardly any action that may not conceivably affect others." For Hayek, "the important question is whether the actions of other people that we wish to see prevented would actually interfere with the reasonable expectations of the protected person." That raises its own questions, of course. What are those reasonable expectations? Hayek was clear that "the pleasure or pain that may be caused by the knowledge of other people's actions should never be regarded as a legitimate cause for coercion."

Hayek did know Mill's work well, and had a great deal to say about it (Hayek [1951], 2015), as we will see in more detail in Section 5. Recall that Hayek never claimed that individuals are unerring in their judgments about what is best for them. Far more cautiously, and characteristically, he said that "nobody can know *who* knows best," and "the only way we can find out is through a social process" (Hayek, 1948). It is speculative to wonder, but as a restriction on government action, the Harm Principle might sometimes have seemed, to Hayek, to be a bit too stringent. Is it so clear that he would have opposed motorcycle helmet laws? Laws requiring people to buckle their seatbelts? At the same time, Hayek clearly thought that the Harm Principle is too permissive. If you do not laugh at my jokes, or if you stop reading this Element, have you not harmed me? If you decide to stay home rather than to go to the local restaurant, have you not harmed that restaurant? We have evidence that Hayek did not

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

embrace the Harm Principle for just this reason. Recall the passage just quoted from *The Constitution of Liberty*, and in his seminar at the University of Chicago, Hayek worried over the potential permissiveness of the Harm Principle.

However that may be, Hayek's arguments against coercion, and Mill's defense of the Harm Principle, have been put under considerable pressure by behavioral findings. Emphasizing the occasional human propensity to blunder, and the incentives of sellers to exploit those blunders, some people have questioned whether the price system is always so marvelous (Akerlof & Shiller, 2015), and have been asking whether new institutional arrangements, involving mandates and bans, have a fresh justification (Conly, 2013; Bubb & Pildes, 2014).

Some of the answers seem profoundly anti-Hayekian; because coercion is involved, they might well be taken, on Hayekian premises, to be indefensible abridgements of liberty. But the motivation for those answers is clear: If we know that people's choices lead them in the wrong direction, why should we insist on freedom of choice? In the face of human errors, is it not odd, or even perverse, to insist on that form of freedom? Is it not especially odd to do so if we know that in many contexts, people choose wrongly, thus injuring their future selves? Ought not institutional design to take behavioral biases into account?

2.1 Information Please

No one doubts that in some cases, consumers, workers, investors, and others lack relevant information; we have seen that Hayek himself emphasized that point, not only in speaking of the dispersed nature of knowledge, but also and more relevantly, in defending government information campaigns. Thus (Hayek [1960], 2011, p. 489):

> Where . . . most individuals do not even know that there is useful knowledge available and worth paying for, it will often be an advantageous investment for the community to bear some of the costs of spreading such knowledge. We all have an interest in our fellow citizens' being put in a position to choose wisely, and if some have not yet awakened to the possibilities which technological developments offer, a comparatively small outlay may often be sufficient to induce the individuals to take advantage of new opportunities and thence to advance further on their own initiative.

Is Hayek endorsing a kind of nudging (Thaler & Sunstein, 2021)? No doubt about it (so long as we include, as we should, educative nudges as nudges). In other cases, we can identify a behavioral market failure, in the sense that people fall prey to an identifiable behavioral bias, and their choices make their lives go worse by their own lights. When this is so, a corrective response might be put on

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

the table, perhaps in the form of a nudge, perhaps in the form of a tax, perhaps in the form of a mandate. Notwithstanding Hayek and Mill, some kind of planning institution, or choice architect, might have more information than choosers do, and might not fall prey to a behavioral bias (Thaler & Sunstein, 2021). It is important, of course, to emphasize the word "might." Planners might claim to find a behavioral bias when there is no such thing; planners might have (bad) incentives of their own; planners might be subject to the influence of well-organized private groups; planners might themselves be subject to behavioral biases; social practices might work against mistakes (a Hayekian point). Still, a corrective response would seem to be on the table.

But putting a response on the table is one thing; deciding whether to impose it, and choosing its content, is another. Hayek did not speak in any detail to the question of how to handle individual departures from rationality. He seemed to think that evolved practices, including markets, norms, and traditions, would work well enough. As we have seen, he did not argue that apparent departures were no such thing; on the contrary, he insisted that they are both important and real.[3] While he offered serious objections to planning, he did not explore such questions as whether default rules are an appropriate response to inertia or procrastination, whether salient disclosures are essential to overcome limited attention, or whether fuel economy mandates might be an appropriate response to present bias and myopic loss aversion. Admittedly, all of these run into Hayekian objections. But his high-level concerns about coercion, and about the deficiencies of planners, cannot resolve concrete questions of this kind.

2.2 Hayekian Behavioral Economics?

If there is such a thing as Hayekian behavioral economics, and if institutional design might be based on it, it would firmly reject the idea that public officials should be content to identify individual errors and declare victory. An approach to institutional design with Hayekian roots, and consistent with some things that Hayek actually said, might engage in a comparative analysis: how costly are those errors, compared with the errors that would be introduced by corrective efforts?

[3] See Rizzo and Whitman (2019) for some arguments to the latter effect. Very briefly: It is true that an outside observer might not understand the utility function of a chooser. It is also true that what an outsider might take to be present bias, inertia, or limited attention might be nothing of the sort. These points are legitimate cautionary notes about behaviorally informed interventions. At the same time, departures from rationality are real. We might be able to explain them, ex post, but those explanations are sometimes what Matthew Rabin calls "explain-away-tions" (Milkman & Thaler, 2016), that is, quite implausible efforts to explain things away, or what Charles Black called "one-step-ahead-of-the-sheriff" arguments, that is, desperate arguments, trying to escape from something (here: data). Departures from rationality are not merely real; they can create serious trouble, including terrible suffering and unnecessary death. What are we going to do about that?

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

To engage in that analysis, one would have to know something about the relevant institutions. If a decision is made to proceed with some kind of remedy, a distinctly Hayekian approach, rooted also in both Kantian and utilitarian thinking, might try to reduce the knowledge problem by asking about *what individual choosers do under epistemically favorable conditions*. In fact, a stream of research is asking exactly that question.[4] In practice, it can be disciplined by asking five subsidiary questions (Allcott & Sunstein, 2015):

(1) *What do consistent choosers, unaffected by clearly irrelevant factors or frames, choose (Goldin, 2015; Goldin, 2017)?*[5]
(2) *What do informed choosers choose?*
(3) *What do active choosers choose? (If we focus on active choosers, we will protect against the possibility that outcomes are a product of inertia or procrastination.)*
(4) *In circumstances in which people are free of (say) present bias or unrealistic optimism, what do they choose?*
(5) *What do people choose when their viewscreen is broad, and they do not suffer from limited attention?*

We should not insist that Hayek, insistent on our inevitable ignorance, would have been comfortable with these questions, but they might well be seen as a neo-Hayekian program, a form of Hayekian behavioral economics. That program is also distinctive. Those involved in behavioral economics have not typically focused on those questions. In principle, the best approach would be for relevant institutions to ask all five questions. Active choosers who are uninformed might blunder; the same is true of informed choosers who procrastinate or suffer from inertia. If public institutions can learn what consistent, informed, and active choosers, uninfluenced by present bias or limited attention, choose, they might have real guidance (assuming that inconsistent, uninformed, and passive choosers, influenced by present bias or limited attention, are not otherwise differently situated from the former group).

All of the questions can be answered empirically, and what we might call neo-Hayekian behavioral economics is trying to do exactly that. For example:

(1) Should employers offer opt-in savings plans, or opt-out plans? Suppose that many employees are affected by the frame; whether they end up in a savings

[4] From varying perspectives, but within the same extended family, see Allcott, Lockwood, & Taubinsky (2019); Bernheim and Rangel (2007); Bernheim and Rangel (2009); Bernheim (2016); Bernheim and Taubinsky (2018); Goldin (2015); Allcott and Sunstein (2015).

[5] Note, however, that even if consistent choosers are unaffected by frames, they might be affected by some bias, such as present bias or optimistic bias.

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

plan depends on whether it is opt-in or opt-out. Suppose that many others are unaffected by the frame; they choose consistently. If the consistent choosers are not different from the inconsistent ones, except for the fact that the latter are affected by the frame, we have a reason to think that the choices of the consistent choosers are the right ones (Goldin, 2015, 2017).

(2) A simple absence of information might lead consumers to fail to choose fuel-efficient motor vehicles, suggesting that some kind of nudge, or perhaps even a mandate, would be a good idea. Experiments might be designed to provide consumers with relevant information and see what they choose (Allcott & Knittel, 2019). The choices of informed consumers might be taken as the foundation for analysis of the value of an intervention.

(3) Suppose that most consumers make an active choice to enroll in certain programs when those programs are designed so as to promote active choosing. If so, there is at least some reason to think that such programs are in consumers' interests. If most consumers do not enroll in such programs when active choosing is not promoted, we have reason to think that their failure to do so might be a product of inertia or inattention (Adusumalli et al., 2021).

(4) Experiments might be designed to make the potential economic savings of (say) energy-efficient appliances highly salient, at least potentially overcoming present bias and limited attention (Allcott & Taubinsky, 2015). If consumers choose or do not choose energy-efficient appliances in such circumstances, we will have learned something about what is likely to increase their welfare – not everything, but something.

(5) When people are given clarity about a wide range of product characteristics, we might investigate what they choose. Suppose, for example, that hidden fees are made clear and conspicuous, so that people with limited attention can see them. What do people choose then?

As noted, there might be heterogeneity in the relevant population, making it challenging to generalize from what some part of a population does. But suppose that there is not a lot of heterogeneity. In principle and sometimes in practice, efforts to answer these subsidiary questions should help public institutions with welfare analysis, where it is often challenging to know how to proceed when behavioral findings seem to cast doubt on standard uses of revealed preferences (Goldin, 2015).[6] And in fact efforts to answer the

[6] For a very different perspective, see Sugden (2019), which might be taken to offer a Hayekian approach to behavioral economics, one that is far less interested in correction of individual error

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

subsidiary questions might justify some kind of response, not because planners like responses in the abstract, but *because planners are building on the choices of (the right) choosers*. That, in short, can be taken as a form of Hayekian behavioral economics.

Why not?

2.3 Learning and Discovery

To be sure, there are serious Hayekian concerns. Some of the most fundamental involve learning and discovery. Through choosing, people find out what is sensible and what is not sensible, and what suits their situation and what does not suit their situation. We all benefit from that. If people are nudged through automatic enrollment, they do not make active choices for themselves, and that might be thought to truncate the process of discovery (Thaler & Sunstein 2021; Sunstein, 2027). This concern would seem to be strongest in the case of coercion, not in the case of nudges, and I am not speaking here of coercion. (We will get to that in the next section.) Still, there is a Hayekian concern whenever the relevant design reduces learning. If people are automatically enrolled in a savings plan, they may not learn about savings plans.

The best answer is that learning should not be a trump card. GPS devices should not be outlawed because they prevent people from learning (Sunstein, 2027). People might not want to make active choices; forcing them to do so is a form of paternalism. If people are content to rely on a default rule, we ought to respect their choices, in part on broadly Hayekian grounds. Unless they are to force people to choose, government officials often *have* to select one or another default rule; they cannot dispense with one. A default rule that builds on the questions sketched above is far better, on Hayekian grounds, than one that is pulled out of some kind of hat.

There are also questions about who is asking and answering the five questions. What do they know? Are they in a good position to answer them? These are, of course, core Hayekian concerns, and they might be decisive. I do not, speaking just for myself, think that they are, in part because some kind of "choice architecture" is inevitable (Thaler & Sunstein, 2021). But if we are

than the approach defended here. Sugden urges that the approach I am defending is "constructivist." Following Hayek, Sugden rightly contrasts constructivists, like Descartes, with evolutionary thinkers, like David Hume and Adam Smith. Constructivists celebrate reason and our ability to assess institutions and codes of morality in its light. Evolutionary thinkers celebrate undesigned, unintended orders, and do not think that constructivists can understand them or improve on them. Sugden thinks that in trying to produce a kind of Hayekian behavioral economics, I have ventured into the constructivist camp, and so am not really Hayekian. Maybe so, but in my view, Sugden is engaged in name-calling here. If there is a little constructivism in my approach, and if some Hayek-risen-from-the-grave might not like it, so what? For discussion, see Sunstein (2020).

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

intensely skeptical of what planners can know, we might conclude that the Hayekian concerns are reason to abandon the whole project.

2.4 Problems and Solutions

No one believes that fines or subsidies are a sufficient approach to the problem of violent crime. No one thinks that people get to choose whether to steal or to assault. Hayek certainly did not think that. In the face of a standard market failure, government intervention has a familiar justification; consider the problem of air pollution. It is true that even in such contexts, default rules may have an important role; consider the possibility of default rules in favor of clean energy (Pichert & Katsikopoulos, 2008; Ebeling & Lotz, 2015). But the effects of defaults, taken by themselves, might well prove too modest for the problem at hand, and they hardly exhaust the repertoire of appropriate responses.

The price system is indeed a marvel, but notwithstanding Hayek's arguments, the price system is not always entirely marvelous. In some areas, including occupational safety and energy policy, there are behavioral market failures (Akerlof & Dickens, 1982; Bar-Gill, 2012; Mullainathan & Shafir, 2013; Bohnet, 2016). Recall that if people are suffering from present bias, unrealistic optimism, limited attention, or a problem of self-control, prices might not capture important factors. If the result is a serious welfare loss for those people, there is an argument for some kind of public response, potentially including institutions that impose mandates. When people are running high risks of mortality or otherwise ruining their own lives, it might make sense to adopt a mandate or a ban, certainly on welfare grounds. (Admittedly, it is not at all clear that Hayek would agree; consider his sharp remarks about coercion.) After all, people have to get prescriptions for certain kinds of medicines, and even in freedom-loving societies, people are forbidden from buying certain foods, or running certain risks in the workplace, simply because the dangers are too high.

In short, behavioral findings suggest that the price signal might be misleading in important respects. Whatever Hayek thought of particular restrictions along these lines, they were not his main target. In the end, some occupational safety and health regulations must stand or fall on behavioral grounds; they forbid workers from facing certain risks, perhaps because unrealistic optimism or present bias might lead them to do so unwisely (Akerlof & Dickens, 1982). On Hayekian grounds, the best response might well be for public institutions to provide information. But we could certainly identify cases in which the best approach is a mandate or a ban, because that response is preferable, from the standpoint of social welfare, to any alternative, including information, economic incentives, or defaults. It is also true, of course, that the best approach

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

might be to do nothing – to allow market failures, including behavioral market failures, on the ground that the cure would be worse than the disease.

3 Hayekian Fuel Economy Standards

I now turn to a case study involving fuel economy mandates. I explore the possibility of rejecting the price signal and defending such mandates, as opposed to economic incentives, by reference not only to externalities but also to behavioral market failures, captured in insufficient consumer attention, ex ante, to economic savings (reduced costs of operating a car) and time savings (fewer trips to the gas station). The more general goal is to ask whether expert institutions that impose such mandates may reduce "internalities," understood as the costs that choosers impose on their future selves (Allcott & Sunstein, 2015). Note well: Hayek did not trust experts much, so we have to approach that question with fear and trembling.

Nonetheless, I intend the issue of fuel economy mandates to be exemplary. A similar analysis might be made of cigarette taxes, taxes on sugar-sweetened beverages (Allcott, Lockwood, & Taubinsky, 2019), savings policy, mask mandates, and many other problems. Admittedly, this is a bit of a venture, and perhaps an adventure. After sketching the argument, I will explore whether, on broadly Hayekian grounds, the argument might be found convincing. As we will see, the verdict will be ambivalent. (I hope you won't mind.)

Here is my starting point: In principle, fuel economy mandates, perhaps imposed by institutions charged with environmental protection, might simultaneously reduce internalities and externalities. On imaginable assumptions, such mandates might turn out to have higher net benefits than carbon taxes, because the former, unlike the latter, deliver consumer savings.[7] To say the least, this is not a conventional view, because fuel economy standards are a highly inefficient response to the externalities produced by motor vehicles, especially when compared to optimal corrective taxes (Karplus et al., 2013).

In the end, everything turns on whether the imaginable assumptions turn out to be true. What behavioral economists consider to be errors might be viewed, by some neo-Hayekians (but not, I speculate, Hayek himself!), as not errors at all. My goal is not to run the numbers or to reach a final conclusion, but to make three more general points, which are that (1) mandates might turn out to be justified on welfare grounds, (2) the standard economic preference for economic

[7] Bubb and Pildes (2014) similarly contend that fuel economy regulation might be justified by reference to behavioral considerations, but they focus only on externalities. The conclusion is much easier to justify by reference to internalities, which Bubb and Pildes bracket in their provocative discussion.

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

incentives misses something of considerable importance, and (3) a Hayekian approach allows us to make some progress in assessing (1) and (2).

Most motor vehicles emit pollution, including greenhouse gases, and the use of gasoline increases national dependence on foreign oil. Recall Hayek's acknowledgement: "Nor can certain harmful effects of deforestation, of some methods of farming, or of the smoke and noise of factories, be confined to the owner of the property in question or to those who are willing to submit to the damage for an agreed compensation. In such instances we must find some substitute for the regulation by the price mechanism" (Hayek [1944], 2007, p. 87). With respect to greenhouse gases, we are dealing with something akin to "the smoke and noise of factories."

But what is the right substitute? This is a question of institutional design. On broadly Hayekian grounds, some kind of cap-and-trade system or corrective tax is the best response, designed to ensure that drivers internalize the social costs of their activity. The reason is largely epistemic: as compared to regulatory mandates, cap-and-trade systems and corrective taxes allow for competition and leave the private sector with a great deal of flexibility and a capacity for plenty of innovation (Potts, 2018). They can find their own means to the chosen goal, which means that economic incentives of some kind, and not mandates, are the appropriate instrument (Stewart & Wiener, 2003). They can compete over means, and competition is a discovery procedure (Hayek, 2014b). For any given reduction in pollution levels, incentives impose a lower cost (Ackerman & Stewart, 1985). They preserve the price signal and the marvel that is the market. They allow competition to serve as a discovery procedure. The choice between cap-and-trade and carbon taxes raises a host of important questions, but they are not at issue here.[8]

For obvious reasons, a great deal of recent analysis has been focused on greenhouse gas emissions and how best to reduce them (Nordhaus, 2015). In principle, regulatory institutions have a host of options. They might create subsidies (say, for electric cars). They might use nudges (say, by providing information about greenhouse gas emissions on fuel economy labels) (Sunstein & Reisch, 2014). They might impose regulatory mandates (say, with fuel economy and energy efficiency standards). Careful analysis suggests that carbon taxes can produce reductions in greenhouse gas emissions at a small fraction of the cost of fuel economy mandates (Karplus et al., 2013; Davis & Knittel, 2019; Knittel, 2019). On one account, "a fuel economy standard is shown to be at least six to fourteen times less cost effective than a price

[8] For a defense of carbon taxes, see Nordhaus (2015).

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

instrument (fuel tax) when targeting an identical reduction in cumulative gasoline use" (Karplus et al., 2013).

These are points about how best to reduce externalities, and on Hayekian grounds, they seem decisive. But behaviorally informed regulators focus on consumer welfare, not only externalities. They are concerned about a different kind of market failure. They speculate that at the time of purchase, many consumers might not give sufficient attention to the costs of driving a car (Bubb & Pildes, 2014). Even if they try, they might not have a sufficient understanding of those costs, because it is not simple to translate differences in miles per gallon (MPG) into economic and environmental consequences (Larrick & Soll, 2008). An obvious response, preserving freedom of choice and in a more Hayekian spirit, would be disclosure, in the form of a fuel economy label that would correct that kind of behavioral market failure (EPA, 2015). In principle, such a label, if behaviorally informed, should solve the problem, and it avoids coercion. Consider Hayek's words on the latter topic, again sounding quite Kantian (Hayek [1960], 2011, p. 71):

> By "coercion" we mean such control of the environment or circumstances of a person by another that, in order to avoid greater evil, he is forced to act not according to a plan of his own but to serve the ends of another. Except in the sense of choosing the lesser evil in a situation forced on him by another, he is unable either to use his own intelligence or knowledge or to follow his own aims and beliefs. Coercion is evil precisely because it thus eliminates an individual as a thinking and valuing person and makes him a bare tool in the achievement of the ends of another.

Labels do not do that. They do not coerce, and they also do not discriminate (a particular problem for Hayek, who opposed "commands," which do discriminate). They allow for the use of dispersed information. They do not tinker with prices or quantities. In short: labels should be used to promote consumer welfare, by increasing the likelihood that consumers will make optimal choices, and corrective taxes should be used to respond to externalities. This might even be seen as a broadly Hayekian approach to the problem.

Such a label might, for example, draw consumers' attention to the long-term costs or savings associated with motor vehicles. Focused on internalities, the existing label in the United States does exactly that; it specifies annual fuel costs and also five-year costs (or savings) compared to the average vehicle. Use of such a label has evident appeal, and it might well be right. But there are concerns. Once more, one set of them is unquestionably Hayekian: Who is producing the label, what do they know, and will companies commit fraud? We need to have some confidence that policymakers know what they are doing.

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

There have been incidents of fraud, above all from Volkswagen, and we might fear that they are more pervasive than we know.

It would also be possible to wonder whether a label will be sufficiently effective; this is an empirical question, not resolvable in the abstract. Perhaps some or many consumers will pay too little attention to the label, and hence will not purchase cars that would save them a significant amount of money (Bubb & Pildes, 2014). And if some or many consumers are genuinely inattentive to the costs of operating a vehicle (at the time of purchase), then it is possible that fuel economy standards, with a level of stringency that would be difficult to defend on standard economic grounds, might turn out to be justified.

In support of that argument, let us go directly to the merits, and put Hayek to one side for a moment. It would be useful for public institutions to focus directly on two kinds of consumer savings from fuel economy standards, involving internalities rather than externalities: money and time. In fact, the vast majority of the quantified benefits from recent fuel economy standards come not from environmental improvements, but from money saved at the pump; turned into monetary equivalents, the time savings, coming from the fact that people do not have to go to gas stations, are also significant. Under the Obama Administration, at least, the Department of Transportation found consumer savings of about $529 billion; time savings of $15 billion; energy security benefits of $25 billion; carbon dioxide emissions reductions benefits of $49 billion; other air pollution benefits of about $14 billion; and less than $1 billion from reduced fatalities (NHTSA, 2012). The total projected benefits were $633 billion over fifteen years, of which a remarkable 84 percent come from savings at the pump, and no less than 86 percent from those savings along with time savings (Bento et al., 2020). Pause over that if you would. Only about 14 percent of the total savings come from reducing externalities! That should make contemporary Hayekians pretty nervous. We will see why in a moment.

In its own rulemaking, the Trump Administration rethought those numbers by reference to recent work[9] raising questions about whether consumers are insufficiently attentive to the economic savings, but it projected the consumer savings to be in the same general vicinity and actually even higher. The Biden Administration produced numbers that are broadly similar to those in predecessor administrations, in the sense that once again, the strong majority of the monetized benefits come from consumer savings.[10] The Trump Administration

[9] *See* Allcott & Knittel (2019); Sallee et al. (2016); Busse et al. (2013).

[10] *See* Revised 2023 and Later Model Year Light-Duty Vehicle Greenhouse Gas Emissions Standards, Final Rule, 86 Fed. Reg. 74500 (Dec. 30, 2021) (to be codified at 40 CFR Parts 86 and 600) https://www.govinfo.gov/content/pkg/FR-2021-12-30/pdf/2021-27854.pdf. See in particular *id.* at 74500 et seq., discussing the evaluation of consumer impacts.

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

has reoriented its analysis on the authority of recent work (Busse et al., 2013; Sallee et al., 2016; Allcott & Knittel, 2019), raising questions about whether consumers are insufficiently attentive to the economic savings.

Hayekians will have many questions and doubts here. Can we trust any of these numbers? Are they not speculative in the extreme? Perhaps public officials – planners if not socialists – just do not know enough to get the numbers close to right. A further problem, and an even more fundamental one, is that on standard economic grounds, which modern Hayekians should be inclined to support, *it is not at all clear that these consumer benefits are entitled to count in the analysis.* The reason is that they are purely private savings and do not involve externalities in any way (Gayer & Viscusi, 2013). Shouldn't government simply let people buy the cars they want? In deciding which cars to buy, consumers can certainly take into account the private savings from fuel-efficient cars. If they choose not to buy such cars, it might be because they do not value fuel efficiency as compared to other vehicle attributes (such as safety, aesthetics, and performance). Where is the problem? Where is the market failure? Isn't the government engaged in pure paternalism, a form of coercion?

It is true that the paternalism and the coercion could be worse. With a fuel economy standard, the government does not tell people: You must buy this car and not that one. Any such standard allows plenty of room for freedom of choice, for competition, and for innovation. Still, the whole fleet is required to become more fuel-efficient, which means that we are not simply relying on the market. And a main reason that we are not doing that is that we think that consumers are not making the right choices on their own.

If the problem lies in a lack of information, the standard economic prescription is the same as the behaviorally informed one: *Fix the label and provide that information so that consumers can easily understand it*. We can understand this prescription in Hayekian terms as a plea for an educative nudge (Thaler & Sunstein, 2021), rather than a more aggressive kind of intervention. And indeed, educative nudges are a key tool in the toolbox of the behaviorally informed policymaker. They might be seen as part of the same general program favored by those who prefer "boosts," understood as efforts to inform and educate people so as to promote their own agency (Hertwig, 2017). Such efforts might be seen as Hayekian in spirit.

We have seen, however, that even with the best fuel economy label in the world, consumers might turn out to be insufficiently attentive to the benefit of improved fuel economy at the time of purchase, not because they have made a rational judgment that they are outweighed by other factors, but simply because they focus on other variables, such as performance, size, and cost (Gabaix & Laibson, 2006). The problem may be one of insufficient attention

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

(Gabaix, 2017). A behavioral hunch, discussed below, is that automobile purchasers do not give adequate consideration to economic savings.[11] Apart from savings, there is the question of time: How many consumers think about time savings when they are deciding whether to buy a fuel-efficient vehicle?

Such questions raise a host of empirical issues. But if consumers are not paying enough attention to savings in terms of money and time, a suitably designed fuel economy mandate might turn out to be justified, because it would produce an outcome akin to what would be produced by consumers who are at once informed and attentive (Sunstein, 2019b). Energy efficiency requirements might be justified in similar terms, and indeed, the argument on their behalf might be stronger.[12] If the benefits of mandates greatly exceed their costs, and if there is no significant consumer welfare loss (in the form, for example, of reductions in safety, performance, or aesthetics), then the mandates would seem to serve to correct a behavioral market failure. And indeed, the US Government has so argued (EPA, 2010). The central conundrum has been referred to as the Energy Paradox in this setting (and in several others). In short, the problem is that consumers appear not to purchase products that are in their economic self-interest. There are strong theoretical reasons why this might be so:

- Consumers might be myopic and hence undervalue the long-term.
- Consumers might lack information or a full appreciation of information even when it is presented.
- Consumers might be especially averse to the short-term losses associated with the higher prices of energy-efficient products relative to the uncertain future fuel savings, even if the expected present value of those fuel savings exceeds the cost (the behavioral phenomenon of "loss aversion").
- Even if consumers have relevant knowledge, the benefits of energy-efficient vehicles might not be sufficiently salient to them at the time of purchase, and the lack of salience might lead consumers to neglect an attribute that it would be in their economic interest to consider.

[11] The hunch is questioned by Allcott and Knittel (2019); Sallee, West, and Fan (2016); Busse, Knittel, and Zettelmeyer (2013). The hunch is supported in Gillingham, Houde, and van Benthem (2019). A sharp, balanced discussion can be found in Graham, Wiener, and Robinson (2019), with what seems to me a prudent conclusion: "it seems that agency analysts should adopt a middle-ground position between full consumer valuation of fuel economy and no consumer valuation of fuel economy, and perform sensitivity analyses with different partial degrees of consumer valuation."

[12] For suggestive evidence, see Newell and Siikamäki (2015). Note that the miles per gallon measure is hardly hidden, and there is nothing quite as salient for energy efficiency.

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

– In the case of vehicle fuel efficiency, and perhaps as a result of one or more of the foregoing factors, consumers may have relatively few choices to purchase vehicles with greater fuel economy once other characteristics, such as vehicle class, are chosen.

Of course, public institutions should be cautious before accepting a behavioral argument on behalf of mandates or bans. Hayek himself was highly skeptical about reliance on the wisdom of experts: "One thing, in fact, which the work on this book has taught me is that our freedom is threatened in many fields because of the fact that we are much too ready to leave the decision to the expert or to accept too uncritically his opinion about a problem of which he knows intimately only one little aspect" (Hayek [1960], 2011, p. 50; see also Scheall, Butos, & McQuade, 2019). On this ground, there is a fair Hayekian objection to the behavioral defense of fuel economy mandates.

Behavioral biases have to be demonstrated, not simply asserted, and Hayekians might be skeptical here; as I have noted, important research suggests that consumers do pay a lot of attention to the benefits of fuel-efficient vehicles.[13] Some of that research finds that with changes in gas prices, consumers adjust their purchasing decisions, strongly suggesting that in choosing among vehicles, consumers are highly attentive to fuel economy (Busse, Knittel, & Zettelmeyer, 2013; Sallee, West, & Fan, 2016). Other research points in the same direction. It finds that when aggressive steps are taken to inform consumers of fuel economy, they do not choose different vehicles, which suggests that a lack of information, and perhaps a lack of salience, are not causal factors here (Allcott & Knittel, 2019).

On the other hand, some evidence cuts the other way. A large-scale study of actual behavior finds that after a significant correction of erroneously stated MPG measures, consumers were relatively unresponsive (Gillingham, Houde, & van Benthem, 2019). As Gillingham et al. write, "Using the implied changes in willingness-to-pay, we find that consumers act myopically: consumers are indifferent between $1 in discounted fuel costs and 15–38 cents in the vehicle purchase price when discounting at 4%." Puzzlingly, many consumers do not buy hybrid vehicles even in circumstances in which it would seem rational for them to do so (Duncan, 2019). According to the leading study, a significant number of consumers choose standard vehicles even when it would be in their economic interest to choose a hybrid vehicle, and even when it is difficult to identify some other feature of the standard vehicle that would justify their choosing it.

[13] For valuable, inconclusive discussions, see Allcott (2016); Allcott & Greenstone (2012).

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

It is also possible to think that even if consumers are responsive to changes in gasoline prices, they are still myopic with respect to choices of vehicles that have technological advances. Graham puts it crisply (Graham, Wiener, & Robinson, 2019):

> Consumers are more familiar with changes in fuel price than with changes in technology, since consumers experience fuel prices each time they refill their tank. Vehicle purchases are much less common in the consumer's experience, especially purchases that entail major changes to propulsion systems. Many consumers – excluding the limited pool of adventuresome "early adopters" – may be reticent to purchase vehicles at a premium price that are equipped with unfamiliar engines, transmissions, materials, or entirely new propulsion systems (e.g., hybrids or plug-in electric vehicles), even when such vehicles have attractive EPA fuel-economy ratings.

More broadly, the government's numbers, finding no significant consumer welfare loss from fuel economy standards, are consistent with the suggestion that consumers are suffering from some kind of behavioral bias. At the same time, the government's numbers, projecting costs and benefits, might be wrong, and a faithful Hayekian would suspect so (Gayer & Viscusi, 2013). Engineering estimates might overlook some losses that consumers will actually experience (a Hayekian worry). No one doubts that consumers have highly diverse preferences with respect to vehicles, and even though they are not mere defaults, fuel economy standards should be designed to preserve a wide space for freedom of choice (Greenstone, Sunstein, & Ori, 2020). Appropriate standards ensure that such space is maintained. Economic incentives have inherent advantages on this count.

A real question, of course, is the magnitude of net benefits from alternative approaches. As we have seen, Hayek at one point urged that it is both appropriate and necessary to ask that question. If the consumer savings are assumed to be very large, fuel economy standards are likely to have correspondingly large net benefits. To give a very rough, intuitive sense of how to think about the comparative question, let us suppose that the US government imposed an optimal carbon tax. Simply for purposes of analysis, suppose that it is $50 per ton, understood to capture the social cost of carbon.[14] Suppose that in relevant sectors, including transportation, a certain number of emitters decide to reduce their emissions, on the grounds that the cost of reducing them is (on average) $Y, which is lower than $50. The net benefit of the carbon tax would be $50 minus Y, multiplied by the tons of carbon emissions that are eliminated. It is imaginable that the resulting figure would be very high. But it is not necessarily higher than the net benefits of well-designed fuel economy standards.

[14] A great deal depends on whether a domestic or global figure is chosen. For discussion, see Gayer and Viscusi (2013); Kotchen (2018).

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

With these qualifications, an argument for fuel economy standards, made by reference to behavioral market failures and to internalities in particular, is at least plausible. In this context, nudges (in the form of an improved fuel economy label) and mandates (in the form of standards) might march hand-in-hand. It is true that if the goal is only to reduce externalities, a carbon tax is far better than a regulatory mandate. It is emphatically true that for Hayekian reasons, the best approach to internalities should not be coercion but instead appropriate disclosure, designed to promote salience and to overcome limited attention. Even if we trust the empirical findings, we might not trust regulatory institutions and their capacity to assess them (Schubert, 2017). But with an understanding of behavioral findings, a regulatory approach, promoting consumer welfare as well as reducing externalities, might turn out to have higher net benefits than the standard economic remedy of corrective taxes and disclosure.

With respect to both content and institutional design, everything turns on what the evidence shows, and on the particular numbers. But in principle, regulation of other features of motor vehicles could also be justified in behavioral terms; certain safety equipment might not be sufficiently salient to consumers at the time of purchase, and some such equipment might fall in the category of experience goods (Sunstein, 2019b). Credit markets can be analyzed similarly (Sarin, 2019). The broadest point is that while a presumption in favor of freedom of choice makes a great deal of sense (Glaeser, 2006), it is only a presumption. It might be overcome, especially when it can be shown that behavioral biases are having significant effects.

How can that be done? What are the appropriate institutional arrangements? Can an approach be adopted that might fairly claim to be Hayekian? Maybe not. In view of Hayek's skepticism about coercion, planner ignorance, and top-down expertise, the answer might ultimately turn out to be "no" (Schubert, 2017). That would be a reasonable view; it is probably right, at least if we are trying to stay close to what Hayek actually wrote.

But return to the five questions in Section 2. At least in the first instance, and possibly in the last, behaviorally informed policy ought to be based not on the preferences and values of social planners, but on learning from the choices of informed and unbiased choosers. It might well turn out to be possible to identify those choices. Once we have done so, we might be on the road toward identifying appropriate interventions, whether they involve nudges, taxes, subsidies, or mandates. True, it would be extravagant to claim that those interventions, defended by reference to people's choices under epistemically favorable conditions, are Hayekian. But it may not be extravagant to say that they are in Hayek's general spirit, and respectful of his most fundamental concerns.

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

4 The AI Calculation Debate and the Socialist Calculation Debate Are the Same Thing

Can socialism work?

The Socialist Calculation Debate is about that question (Boettke, Candela, & Truitt, 2024). Very simply, and once upon a time: Some people, including Otto Neurath and Abba Lerner, argued that central planners could dispense with the price signal and set prices by trial-and-error. Private property, in their view, was not necessary. Austrian economists, including and perhaps above all, Hayek, disagreed. They insisted that socialist planners cannot possibly know what they need to know in order to make socialism work. They cannot rationally set prices and wages for next year or even next month. They are irremediably ignorant. Their knowledge is hopelessly inadequate. Any judgments on their paper will be confounded by circumstances.

These claims bear on numerous questions. Can planners, socialist or not, make a system of rent control work? Can planners, socialist or not, choose maximum hours laws for truck drivers or airline pilots? Can they decide to favor electric motor vehicles, and if so, exactly how? Can they have some success in deciding what energy sources to promote over the coming decades? Some people think that the answers to these questions are "yes," or something like that. Hayek thought that the answer was clearly no. The calculations are simply not possible. Hayek won the Socialist Calculation Debate. Actually, it was a smashing victory (Boettke, Candela, & Truitt, 2024).

I have referred to a paper Hayek published in 1964 (and in various places, including Hayek, 1964 and Hayek, 1967), in which he explored the limits of prediction, with reference to the theory of evolution. Hayek emphasizes that Darwin's theory points to a process or mechanism that need not have produced the same organisms that we observe on earth. The theory of evolution describes "a range of possibilities," one that is extremely wide. And even if we knew (as we do not) everything about (1) the mechanism of mutation, (2) the circumstances in which particular mutations would appear, and (3) the precise advantages that any mutation would confer, we still would not be able "to explain why the existing species or organisms have the particular structures which they possess, nor to predict what new forms will spring from them" (Hayek, 1967, p. 33). This is a striking claim, and it is not intuitive.

Hayek explains that the reason for our ignorance is "the actual impossibility of ascertaining the particular circumstances which, in the course of two billion years, have decided the emergence of the existing forms, or even those which, during the next few hundred years, will determine the selection of the types which will survive" (Hayek, 1967, p. 33). The number of relevant facts is

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

simply too large. It is not possible to insert them into some formula that could then spit out some predictions. Note that Hayek's claim is very strong. He refers to an "actual impossibility," thus suggesting that the problem cannot be solved.

In Hayek's account, complex social phenomena have the same characteristics. In the social domain, "individual events regularly depend on so many concrete circumstances that we shall never in fact be in a position to ascertain them all; and that in consequence the ideal of prediction and control must largely remain beyond our reach." To drive the point home, Hayek observes that "almost any event in the course of a man's life may have some effect on almost any of his future actions," which "makes it impossible" for us to "translate our theoretical knowledge into predictions of specific events." Hayek acknowledges that the advances of science have produced a great deal of exuberance, but he is exuberant in his own way in offering this conclusion: "It is high time, however, that we take our ignorance more seriously" (Hayek, 1967, pp. 34, 39).

I am about to flip a coin. Can you predict whether it will be heads or tails?

You might have gotten it right. But if so, you were lucky. Whether it would come up heads or tails depends on an assortment of factors that could not be identified before I flipped the coin.

In principle, of course, reliable predictions could be made; if you knew everything relevant to whether a coin will come up heads or tails, you could of course make an accurate prediction. So let's not talk about randomness. The problem is the absence of essential information. It is our ignorance. No wonder that the Apostle of Ignorance drew attention to that problem.

4.1 Sampling on the Dependent Variable

A number of years ago, a brilliant law student – let's call her Jane – came to my office with an intriguing research project. She wanted to study the sources of success. Jane's plan was to contact dozens of spectacularly successful people in multiple fields (business, politics, music, literature) and to see what they had in common. Maybe all of them had difficult childhoods. Maybe none of them had difficult childhoods. Maybe all of them were quick to anger. Maybe none of them was quick to anger. Maybe all of them developed a passion in high school. Maybe none of them developed a passion in high school. Maybe all of them were impatient. Maybe none of them was impatient.

Jane was energetic as well as astonishingly smart. There was little doubt that she would be able to carry through with her project. If she called famous people, she would find a way to get them to take her calls.

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

Still, something was wrong with what she had in mind, and Hayek helps to explain why. Success emerges from a complex system, and you cannot predict it, readily or at all, in advance. If you think that you can, and if you use Jane's methods to show that you can, you will have missed something essential, and you will have gotten all confused. You will not have learned about the sources of success. Not really (Sunstein, 2024). (Parenthetical note: Could you have predicted that I would be writing this Element ten years ago? Five years ago? Two years ago? I certainly could not have.)

Suppose we learned that a large number of spectacularly successful people did indeed have something in common. Would we know that what they had in common was responsible for their spectacular success?

Not at all, and Hayek helps explain why (Hayek, 1964). There might be plenty of people (hundreds, thousands, millions) who share that characteristic, and who did not end up spectacularly successful. The shared characteristic might not be sufficient for success. Imagine, for example, that spectacularly successful people turn out to be quick to anger. Plenty of people who are quick to anger do not succeed. Maybe they never got a chance. Maybe they got mad at the wrong person at the wrong time. Maybe they were born in poverty. Maybe they didn't have the right skin color.

If we learn that spectacularly successful people tend to be quick to anger, have we learned anything at all? Maybe not.

The problem with Jane's project has a name: *sampling on the dependent variable*. Countless successful business books follow a path identical to that proposed by Jane. They try to figure out what characteristics are shared by inventors, innovators, leaders, or other successful types. If they find a shared characteristic, they urge that they have discovered a secret or clue of some kind. Maybe so. But maybe not. Probably not. This conclusion is very much in line with Hayek's arguments about scientism. It has important implications for economics, suggesting that predictions are sometimes impossible. And if predictions are impossible, policymakers will be groping in the dark.

4.2 Challenges

Could AI have predicted in (say) 2006 that Barack Hussein Obama would be elected president of the United States in 2008? Or could AI have predicted in (say) 2014 that Donald Trump would be elected president of the United States in both 2016 and 2024? Could AI have predicted in (say) 2005 that Taylor Swift would become a worldwide sensation? The answer to all of these questions is obvious: "No." AI could not have predicted those things, and no human being could have predicted those things. But why?

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

To test your answer, here are five challenges:

(1) Consider the question whether two people are going to fall in love. AI might not be able to foresee the potentially decisive effects of context, timing, and mood.
(2) Consider the question whether a song will become a big hit. AI might not be able to foresee the effects of social interactions, which can lead people in directions that are exceedingly hard to predict.
(3) Consider the question whether a social movement, on the left or right, will arise in a specified month or year. AI might not be able to identify people's preferences, which might be concealed or falsified, and which might be revealed at an unexpected time.
(4) Consider the question whether fossil fuels will be phased out by 2048. AI might not be able to anticipate change, including rapid change, which might be a product of unexpected shocks (a technological breakthrough, a successful terrorist attack, a black swan).
(5) Consider the question whether a new start-up will do well. AI might not have local knowledge, or knowledge about what is currently happening or likely to happen on the ground.

The Hayekian argument, in short: We need to take AI's ignorance more seriously. The Socialist Calculation Debate and the AI Calculation Debate are the same thing. Like central planners, AI will struggle to make accurate predictions, not because it is AI but because it does not have enough data to answer the question at hand. Those cases often, though not always, involve complex systems. (Hayek appears to have missed the "though not always" part; return to the case of coin flips.) Predictions about romantic attraction, about the success of cultural products, and about coming revolutions are cases in point. They help to confirm Hayek's claims about the ball game (see Section 1) and about why prediction is so often doomed in principle.

4.3 Life Trajectories

In 2020, a large team of researchers – 112, to be exact – engaged in an unusually ambitious project. They wanted to see if life trajectories could be predicted. To do that, they challenged the world. Their challenge had a simple name: The Fragile Families Challenge (Salganik et al., 2020).[15]

[15] Matthew Salganik et al., *Measuring the predictability of life outcomes with a scientific mass collaboration*, 117 PNAS (2020), available at https://www.pnas.org/cgi/doi/10.1073/pnas.1915006117.

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

The challenge began with an extraordinary dataset, known as the Fragile Families and Child Wellbeing Study, which was specifically created in order to enable social science research. That study, which is ongoing, offers massive amounts of data about thousands of families, all with unmarried parents. Each of the mothers gave birth to a child in a large city in the United States around 2000. The data were collected in six "waves," at birth and at the ages of one, three, five, nine, and fifteen. Each collection produced a great deal of information, involving child health and development, demographic characteristics, education, income, employment, relationships with extended kin, father–mother relationships, and much more. Some of the data were collected by asking a battery of questions to both the mother and the father. Some of it came from an in-home assessment (at ages three, five, and nine) that included measurements of height and weight, observations of the neighborhood and home, and various tests of vocabulary and reading comprehension. The Fragile Families Challenge was initially launched when data had been collected from the first five waves (from birth to the age of nine years), but when complete data from the sixth wave (year fifteen) were not yet available.

That was a terrific advantage, because it allowed the researchers to create the Challenge, which was to predict the following outcomes:

(1) Child grade point average
(2) Child grit (determined by a self-reported measure that includes perseverance)
(3) Household eviction
(4) Household material hardship
(5) Layoff of the primary caregiver
(6) Participation in job training by the primary caregiver

Those who took the challenge were given access to background material from the first five waves, and also to data on one-half of the families from the sixth wave. The material contained data on a total of 4,262 families, with a whopping 12,942 variables about each family. The central task was to build a model, based on the data that were available, that would predict outcomes during the sixth wave for those families whom data were not available.

The researchers sought to recruit a large number of participants in the Fragile Families Challenge. They succeeded. In the end, they received 457 initial applications, which were winnowed down to 160 teams. Many of the teams used state-of-the-art machine-learning methods, explicitly designed to increase accuracy. The central question was simple: Which of the 160 teams would make good predictions?

The answer is: None of them! True, the machine-learning algorithms were better than random; they were not horrible. But they were not a lot better than

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

random, and for single-event outcomes – such as whether the primary caregiver had been laid off or had been in job training – they were only *slightly* better than random. The researchers conclude that "low predictive accuracy cannot easily be attributed to the limitations of any particular researcher or approach; hundreds of researchers attempted the task, and none could predict accurately."

Notwithstanding their diverse methods, the 160 teams produced predictions that were pretty close to one another – and not so good. As the researchers put it, "the submissions were much better at predicting each other than at predicting the truth." A reasonable lesson is that even with the aid of AI, we really do not understand the relationship between where families are in one year and where they will be a few years hence. Seeming to draw that lesson, the authors of the Fragile Families Challenge suggest that their results "raise questions about the absolute level of predictive performance that is possible for some life outcomes, even with a rich data set." You can learn a great deal about where someone now is in life, and still, you, or AI, might not be able to say very much at all about specific outcomes in the future. That sounds a lot like Hayek, writing in 1964.

Let us focus on unmeasured variables, which are Hayek's concern as well. Take a girl who is ten years old and learn everything you can about her: her family, her demographics, her neighborhood, her schooling, her sports. Now predict various things about her life at the age of twenty-one. Do you have much confidence in your prediction? Hayek's point: You shouldn't. The number of variables that can move a life in one direction or another is very high, and it is not possible to foresee them in advance. Someone might break a leg at a crucial moment, meet an amazing music teacher, find a new friend, hear a song on the radio on Sunday morning, or see something online or on the news that changes everything. These points are essentially Hayek's claims about why certain predictions are doomed to fail.

4.4 Love and Romance

Can AI algorithms predict whether you will fall in love with a stranger? Can AI help people to find romantic partners?

Thus far, the results on such counts are not promising. Samantha Joel and colleagues find that AI algorithms struggle to predict "the compatibility elements of human mating . . . before two people meet," even if one has a very large number of "self-report measures about traits and preferences that past researchers have identified as being relevant to mate selection" (Joel et al., 2017). Again sounding a lot like Hayek, Joel and her colleagues suggest that romantic attraction

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

may well be less like a chemical reaction with predictable elements than "like an earthquake, such that the dynamic and chaos-like processes that cause its occurrence require considerable additional scientific inquiry before prediction is realistic."

What are "dynamic and chaos-like processes"? It is worth pondering exactly what this means. Most modestly, it might mean that AI needs far more data in order to make accurate predictions – far more, at least, than is provided by self-report measures about traits and preferences. Such measures might tell us far too little about whether one person will be attracted to another. Perhaps we need more data about the relevant people, and perhaps we should focus on something other than such measures. It is possible that AI cannot make good predictions if it learns (for example) that Jane is an extrovert and that she likes football and Chinese food. It is possible that AI algorithms would do a lot better if they learned that Jane fell for John, who had certain characteristics that drew her to him, and also for Tom and Frank, who had the same characteristics. If so, perhaps she is most unlikely to fall for Fred, who has none of those characteristics, but quite likely to fall for Eric, who shares those characteristics with John, Tom, and Frank.

On this view, the right way to predict romantic attraction is to say, "if you like X and Y and Z, you will also like A and B, but not C and D!" Or perhaps we should ask whether people who are like Jane, in the relevant respects, are also drawn to Eric. Of course it would be necessary to identify the relevant respects in which people are like Jane, and that might be exceedingly challenging.

More radically, and following Hayek, we might read the findings by Joel and her colleagues to suggest that romantic attraction is not predictable by AI algorithms for a different reason: It depends on so many diverse factors, and on so many features of the particular context and the particular moment, that any form of AI will not be able to do very well in specifying the probability that Jane will fall for Eric. The reference to "dynamic and chaos-like processes" might be a shorthand way of capturing current mood, weather, location, time of day, background sounds, and an assortment of other factors that help produce a sense of romantic connection or its absence.

Think a bit, if you would, about what made you feel a romantic spark. Did someone look at you in a certain way, at a certain moment, and then boom? What led to that boom? Do you even know?

Jane might smile at a specific time at a street corner and look Eric in the eye, and Eric's heart might flutter; or Jane might not look Eric in the eye at that moment because she is distracted by something that happened in the

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

morning. Eric might say something witty as sandwiches arrive, because of something he read in the paper that morning, and that might initiate a chain of events that culminates in marriage and children. Or Jane might make a bad joke at a bad time, and Eric might think, "This really will not work." For romance, so much may depend on factors that cannot be identified in advance. This is the sense in which AI is sometimes like centralized planners: It does not have relevant information about time and place. (Again, there does not seem to be anything like the price system to replace them with.)

4.5 Careful

We do have to be careful here. AI might be able to say that there is essentially no chance that Jane will like Carl, because there are things about Carl that we know, in advance, to be deal-breakers for Jane. Jane might not be drawn to short men or to tall men; she might not be attracted to much older men or to much younger men; she might not be attracted to men. An algorithm might be able to say that there is some chance that Jane will like Bruce; there is nothing about Bruce that is a deal-breaker for her, and there are some clear positives for her. Perhaps an algorithm can specify a range of probability that Jane will fall for Bruce; perhaps the probability of a romantic connection (suitably defined) is more than 10 percent but less than 70 percent. An algorithm might be able to say that Eric is within the category of "it might well happen" for Jane, because Eric is in some sense "her type."

Perhaps an algorithm can specify a range of probability for Jane and Eric; perhaps the probability of a romantic connection (suitably defined) is more than 20 percent but less than 80 percent. The real question is whether and to what extent AI will eventually be able to do much better than that. We might speculate that the importance of particular factors – the concrete circumstances – is such that there are real limits on AI's predictive power (even if AI might be able to outperform human beings, whose own predictive power is sharply limited in this context).

The topic of romantic attraction is intriguing in itself, and on Hayekian grounds, it can be seen as overlapping with an assortment of other prediction problems: whether you will enjoy living in Paris; whether you will become friends with a coworker; whether you will like a new job; whether a pandemic will occur in the next five years; whether a recession will occur in the next six months; whether a new movie will make a specified amount of money; whether a new book will hit the bestseller list. It is generally agreed that in stable environments with fixed rules, AI algorithms, armed with a great deal of data,

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

are able to make pretty good predictions. But if the future is unlikely to be like the past, there is a real question of whether, where, and when AI algorithms will do well, or even outperform human beings.[16] One problem might be the sheer number of relevant factors, not knowable in advance, that might produce one or another outcome; this is why the case of romantic attraction has general lessons. Another problem might be an external shock or unexpected event, which might turn everything around (a technological innovation, a terrorist attack, a pandemic, or in the context of romance, a recent loss for Jane, which opens her heart). Yet another problem might be how factors interact with each other (lasagna followed by chocolate cake might spell romance). We are speaking here of the essentially unpredictable nature of many events, because of the absence of necessary data.

4.6 Revolution Is in the Air

In work that predated the rise of AI, the economist Timur Kuran wrote a deeply Hayekian essay, urging that revolutions are unpredictable by their very nature (1995). Kuran argued that an underlying problem lies in "preference falsification": People do not disclose their preferences, which means that we cannot know whether they will, in fact, be receptive to a revolutionary movement. If we do not know what people's preferences are, we will not know whether they might be willing to participate in a rebellion once the circumstances become propitious.

Kuran added that we cannot observe people's *thresholds* for joining such a movement. How many people would be willing to join when a movement is at its early stages? Who will require something like strong minority support before joining it? Kuran also noted that social interactions are critical, and they too cannot be anticipated in advance. For a revolution to occur, people must see other people saying and doing certain things at certain times. How can we know, before the fact, who will see whom, and when, and doing what? The Hayekian answer might well be that we do not have the data, which means that we cannot possibly do that.

Kuran was not writing about AI, but it is unlikely to be able to do that, either. AI will find it challenging or impossible to learn what people's preferences are, and it might not be able to learn about thresholds. Even if it could do both, it

[16] *See* Gigerenzer (2022). The treatment in this book is helpful in important ways, but is too upbeat, I think, on people's ability to make accurate predictions through the use of heuristics in circumstances of genuine uncertainty. The Socialist Calculation Debate, the AI Calculation Debate, and the Heuristics Under Uncertainty Debate should all be resolved in favor of taking ignorance really seriously. Daniel Kahneman et al. (2021) has relevant discussion, above all in chapters 11 and 12.

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

would not (to say the least) have an easy time obtaining the data that would enable it to predict social interactions, and AI might not even be able to identify their probability. In some ways, the challenge of predicting a revolution is not so different from the challenge of predicting a romantic spark.

It is true that we, and AI, might be able to learn something about when a revolution is improbable in the extreme, and also about when a revolution is at least possible. For one thing, we might be able to make at least some progress in identifying private preferences – for example, by helping people feel safe to say that they dislike the status quo, perhaps by showing sympathy with that view, or perhaps by guaranteeing anonymity. AI might well be able to help on that count. Kuran wrote before the emergence of social media platforms, which give us unprecedented opportunities to observe hitherto unobservable preferences – for example, via prompts, posts, and Google searches, which might reveal widespread dissatisfaction with the current government.

Perhaps AI can say something about probabilities, based on data of this kind. But AI will not be able to tell us a whole lot, because its knowledge of preferences and thresholds will be limited, and because it will not be able to foresee social interactions. The general analysis should not be limited to revolutions. Preference falsification, diverse thresholds, and social interactions – one or more of these are in play in many domains.

When will marriages break up? When will employees engage in some kind of revolt? When will we see something like #MeToo? When will a populist movement emerge and succeed? AI might be able to tell us something, but not nearly everything. Hayek (1964) explains why.

4.7 Hits!

Consider the question whether books, movies, or musical albums are likely to succeed. Of course we might know that a new album by Taylor Swift is likely to do well, and that a new album by a singer who is both terrible and unknown is likely to fail. A few decades ago, I was part of a rock group called Serendipity. You haven't heard of us, and we were terrible; there was no chance that we could succeed. You don't need AI to know that. But across a wide range, a great deal depends on social interactions and apparent accidents, and on who says or does what exactly when. In such circumstances, AI might not be able to help much. The point overlaps with, and might be identical to, Hayek's claims about complexity and the impossibility of prediction.

This point clearly emerges from research from a number of years ago, when Matthew Salganik, Duncan Watts, and Peter Dodds investigated the sources of cultural success and failure (2006). Their starting point was that those who sell

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

books, movies, television shows, and songs often have a great deal of trouble predicting what will succeed. Even experts make serious mistakes. Some products are far more successful than anticipated, whereas some are far less so. This seems to suggest, very simply, that those that succeed must be far better than those that do not. But if they are so much better, why are predictions so difficult? Why do the best analysts fail? No one anticipated the success of the Harry Potter series; the Beatles couldn't get a record deal; the rise of Donald Trump was a shock.

To explore the sources of cultural success and failure, Salganik and his coauthors created an artificial music market on a preexisting website. The site offered people an opportunity to hear forty-eight real but unknown songs by real but unknown bands. One song, by a band called Calefaction, was 'Trapped in an Orange Peel.' Another, by Hydraulic Sandwich, was 'Separation Anxiety.' The experimenters randomly sorted half of about 14,000 site visitors into an 'independent judgment' group, in which they were invited to listen to brief excerpts, to rate songs, and to decide whether to download them. From those 7,000 visitors, Salganik and his coauthors could obtain a clear sense of what people liked best. The other 7,000 visitors were sorted into a 'social influence' group, which was exactly the same except in just one respect: the social influence group could see how many times each song had been downloaded by other participants.

Those in the social influence group were also randomly assigned to one of eight subgroups, in which they could see only the number of downloads in their own subgroup. In those different subgroups, it was inevitable that different songs would attract different initial numbers of downloads as a result of unknown factors. For example, 'Trapped in an Orange Peel' might attract strong support from the first listeners in one subgroup, whereas it might attract no such support in another. 'Separation Anxiety' might be unpopular in its first hours in one subgroup but attract a great deal of favorable attention in another.

The research questions were simple: would the initial numbers affect where songs would end up in terms of total number of downloads? Would the initial numbers affect the ultimate rankings of the forty-eight songs? Would the eight subgroups differ in those rankings? You might hypothesize that after a period, quality would always prevail – that in this relatively simple setting, where various extraneous factors (such as reviews) were highly unlikely to be at work, the popularity of the songs, as measured by their download rankings, would be roughly the same in the independent group and in all eight of the social influence groups. (Recall that for purposes of the experiment, quality is being measured solely by reference to what happened within the control group.)

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

It is a tempting hypothesis, but that is not at all what happened! 'Trapped in an Orange Peel' could be a major hit or a miserable flop, depending on whether a lot of other people initially downloaded it and were seen to have done so. To a significant degree, everything turned on initial popularity. Almost any song could end up popular or not, depending on whether or not the first visitors liked it. Importantly, there is one qualification: the songs that did the very best in the independent judgment group rarely did very badly, and the songs that did the very worst in the independent judgment group rarely did spectacularly well. But otherwise, almost anything could happen.

The Hayekian lesson is that success and failure in cultural markets are exceedingly hard to predict, whether we are speaking of AI or human beings. It is exceedingly difficult to know, in advance, whether a cultural product will benefit from the equivalent of early downloads. True, knowing that is not unknowable in principle. If you knew everything about everything (not merely the initial conditions but also how complex systems, including interaction effects, will unfold over time), you would know that, just as if you knew everything about everything, you would know whether a specific coin will come up heads or tails if I toss it in the air in the next second, or whether Carl and Eleanor are going to fall in love if they have lunch next Tuesday, or whether there is going to be a revolution in a specific nation in February of next year. But it is not easy to know everything about everything.

Because one song could do well in one world and badly in another similar world, it cannot easily be predicted in advance whether a song will do well in some other world that concerns us. Early popularity might well be crucial, and early popularity can turn on variables that are challenging to identify in advance. Because of the sheer number of variables that can produce success or failure, AI will struggle to make successful predictions at early stages (though it can do better if it is given data in real time and on an ongoing basis, which might enable us to say, shortly before success or failure, where we are heading). And in the case of financial markets, there is a special problem: Once it is made, a prediction by AI will automatically be priced into the market, which will immediately make that prediction less reliable, and possibly not reliable at all.

4.8 Cute

What about business? What about products? Where do people want to travel? (Paris, Berlin, London, Copenhagen, Vienna, Prague, Beijing, Dublin, Amsterdam, Boston, New York?) Where do people want to study? What objects do people like

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

or not like? With respect to products, an experiment modeled on the Music Lab found the same pattern (Epstein et al., 2021). The experiment involved "Meet the Ganimals," an online platform where people can generate and curate "ganimals," which are AI-generated hybrid animals. People can also say how much they like particular ganimals and rate them in terms of cuteness, creepiness, realism, and other variables.

As in the Music Lab experiment, people were sorted into (1) groups with independent conditions, in which they made evaluations entirely on their own, and (2) groups with social influence conditions, in which they could see what other people thought. Just as in the Music Lab experiment, participants were randomly assigned to one of multiple online "worlds," each of which evolved independently of the others. Participants saw only ganimals discovered and votes cast by others in their online world, and the ranking of ganimals was based only on votes in that world.

You might think that some ganimals really are adorable and that others really are not, and that in the end, the adorable ones would be counted as adorable, and the not-adorable ones would be counted as not-adorable. But here again, social influences greatly mattered. In the social influence worlds, outcomes turned out to be more unequal and highly unpredictable. Without social influences, different groups converged in their enthusiasm toward precisely the same set of ganimal features. (If you are curious: ganimals with eyes, a head, and dog-like features.) But with social influences, groups rapidly evolved into diverse local cultures that dramatically diverged from that in the independent judgment conditions. One ganimal could be spectacularly popular in one group and essentially unknown in another. The findings were very similar to those in the Music Lab.

Shall we draw the Hayekian lesson (Caldwell, 2003)? Many markets have a lot in common with the market for ganimals. People aren't going to think that a gruesome ganimal is adorable. If you have something with eyes, a head, and dog-like features, it might be golden. But maybe not. Diverse local cultures can arise, and a fabulous product might get attention in one or more of them, and no attention at all in another. Could AI predict which products will get attention in which cultures? Maybe so. But maybe not, if social interactions, based on an assortment of factors on which data cannot be obtained in advance, turn out to be crucial. The success of Barack Obama in 2008 and Donald Trump in 2016 depended on such factors, and the same is true for the rise of Jane Austen, The Beatles, Star Wars, and Taylor Swift.[17]

[17] I discuss these issues in Sunstein (2024), and draw on some passages from that book here.

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

4.9 Knightian Uncertainty

In 1921, Frank Knight wrote: "Uncertainty must be taken in a sense radically distinct from the familiar notion of Risk, from which it has never been properly separated …. The essential fact is that 'risk' means in some cases a quantity susceptible of measurement, while at other times it is something distinctly not of this character; and there are far-reaching and crucial differences in the bearings of the phenomena depending on which of the two is really present and operating" (pp. 19–20). Knight was referring to what is now called "Knightian uncertainty": circumstances in which probabilities cannot be assigned to future events. The social theorist Jon Elster offers an example: "One could certainly elicit from a political scientist the subjective probability that he attaches to the prediction that Norway in the year 3000 will be a democracy rather than a dictatorship, but would anyone even contemplate *acting* on the basis of this numerical magnitude?" (1983, p. 199).[18]

Hayek does not appear to have said anything about Knightian uncertainty; perhaps he never engaged with the subject. But Knight and Hayek were operating along parallel tracks. Hayek's claims about prediction can be taken in Knightian terms. If we could assign probabilities to very difficult prediction problems, we would not, of course, solve those problems, but we would make a lot of progress on them. Some of Hayek's claims about prediction problems can be seen as pointing to the problem of Knightian uncertainty, and as giving reasons for thinking that it exists (Hayek, 1964).

Regulators, ordinary people, and AI are sometimes acting in situations of Knightian uncertainty (where outcomes can be identified but no probabilities can be assigned) rather than risk (where outcomes can be identified and probabilities assigned to various outcomes) (Knight, 1921).[19] Some people appear to

[18] See also Elster (1979).

[19] *See id.*; Bewley (1988); Davidson (1991). Knightian uncertainty is sometimes described as "radical uncertainty" or "deep uncertainty"; I bracket possible differences here. See Vincent Marchau et al., eds. (2019). It is also important to note that Keynes and Knight had different concerns; their differences are not relevant for my purposes here. See Dimond (2022) and Gerrard (2022), and in particular this:

> Keynes and Knight both grasped the essential difference between probability-as-risk and probability-as-uncertainty, but they travelled along vastly different roads to get there. Knight contextualised risk and uncertainty in the economic theory of profit as the reward for successful entrepreneurial action under uncertainty. The consequence of Knight's emphasis on context is that the philosophical foundations of his approach are less developed. Keynes's road was much longer, more circuitous and initially primarily concerned with the philosophical foundations, culminating in *A Treatise on Probability* before more fully contextualising his logical theory of probability in the behaviour of the economic system as a whole. The different roads followed by Keynes and Knight have had one crucial consequence. Keynes's greater emphasis on the

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

think that AI creates an uncertain risk of catastrophe, including extinction of the human race (Statement on AI Risk, 2024). Consider in this regard a passage from John Maynard Keynes (Keynes, 1937, pp. 213–214):

> By "uncertain" knowledge, let me explain, I do not mean merely to distinguish what is known for certain from what is only probable. The game of roulette is not subject, in this sense, to uncertainty; nor is the prospect of a Victory bond being drawn. Or, again, the expectation of life is only slightly uncertain. Even the weather is only moderately uncertain. The sense in which I am using the term is that in which the prospect of a European war is uncertain, or the price of copper and the rate of interest twenty years hence, or the obsolescence of a new invention, or the position of private wealth-owners in the social system in 1970. About these matters there is no scientific basis on which to form any calculable probability whatever. We simply do not know.

Sounding a lot like Knight and Hayek (his frequent adversary),[20] Keynes urged that some of the time, we cannot assign probabilities to imaginable outcomes. "We simply do not know." Keynes immediately added, however, with evident bemusement, that "the necessity for action and for decision compels us as practical men to do our best to overlook this awkward fact and to behave exactly as we should if we had behind us a good Benthamite calculation of a series of prospective advantages and disadvantages, each multiplied by its appropriate probability, waiting to be summed" (1937).

How on earth, he wondered, do we manage to do that? Keynes listed three techniques (and it is worth considering their role in AI):

(1) We assume that the present is a much more serviceable guide to the future than a candid examination of past experience would show it to have been hitherto. In other words, we largely ignore the prospect of future changes about the actual character of which we know nothing.
(2) We assume that the existing state of opinion as expressed in prices and the character of existing output is based on a correct summing up of future prospects, so that we can accept it as such unless and until something new and relevant comes into the picture.
(3) Knowing that our own individual judgment is worthless, we endeavor to fall back on the judgment of the rest of the world which is perhaps better informed. That is, we endeavor to conform with the behavior of the

philosophical issues led him ultimately to treat uncertainty as relating to the weight of argument (i.e., the evidential base), not probability *per se*, whereas Knight defined uncertainty in terms of probability (i.e., the degree of belief), not the evidential base that determined the degree of belief.

[20] On some similarities and differences, see Packard et al. (2021).

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

> majority or the average. The psychology of a society of individuals, each of whom is endeavoring to copy the others, leads to what we may strictly term a conventional judgment.

Keynes did not mean to celebrate those techniques. Actually, he thought that they were ridiculous. We might know, for example, that technological innovations have not produced horrific harm in the past, and so we might think that AI will not produce such harm in the future (strategy (1)). As good Hayekians, AI might look at the price signal to assess the risks associated with climate change (strategy (2)). AI might follow the wisdom of crowds to assess the likelihood of a pandemic (strategy (3)). But under circumstances of uncertainty, should we trust any of these? "All these pretty, polite techniques, made for a well-panelled Board Room and a nicely regulated market, are liable to collapse," Keynes urged, because and when "we know very little about the future."

Keynes emphasized the difficulty or impossibility of assigning probabilities to outcomes, but he also signaled the problem of *ignorance*, in which we are unable to specify either the probability of bad outcomes or their nature – where we do not even know the kinds or magnitudes of the harms that we are facing.[21] One reason, which should now be familiar, might be that we are dealing with a novel, unique, or nonrepeatable event. A more Hayekian reason, also familiar, might be that we are dealing with a problem involving interacting components of a system, in which we cannot know how components of the system are likely to interact with each other, which means that predictions are highly unreliable.

4.10 Back to the Future

There are some prediction problems on which AI will not do well; the reason lies in an absence of adequate data, and in a sense in what we might see as the intrinsic unpredictability of (some) human affairs. I have referred to disparate challenges, but all of them are closely connected to the knowledge problem, and in particular to the unfathomably large number of factors that account for some kinds of outcomes and the critical importance of social interactions.

In some cases, AI will be able to make progress over time. But Hayek was right. In important cases, we are dealing with complex phenomena, and the real problem is that the relevant data are simply not available in advance, which is why accurate predictions are impossible – not now, and not in the future, either. Economists, acting as policymakers or policy advisers, would do well to keep this in mind. And regulators, attempting to assess proposed courses of action,

[21] On ignorance and precaution, see Harremoës (2003). For general accounts, see Smithson (1989); Aven and Steen (2010); Giang (2015); Sunstein (2021). An invaluable resource is Arrow (1984).

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

would be well-advised to do so as well. Often the answer to a question about the likely course of events is this: *We simply do not know.*

That answer raises an assortment of puzzles about how to proceed when we cannot rule out the possibility of catastrophe. Decision theory, behavioral economics, and public policy continue to struggle with that question, and it is not clear that they have an answer (Sunstein, 2021). What they should know is what they do not and cannot know. Sometimes it is great progress to take our ignorance seriously.

5 Hayek, Mill, and Tradition

How should we think about customs and traditions? Some people think that we should be skeptical of them. Economists worry that traditions can be a product of informational cascades, in which people follow others not because they know that others are acting sensibly, but simply because the actions of others seem to carry information (Sunstein, 2019a). Cascade effects can lead people to do stupid things. We might buy products that do us no good. We might make foolish investments. Informational cascades might lead rational people to perpetuate stupid traditions (Ullmann-Margalit, 1997). Behavioral economists much worry over traditions and conformity effects. They emphasize that if people are boundedly rational, they might end up participating in the creation of shared practices, or traditions, that might not help anyone, or that might hurt a lot of people. People might be scared of things that are not dangerous. "Availability cascades" can lead to fears that are widely shared but baseless (Kuran & Sunstein, 1999).

But Hayek cherished traditions, as undesigned orders from which we greatly benefit. Of course, he accepted the idea that we might be able to improve on them, so long as we approach them with humility. But still, he cherished them. How should we think about that? Hayek did not have a simple answer to that question.

Hayek and John Stuart Mill help to define the liberal tradition, but in both temperament and orientation, they could not be further apart. Mill was a progressive, a social reformer, an optimist about change, in some ways a radical. In some ways, he was a rationalist or a constructivist, of the sort that Hayek despised. Mill believed that, properly understood, liberalism calls for significant revisions in the existing economic order, which he saw as palpably unjust: "The most powerful of all the determining circumstances is birth. The great majority are what they were born to be" (Mill, 1879, p. 31).

Hayek was not exactly a conservative; in fact, he was sharply critical of conservatism on the ground that it was largely oppositional and did not offer an

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

affirmative position. He did not think that conservatives adequately understood the nature and importance of undesigned orders, the problems with scientism, and what follows from an appreciation of the knowledge problem. We have seen that Hayek did not like coercion, including coercion of the sort that conservatives favor. Still, he is a great hero to many conservatives. One reason is that he generally venerated traditions and longstanding practices, seeing them as embodying the views and knowledge of countless people over long periods (Caldwell, 2003).

Hayek greatly admired Burke, who (it will be recalled) attacked the idea that self-styled reformers, equipped with an abstract theory, should feel free to override social practices that have stood the test of time. Hayek also admired Mandeville, who spoke of law this way (Mandeville, 1924, ii, p. 142): "We often ascribe to the excellency of man's genius, and the depth of his penetration, what is in reality owing to the length of time, and the experience of many generations." To be sure, Hayek did not think that traditions can never be changed. But he liked incremental change rather than wholesale change, and he insisted that traditions often capture the wisdom of many people, extending over generations. They are undesigned orders, and as we have seen, he liked undesigned orders a lot.

In a provocative lecture in 1983, Hayek offered, as his title, a sentence from David Hume: "The rules of morality are not the conclusions of our reason" (Hayek, 1983). By this, Hayek did not mean that the rules of morality are a product of passion rather than reason. Instead, he meant that the rules of morality are a product of an evolutionary process that we should regard with reverence. "The traditional rules of morality are," Hayek wrote, "an indispensable condition of the very existence of present mankind which we cannot alter at will to please our tastes, and which we can at most endeavor gradually to develop or improve within a framework that is given to us" (p. 1). Hayek distinguished between David Hume's view, which he shared, and the utilitarian view, "which is precisely the kind of rationalist explanation of morals" that ought to be opposed (p. 2). What is crucial is that "man has neither designed nor does understand his own morals" (p. 4). Hayek singles out "the rules which define the institutions of individual property and those of the family" (p. 5).

Those rules, Hayek thought, are a product of "selective evolution." That means that "the greater part of present humanity owes their very lives to the observation of traditional rules – rules which they did not like because they consisted of restraints on their innate instincts, and were beyond their capacity intellectually to justify" (p. 7). Hayek did not like, and meant to oppose, the views of the Enlightenment rationalists, who sought "precisely to free us from all such traditional beliefs" (p. 7). In his view, the rationalists "imagine that they

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

can invent for us a better moral which will secure for us a more pleasant, more beautiful, and more just world," but they "ignore how much we owe to traditional morals" (p. 8). Hayek rejected the belief "that it was man's intellect which enabled him to design his morals, and therefore also, where the results do not satisfy his desires, to replace them by better ones" (p. 10). This is the fatal conceit.

Still, Hayek did not favor morals legislation. He insisted that "where private practices cannot affect anybody but the voluntary adult actors, the mere dislike of what is being done by others, or even the knowledge that others harm themselves by what they do, provides no legitimate ground for coercion" (Hayek [1960], 2011, p. 212). That sounds a lot like Mill in *On Liberty*. As we have seen, Hayek thought that Mill's Harm Principle was too permissive. But he stood with Mill in rejecting the view that traditions, as such, are an adequate basis for coercion. Indeed, and strikingly, he thought that "the most conspicuous instance of" the illegitimate use of coercion "is that of the treatment of homosexuality." In his account, "private practice among adults, however abhorrent it may be to the majority, is not a proper subject for coercive action for a state whose object is to minimize coercion." That is a striking claim, especially when we consider that it was made in 1960. We might note that Hayek did not give a careful argument on behalf of this conclusion, as he did for free markets. (Was he a pro tanto rationalist?) Still, it is what he thought.

Mill had an abstract theory, one based on a conception of liberty from both government and oppressive social customs, and he thought that society could be evaluated by reference to it. Thus the beginning of *The Subjection of Women* (Mill, 1869, p. 1):

> The object of this Essay is to explain as clearly as I am able, the grounds of an opinion which I have held from the very earliest period when I had formed any opinions at all on social or political matters, and which, instead of being weakened or modified, has been constantly growing stronger by the progress of reflection and the experience of life. That the principle which regulates the existing social relations between the two sexes – the legal subordination of one sex to the other – is wrong in itself, and now one of the chief hindrances to human improvement; and that it ought to be replaced by a principle of perfect equality, admitting no power or privilege on the one side, nor disability on the other.

To that extent, Mill was indeed a rationalist in Hayek's sense. And consider these words, which Hayek could never have written (Mill, 1869, pp. 43–44): "Stupidity is much the same all the world over. A stupid person's notions and feelings may confidently be inferred from those which prevail in the circle by

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

which the person is surrounded. Not so with those whose opinions and feelings are an emanation from their own nature and faculties."

More rationalism, no doubt about it. Understanding Hayek and Mill, and what unifies and separates them, is fundamental to understanding the liberal tradition, and what contemporary Hayekianism might look like.

Against this background, there is every reason to be intrigued by a book with the title *Hayek on Mill* (Hayek [1951], 2015). Hayek died in 1992, but the University of Chicago Press is continuing to publish a multivolume edition of his collected works. Readers have been discovering essays by Hayek that were never published, were not easily available, or were not widely known. What would Hayek have to say about a great champion of liberty, in some ways his intellectual ancestor, a sharp critic of traditions who ended up embracing socialism? What would Hayek, one of the great twentieth-century liberals, say about Mill, the greatest nineteenth-century liberal?

How stunning, then, to find that Hayek offered only a few snippets on that question. His treatment, first published in 1951, grew out of an enormous, uncharacteristic, and somewhat obsessive undertaking. Startlingly, Hayek assembled what remains of the correspondence between Mill and his eventual wife, Harriet Taylor (one or the other destroyed numerous letters, probably including the most interesting), and used it as the basis for a narrative account of their mysterious love affair.

The book raises mysteries of its own. For all his greatness, Hayek was not exactly a romantic writer. Of course he celebrated the operations of free markets, but he did not display a lot of interest in the full range of human emotions. Was he passionate? In a way, about markets and the rule of law (as he understood them). Some liberals, including Mill, have a romantic streak in their writing; Hayek did not. How was it, exactly, that Hayek, of all people, became captivated by the story of John Stuart Mill and Harriet Taylor?

Here is one answer: Hayek's personal life was not free of storms, and some of the storms overlapped with Mill's, in the sense that he had a passionate romantic relationship, eventually leading to a second marriage of his own (Ebenstein, 2018; Caldwell & Klausinger, 2022). A more academic answer is that he had to explain to himself and others why Mill – one of the few thinkers he had to regard as an intellectual equal or superior – moved away from what Hayek celebrated as classical liberalism, which for Hayek was focused on limited government and protection of free markets. But Hayek's interest in the romance itself outpaced his interest in the evolution of Mill's thinking (perhaps because of the beauty and great delicacy of the correspondence). Still, Mill was a romantic liberal, seeing liberalism itself as a kind of romance; Hayek was anything but that, and he saw liberalism as a marvel, but also as anything but a kind of romance.

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

Does Mill's relationship with Harriet Taylor have anything to do with liberalism, liberty, and economic theory? I think so (Mill & Mill [1859], 2026). One of the lessons we can draw from Hayek's work of excavation is that Mill's distinctive form of liberalism, with its emphasis on individual freedom from the confining effect of social norms, owed a large debt to his relationship with Taylor. As we shall see, Hayek himself missed the connection entirely, because his own preoccupations lay elsewhere. Still, Hayek was focused on something that did indeed divide him from Mill, and it had to do with the right attitude toward customs and traditions. Mill and Hayek were, in a way, intellectual siblings, but Hayek's focus on undesigned orders, and the dispersed knowledge they reflect, was not by any means Mill's emphasis. That focus was Hayek's favorite thing; Mill's favorite thing, I think, was individual liberty as he understood it. On Hayek's grounds, Mill suffered from a kind of hubris. In the end, Mill was a kind of planner, and Hayek could not abide that.

Hayek begins with one of his central puzzles, and it involves Taylor rather than Mill: "The literary portrait which in the *Autobiography* John Stuart Mill has drawn for us of the woman who ultimately became his wife creates a strong wish to know more about her" (Hayek [1951], 2015, p. 9). Mill's own account suggests that she must have been "one of the most remarkable women who ever lived." Hayek quotes a very long passage from Mill himself (Mill, 1873, pp. 186–187, as cited in Hayek [1951], 2015):

> In general spiritual characteristics, as well as in temperament and organization, I have often compared her, as she was at this time, to Shelley: but in thought and intellect, Shelley, so far as his powers were developed in his short life, was but a child compared with what she ultimately became. Alike in the highest regions of speculation and in the smaller practical concerns of daily life, her mind was the same perfect instrument, piercing to the very heart and marrow of the matter; always seizing the essential idea or principle. The same exactness and rapidity of operation, pervading as it did her sensitive as [well as] her mental faculties, would with her gifts of feeling and imagination have fitted her to be a consummate artist, as her fiery and tender soul and her vigourous eloquence would certainly have made her a great orator, and her profound knowledge of human nature and discernment and sagacity in practical life, would in [the] times when such a *carrière* was open to women, have made her eminent among the rulers of mankind.

Mill had a lot more to say about Harriet Taylor (1859b, p. v): "Were I [but] capable of interpreting to the world one half the great thoughts and noble feelings which are buried in her grave, I should be the medium of a greater benefit to it, than is ever likely to arise from anything that I can write, unprompted and unassisted by her all but unrivalled wisdom."

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

It is somehow right, and also deeply moving, that an extraordinary edition of *On Liberty*, published in 2026, finally credits Harriet Taylor as the book's coauthor (Mill & Mill [1859], 2026). Mill would have loved that – Hayek not so much. One of Hayek's projects is to discover whether Mill's account of his wife was "sheer delusion" (Hayek [1951], 2015, p. 24).

Mill and Taylor met at a dinner in 1830, when she was just twenty-two, a mother of two boys, and married for four years to John Taylor, eleven years older than she and a junior partner in a family firm of wholesale druggists. Thomas Carlyle called him "an innocent dull good man" (p. 20). An acquaintance describes her, at the time, as "possessed of a beauty and grace quite unique of their kind," with "large dark eyes, not soft or sleepy, but with a look of quiet command in them" (Hayek [1951], 2015, p. 21). (The best words here are "grace" and "quiet command.") She wrote poetry and was soon to produce a number of essays on social usages and conventions, including one that strongly prefigured Mill's attacks, or their attacks, on conformity, decades later, in *On Liberty* (1859) (p. 22).

For his part, Mill was nothing like the dry, somewhat desiccated old man depicted in photographs. Twenty-four at the time, he must have cut a dashing figure, having been described by Carlyle as "a slender, rather tall and elegant youth," who was "remarkably gifted with precision of utterance, enthusiastic, yet lucid, calm" (p. 27). At the same time, his emotional state was not good. In a forlorn letter to a friend, written a year before meeting Taylor, he referred to "the comparative loneliness of my probable future lot," and contended poignantly that there was "now no human being . . . who acknowledges a common object with me" (p. 33–34).

In his autobiography, Mill insisted that it was not until years after meeting Taylor that their relationship "became at all intimate or confidential" (1873) (p. 37). Hardly. Referring to an article published in mid-1831, Taylor's closest friend pointedly wrote her, "Did you or Mill do it?" (p. 37). In the same year, a letter from a mutual friend, written to John Taylor, spoke mysteriously of the need for a "reconciliation" between Mr. Taylor and Mill (p. 38). In 1832, Mrs. Taylor wrote to Mill that they must not meet again, to which Mill responded in French: "*Sa route et la mienne sont séparées, elle l'a dit: mais elles peuvent, elles doivent, se rencontrer. A quelqu' époque, dans quelqu' endroit, que ce puisse être, elle me trouvera toujours ce que j'ai été, ce que je suis encore*" (p. 39). ("Her path and mine are separate, she said so: but they can, they must, come together. At whatever time, in whatever place that might be, she will find me forever as I was, as I am still." If you have ever been in a real romance, those words will speak directly to you.) A few weeks later, their relationship resumed.

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

By 1832, the two had embarked on some kind of love affair. Taylor wrote Mill: "Far from being unhappy or even *low* this morning, I feel as tho' you had never loved me half so well as last night" (Hayek [1951], 2015, p. 48). And later, in response to an apparent confession from Mill (p. 49):

> I am glad that you have said it – I am *happy* that you have – no one with any fineness & beauty of character but must feel compelled to say *all*, to the being they really *love*, or rather with any *permanent* reservation it is *not* love – while there is reservation, however little of it, the love is just *so much* imperfect . . . . *Yes* – these circumstances *do* require greater strength than any other – the greatest – that which you have, & which if you had not I should never have loved you, I should not love you now.

The Taylors agreed to a separation, and Mill and Harriet were able to spend time together. To a close friend, Mill wrote a rapturous letter (Hayek [1951], 2015, p. 51): "I am astonished when I think how much has been restrained, how much untold, unshewn and uncommunicated till now . . . . Not a day has passed without removing some real & serious obstacle to happiness . . . . There will never again I believe be any obstacle to our being together entirely."

Taylor wrote in a similar spirit, stating that "there has been so much more pain than I thought I was capable of, but also O how much more happiness" (p. 51).

Hayek writes that in the middle of the 1830s, Mill and Taylor did not try to conceal their intimacy but confronted with a great deal of malicious gossip, they withdrew almost entirely from social life (2015). In 1834, Carlyle wrote of the rumor that Mill had "fallen *desperately in love* with some young philosophic beauty" and been "lost to all his friends and to himself" (pp. 54–55). Enraged by the gossip, Mill cut off a number of his friendships. He wrote to a friend: ". . . what ought to be so much easier to me than to her, is in reality more difficult – costs harder struggle – to part company with the opinion of the world, and with my former mode of doing good in it" (p. 51). You might reasonably speculate that Mill's eventual writing on conformity, and on the importance of individuality, had some connection with his personal experiences.

It is not entirely clear what happened between Mill and Taylor from the middle 1830s to the late 1840s. What little remains of their correspondence shows a degree of agitation within and between them. Taylor to Mill: "I don't know why I was so low when you went this morning. I was *so* low – I could not bear your going my darling one; yet I should be well enough accustomed to it by now" (p. 93). The two were alert to the reactions of others; Mill seemed especially sensitive on that count. Taylor to Mill, perhaps teasingly: "I was not *quite* wrong in thinking you feared opinions. – I never supposed you dreaded the opinions of fools but only of those who are otherwise wise & good but have

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

not your opinions about Moralities" (p. 94). But there can be little doubt about the intensity of their relationship. Taylor to Mill: "When I think that I shall not hold your hand until Tuesday the time is so long & my hand so useless. Adieu my delight" (p. 103).

In 1848, Taylor returned to London after traveling with Mill, only to discover that her husband had fallen gravely ill with cancer. For a period of two months, she dedicated herself entirely to caring for him and saw Mill not at all, restricting herself to correspondence, some of it angry, even bitter, and full of dignity (Hayek [1951], 2015, p. 152): "You talk of my writing to you 'at some odd time when a change of subject of thought may be rather a relief than otherwise'! *odd time*! Indeed you must be ignorant profoundly of all that *friendship* or *anxiety* means when you can use such pitiful narrow hearted expressions."

And as her husband neared the end: "The sadness & horror of Nature's daily doings exceed a million fold all the attempts of Poets! There is nothing on earth I would not do for him & there is nothing on earth which *can* be done. Do not write" (p. 159). (Those who read Taylor's letters might be excused for falling in a little in love with her.)

In 1851, two years after her husband's death, she and Mill were married. The event must have been joyful, a kind of completion, but as Hayek reports, "the marriage led to the most painful episode in Mill's life, his complete break with his mother and her other children" (p. 167). The occasion for the break is yet another mystery. It must have had something to do with his reaction to their disapproval or their sense of scandal. But who knows? Somewhat unkindly, Hayek describes it as "almost as unintelligible to his relations as to us" (pp. 167–68). Twenty years after the break, his sister Harriet expressed genuine bafflement, reporting that "up to the time of his marriage he had been everything to us," and "it was a frightful blow to lose him at once and for ever, without [one] word of explanation, – only in evident anger" (p. 168). Anger at what, exactly?

The marriage itself was quiet, productive, and supremely happy, but both husband and wife suffered from a series of illnesses. In 1854, Mill believed himself to be gravely sick. He wrote (p. 199): "The only change I find in myself from a near view of probable death is that it makes me instinctively conservative. It makes me feel, not as I am accustomed – oh, for something better! – but oh, that we could be going on as we were before. Oh, that those I love could be spared the shock of a great change!"

But he recovered well, and it was Taylor who became desperately ill four years later. Mill wrote an appeal to a doctor: "I implore you to come immediately. I need hardly say that any expense whatever will not count for a feather in the balance" (p. 252). He never came (perhaps because the letter arrived too late).

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

Shortly after her death, Mill wrote: "It is doubtful if I shall ever be fit for anything public or private, again. The spring of my life is broken. But I shall best fulfil her wishes by not giving up the attempt to do something useful" (Hayek [1951], 2015, pp. 253–54). *On Liberty* was published in 1859 and dedicated "to the beloved and deplored memory of her who was the inspirer, and in part the author, of all that is best in my writings" (p. 256). Mill lived to 1873, and many of his greatest works, some perhaps coauthored (Mill & Mill [1859], 2026), appeared after Taylor's death.

Remarkably, Hayek ends his book saying very little about Taylor's influence on Mill's thought. But in fragments of the book, and in other essays he wrote, we can uncover Hayek's views on the mysteries with which he began. Hayek agrees that Taylor's "influence on his thought and outlook, whatever her capacities may have been, were quite as great as Mill asserts" (p. 373). At the same time, Hayek concludes, "they acted in a way somewhat different from what is commonly believed. Far from it having been the sentimental it was the rationalist element in Mill's thought which was strengthened by her influence" (p. 374).

Recall that Hayek saw rationalism, constructivism, and scientism as essentially the same thing, and it was what he spent a lifetime arguing against. In the end, he thought, Mill was on the wrong side, and it was Taylor who put him there. Recall that in Hayek's view, there is an enduring opposition between true individualism, fundamental to liberalism as he saw it, and rationalism, with the latter associated with socialism and illiberalism. By rationalism, Hayek meant to refer to what I have been emphasizing throughout: the hubristic view that with the aid of reason, human beings can reassess and plan a social order, subjecting it to the control of individual human reason, rather than relying on free markets, spontaneous orders, evolution, traditions, and the workings of the invisible hand as described by Scottish Enlightenment thinkers such as Adam Smith, a founding individualist.[22]

Rationalists, in Hayek's account, end up as collectivists. They foolishly think that human beings can effectively design rules and institutions, that "fatal conceit" that "always tends to develop into the opposite of individualism, namely, socialism or collectivism" (Hayek [1951], 2015).[23] The fatal conceit leads to illiberalism and to an obliteration of free markets. More fundamentally, it is based on what Hayek saw as the defining error of the rationalists. Mill was a liberal, and a friend of freedom, but in some respects, Hayek thought, Mill joined their ranks.

[22] The best discussion of this subject remains Ullmann-Margalit (1978).
[23] *See also* Hayek (1988).

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

As we have seen, Hayek did not believe in complete laissez-faire. He favored a guaranteed minimum for the poor ("some minimum of food, shelter, and clothing, sufficient to preserve health and the capacity to work, can be assured to everybody" ([1944], 2007, p. 148)) and even a comprehensive system of social insurance. His own conception of liberalism allowed all that. Nonetheless, he insisted that rationalism is both arrogant and dangerous, assuming that human beings know what they cannot know, and he believed that Taylor moved Mill in its direction.

To be sure, Hayek was complicated here. He celebrated traditions as undesigned orders, and he defended traditional morals. Still, he did not think that traditions should be frozen in amber. He was open to changes and improvements. Nor did he think that traditional morality was a legitimate basis for coercion. Recall that with respect to the force of law, he took an even stronger pro-liberty line than Mill did, seeing the Harm Principle as insufficiently constraining. It would be fair to wonder whether Hayek fully sorted out his enthusiasm for traditions, and for traditional morals, alongside his opposition to coercion. Perhaps the best way to do that is to say that traditional morals should be celebrated in the civil sphere, but should not be enforced by the state. That is a coherent view, and it does seem to be Hayek's own. He wrote, "Perhaps one of the most important characteristics that distinguish a free from an unfree society is indeed that, in matters of conduct that do not directly affect the protected sphere of others, the rules which are in fact observed by most are of a voluntary character and not enforced by coercion" (Hayek, 1960, p. 213).

At the same time, the prohibition on coercion "does not necessarily mean that in a free society such conduct should also be exempt from the pressure of opinion or disapproval" (id.). Here Hayek distinguished his own view from Mill's, noting that Mill "directed his heaviest attack against such 'moral coercion'" (p. 214). In Hayek's view, Mill "probably overstated the case for liberty." This is a remarkable claim, for Hayek was not wont to say that other people, or that anyone, overstated that case. Still, Hayek was right to separate himself from Mill here. Mill was keenly aware that traditions can be oppressive and stupid, and that they can stifle individuality, even if they are not enforced by law. Hayek did not have a lot to say about that. I think it is fair to say that where Hayek gave pride of place to undesigned orders, Mill put freedom at the center.

It is fair to wonder whether Hayek saw, clearly enough, that traditions may be a product of conformity pressures or cascade effects, leading to terrible things (Ullmann-Margalit, 1997; Sunstein, 2020). Many liberals follow Mill. They might respect traditions, but they insist on critical assessment of them, and they believe that traditions can be unjust or foolish. Perhaps some people have not participated enough or at all, in their creation, and not sufficiently in their

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

perpetuation. (Consider Mill's view on sex equality. Or not just people; consider animal welfare, which greatly concerned Mill, and which did not seem to concern Hayek much.) Perhaps traditions were created by some kind of accident, and then got entrenched. Perhaps they are a product of some kind of cascade, informational or reputational (Kuran & Sunstein, 1999). There are, after all, multiple equilibria. If we converged on one practice, well, maybe it is an accident (a Hayekian point), and maybe we should not be so wedded to it.

In a brilliant essay, Edna Ullmann-Margalit, evidently an admirer of Hayek, nonetheless urges against him (Ullmann-Margalit, 1997),

> Nor can it in general be assumed that [a social practice or norm] fulfills a positive function that contributes to the survival and well-being of the society incorporating it – it may, for example, promote sectarian interests, or it may lack a function (in the relevant sense) altogether. The attributes of lastingness and of overall positive functionality have to be *ascertained*, case by case, rather than presupposed.

She adds that we "need to subject each existing social institution to a critical examination, free of presuppositions, in order to see whether it was indeed of such positive social function and of such pedigree as to make it worthy of respect and preservation." That is a rationalist claim, and Ullmann-Margalit urges that Hayek was wrong not to be open to it.

In my view, Hayek was obtuse in his conclusions about Taylor's influence on Mill. There is a real irony here. The Apostle of Ignorance studied Mill and Taylor for a very long time, and he remained ignorant of something essential to his whole project. He did not get Mill or Taylor quite right (Mill & Mill [1859], 2026).

To be sure, Mill cared deeply about social justice, and he came to embrace what he described as a form of socialism, above all because of the unfairness of "the present economic order of society" (Mill, 1879). But his complex writing on that topic should hardly be seen as an endorsement of centralized government planning. Mill did not quite know what Hayek knew, or quite see what Hayek saw, and Hayek's lifelong obsessions were not Mill's cup of tea. If he were to time travel and read Hayek, I think he would nod with admiration and appreciation – but he would also have been a bit annoyed.

Where Taylor most influenced Mill was on topics that were not the subject of Hayek's main focus. Hayek missed all this. Mill's *The Subjection of Women* was profoundly affected by Taylor's views as expressed in her 1851 essay, *The Enfranchisement of Women*.[24] Taylor sketched many of Mill's central arguments, and others that were more radical still, including an

[24] There continues to be some dispute about whether Mill or Taylor was the true author of *The Enfranchisement of Women*, but the general consensus is in favor of Taylor, and hence that Mill

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

explanation of why married women should work outside the home (Mill [1851], 1868, p. 13): "A woman who contributes materially to the support of the family, cannot be treated in the same contemptuously tyrannical manner as one who, however she may toil as a domestic drudge, is a dependent on the man for subsistence."

On Liberty, now thought to be written by both Mill and Taylor (Mill & Mill [1859], 2026), is widely taken to be an argument for limited government, and so it is. It is also an attack on conformity and tradition, and this is indeed a topic that separated Hayek and Mill, and where Taylor really did have an influence. More precisely, Mill and Taylor were collaborators. It is crucial to recall here that in contending that people may be restrained only to prevent "harm to others," Mill was, and Mill and Taylor were, speaking of the effects of social norms and conventions, not merely of government (1859, p. 17). Much of Mill's attack was on the oppressive quality of public opinion.

Thus Mill writes, "the tyranny of the majority was at first, and is still vulgarly, held in dread, chiefly as operating through the acts of the public authorities." But if society "issues wrong mandates instead of right, or any mandates at all in things with which it ought not to meddle, it practices a social tyranny more formidable than many kinds of political oppression, since, though not usually upheld by such extreme penalties, it leaves fewer means of escape, penetrating much more deeply into the details of life, and enslaving the soul itself." In Mill's view, "there needs protection" not only against the state but "also against the tyranny of the prevailing opinion and feelings; against the tendency of society to impose, by other means than civil penalties, its own ideas and practices as rules of conduct on those who dissent from them." As we have seen, Hayek did not endorse these points; recall his striking view that Mill "probably overstated the case for liberty."

Taylor herself had made similar arguments more than two decades earlier, and it is hard to mistake the connections among her youthful views, Mill's own painful experiences in the 1830s, and their, and his, passionate arguments against the tyranny of custom. Far from valorizing traditions, Mill's particular case for liberty emphasized the immense importance of allowing experiments of living. In his view, "the worth of different modes of life should be proved practically, when any one thinks fit to try them. It is desirable, in short, that in things which do not primarily concern others, individuality should assert itself" (Mill, 1859b, p. 105). That idea is the beating heart of Mill's conception of liberalism.

rightly reported when he reprinted the essay that it was "hers in a peculiar sense, my share in it being little more than that of an editor and amanuensis" (Mill, 1859).

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

Here we can find the sharpest of the divergences between two of the great figures in the liberal tradition. Enthusiastic about individualism, Hayek also prized traditions and customs. He saw them with a kind of reverence, as a product of dispersed knowledge. Here, by contrast, are Mill's words (1859b, p. 105): "Where, not the person's own character, but the traditions or customs of other people are the rule of conduct, there is wanting one of the principal ingredients of human happiness, and quite the chief ingredient of individual and social progress."

This is a timeless claim, to be sure, and it is certainly not Hayekian. It is a form of rationalism (a point for Hayek's characterization of Mill here). It is also intensely autobiographical. Mill and Taylor embarked for many years on a kind of "experiment of living" that was designed to promote their own happiness despite being roundly condemned by "the traditions or customs of other people." But their individuality asserted itself. The worth of their different modes of life was proved practically. There are reasons to think that Mill's arguments against traditions have a lot of force, and that Hayek's praise of invisible hand mechanisms was a bit sloppy (Ullmann-Margalit, 1997).

All this leaves the mystery with which Hayek started: Who was Harriet Taylor? Hayek's own verdict was clear. She "was an unusual person. But the picture Mill has given us of her is throughout determined by his own character and tells us probably more of him than of her" (Hayek [1951], 2015, p. 312). To Hayek, Mill was in the grip of a delusion. Thus Hayek's conclusion (p. 311): "Behind the hard shell of complete self-control and strictly rational behavior there was [in Mill] a core of a very soft and almost feminine sensitivity, a craving for a strong person on which he could lean, and on whom he could concentrate all his affection and admiration."

Though fascinated by the scandal and the romance, Hayek rendered a cold verdict on Taylor herself. Perhaps he was right. But no one can fail to be moved by Mill's competing verdict, his inscription on her grave: "She was the sole earthly delight of those who had the happiness to belong to her . . .. Were there but a few hearts and intellects like hers this earth would already become the hoped-for heaven." We can bow humbly in Mill's direction, and ultimately accept his objections to the pressures of conformity (Sunstein, 2019a), without denigrating Hayek's arguments in favor of traditions as undesigned orders that often (not always!) serve us well.

Epilogue

Hayek's work is like an unimaginably large house, full of rooms, large and small. You enter one of them, the living room downstairs, and you learn about

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

constructivism and how wrong and dangerous it is. You leave and enter another, right next to it, and you learn about the varieties of liberalism. Then you exit and walk down the corridor, and there's a new room, a brighter one, different from the rest, and you learn something about the human mind and how it creates categories. You decide to go upstairs, and there's a small room immediately on the left, and you enter, and now you get a new understanding of the rule of law. In the massive attic, you find out, in the first room you enter, what is wrong with progressive taxation.

Right next to it, with large, comfortable chairs, is a room focused on individualism. In a big room in the large basement, you learn something new about majority rule. In the den on the first floor, there's Hayek himself, telling you, with warmth and precision, about Adam Smith, Bernard Mandeville, Ludwig von Mises, Joseph Schumpeter, John Stuart Mill, John Maynard Keynes, his cousin Ludwig Wittgenstein – and oh, about capitalism too.

In every one of the countless rooms, Hayek will tell you something about human ignorance, at least if you stay there long enough. Initially, you might find that topic a bit discouraging, verging on dispiriting, but in Hayek's hands, it is exhilarating, even thrilling. How and when does innovation work? Why are free markets such a marvel? Why is today so unfathomably different from yesterday? How is competition a discovery procedure? What's wrong with experts? Why don't we know what our world will be like a decade from now? Why do planners get it wrong, even if their motivations are good? Why is freedom so important? Why should we celebrate the rule of law?

Hayek's answers are enduring. Everyone should know them. On the socialist calculation debate, Hayek was right, and his views bear on many calculation debates, including the AI calculation debate. To be sure, behavioral economists have taught us many things that Hayek did not and could not know (even if he did know more than you might suspect, in part from reading Adam Smith, that early behavioral economist). But it would be possible to accept much of what Hayek did know, including about human ignorance, and to use behavioral findings in a way that takes Hayek on board. It is tempting to dismiss traditions. In my view, traditions deserve to be scrutinized. (Hurray for Mill; hurray for a dose of rationalism.) But Hayek's work on undesigned orders suggests the immense importance of approaching long-standing practices with humility and respect.

"It is high time," Hayek wrote, "that we take our ignorance more seriously." Hayek showed that doing that is a kind of liberation for thinkers of all varieties, and even more importantly for the human spirit.

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

References

Ackerman, B. & Stewart, R. (1985). Reforming environmental law. *Stanford Law Review*, 37(5), 1333–1365.

Adusumalli, S., Westover, J. E., Jacoby, D. S., et al. (2021). Effect of passive choice and active choice interventions in the electronic health record to cardiologists on statin prescribing: A cluster randomized clinical trial. *JAMA Cardiology*, 6(1), 40–48.

Akerlof, G. A. & Dickens, W. T. (1982). The economic consequences of cognitive dissonance. *American Economic Review*, 72(3), 307–319.

Akerlof, G. A. & Shiller, R. J. (2015). *Phishing for Phools: The Economics of Manipulation and Deception*. Princeton: Princeton University Press.

Allcott, H. (2016). Paternalism and energy efficiency: An overview. *Annual Review of Economics*, 8(1), 145–176.

Allcott, H. & Greenstone, M. (2012). Is there an energy efficiency gap? *Journal of Economic Perspectives*, 26(1), 3–28.

Allcott, H. & Knittel, C. (2019). Are consumers poorly informed about fuel economy? Evidence from two experiments. *American Economic Journal: Economic Policy*, 11(1), 1–37.

Allcott, H., Lockwood, B. B., & Taubinsky, D. (2019). Should we tax sugar-sweetened beverages? An overview of theory and evidence. *Journal of Economic Perspectives*, 33(3), 202–227.

Allcott, H. & Sunstein, C. R. (2015). Regulating internalities. *Journal of Policy Analysis and Management*, 34(3), 698–705.

Allcott, H. & Taubinsky, D. (2015). Evaluating behaviorally motivated policy: Experimental evidence from the lightbulb market. *American Economic Review*, 105(8), 2501–2538.

Arrow, K. J. (1984). *Individual Choice under Certainty and Uncertainty.* Cambridge: Belknap Press.

Aven, T. & Steen, R. (2010). The concept of ignorance in a risk assessment and risk management context. *Reliability Engineering & System Safety*, 95(11), 1117–1122.

Bar-Gill, O. (2012). *Seduction by Contract: Law, Economics, and Psychology in Consumer Markets*. Oxford: Oxford University Press.

Bento, A. M., Jacobsen, M. R., Knittel, C. R., & van Benthem, A. A. (2020). Estimating the costs and benefits of fuel-economy standards. NBER Book Chapter No. c14288. www.nber.org/books-and-chapters/environmental-

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

and-energy-policy-and-economy-volume-1/estimating-costs-and-benefits-fuel-economy-standards.

Bergh, A. (2020). Hayekian welfare states: Explaining the coexistence of economic freedom and big government. *Journal of Institutional Economics*, 16(1), 1–12.

Bernheim, B. D. (2016). The good, the bad, and the ugly: A unified approach to behavioral welfare economics. *Journal of Benefit-Cost Analysis*, 7(1), 12–68.

Bernheim, B. D. & Rangel, A. (2007). Toward choice-theoretic foundations for behavioral welfare economics. *American Economic Review*, 97(2), 464–470.

Bernheim, B. D. & Rangel, A. (2009). Beyond revealed preference: Choice-theoretic foundations for behavioral welfare economics. *Quarterly Journal of Economics*, 124(1), 51–104.

Bernheim, B. D. & Taubinsky, D. (2018). Behavioral public economics. In B. D. Bernheim, S. DellaVigna& D. Laibson, eds., Vol. I of *Handbook of Behavioral Economics: Applications and Foundations*. New York: Elsevier, pp. 381–516.

Bewley, T. F. (1988). Knightian Uncertainty. In D. P. Jacobs, E. Kalai, M. I. Kamien, and N. L. Schwartz, eds., *Frontiers of Research in Economic Theory: The Nancy L. Schwartz Memorial Lectures, 1983–1997*. Cambridge: Cambridge University Press, pp. 71–81.

Boettke, P. J. (2018). *F. A. Hayek: Economics, Political Economy and Social Philosophy*. London: Palgrave Macmillan.

Boettke, P. J., Candela, R. A., & Truitt, T. L. (2024). *The Socialist Calculation Debate*. Cambridge: Cambridge University Press.

Bohnet, I. (2016). *What Works: Gender Equality by Design*. Cambridge, MA: Belknap Press of Harvard University Press.

Bubb, R.& Pildes, R. H. (2014). How behavioral economics trims its sails and why. *Harvard Law Review*, 127(6), 1593–1678.

Burke, E. (2014 [1790]) *Revolutionary Writings: Reflections on the Revolution in France and the First Letter on a Regicide Peace*. Cambridge: Cambridge University Press.

Busse, M., Knittel, C., & Zettelmeyer, F. (2013). Are consumers myopic? Evidence from new and used car purchases. *American Economic Review*, 103(1), 220–256.

Caldwell, B. (2003). *Hayek's Challenge: An Intellectual Biography*. Chicago: University of Chicago Press.

Caldwell, B. (2008). Hayek on Mill. *History of Political Economy*, 40(4), 689–704.

Caldwell, B. & Klausinger, H. (2022). *Hayek: A Life*, 1899–1950. Chicago: University of Chicago Press.

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

Conly, S. (2013). *Against Autonomy: Justifying Coercive Paternalism*. Cambridge: Cambridge University Press.

Davidson, P. (1991). Is probability theory relevant for uncertainty? A post-Keynesian perspective. *Journal of Economic Perspectives*, 5(1), 129–143.

Davis, L. W. & Knittel, C. R. (2019). Are fuel economy standards regressive? *Journal of the Association of Environmental and Resource Economists*, 6 (S1), S37–S65.

Dimond, R. (2020). Keynes, Knight, and fundamental uncertainty: A double centenary, 1921–2021. *Review of Political Economy*, 33(4), 570–584.

Duncan, D., Ku, A. L., Julian, A., Carley, S., et al. (2019). Most consumers don't buy hybrids: Is rational choice a sufficient explanation? *Journal of Benefit-Cost Analysis*, 10(1), 1–38.

Dworkin, R. (1985). *A Matter of Principle*. Harvard University Press. https://doi.org/10.2307/j.ctv1pncpxk.

Ebeling, F. & Lotz, S. (2015). Domestic uptake of green energy promoted by opt-out tariffs. *Nature Climate Change*, 5(9), 868–871.

Ebenstein, L. (2018). Hayek's divorce and move to Chicago. *Econ Journal Watch*, 15(3), 301–321.

Elster, J. (1979). Risk, uncertainty, and nuclear power. *Social Science Information*, 18(3), 371–400.

Elster, J. (1983). *Explaining Technical Change: A Case Study in the Philosophy of Science*. Cambridge: Cambridge University Press.

Environmental Protection Agency (EPA) (2010). Light-duty vehicle greenhouse gas emission standards and corporate average fuel economy standards. www.govinfo.gov/content/pkg/FR-2010-05-07/pdf/2010-8159.pdf.

Environmental Protection Agency (EPA) (2015). Learn about the fuel economy label. www.epa.gov/greenvehicles/learn-about-fuel-economy-label.

Epstein, Z., Groh, M., Dubey, A., Pentland, A., et al. (2021). Social influence leads to the formation of diverse local trends. *Proceedings of the ACM on Human-Computer Interaction*, 5, 409.

Frantz, R. (2020). Before Kahneman and Tversky, there was Friedrich Hayek. https://cosmosandtaxis.files.wordpress.com/2020/02/frantz_ct_vol7_iss5_6-2.pdf.

Frantz, R. & Leeson, R. (2013). *Hayek and Behavioral Economics*. London: Palgrave Macmillan.

Gabaix, X. (2017). Behavioral inattention. NBER Working Paper No. w24096. Retrieved from: www.nber.org/papers/w24096.

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

Gabaix, X. & Laibson, D. (2006). Shrouded attributes, consumer myopia, and information suppression in competitive markets. *Quarterly Journal of Economics*, 121(2), 505–540.

Gayer, T. & Viscusi, W. K. (2013). Overriding consumer preferences with energy regulations. *Journal of Regulatory Economics*, 43(3), 248–264.

Gerrard, B. (2022). The road less travelled: Keynes and knight on probability and uncertainty. *Review of Political Economy*, 36(3), 1253–1278.

Giang, P. H. (2015). Decision making under uncertainty comprising complete ignorance and probability. *International Journal of Approximate Reasoning*, 62, 27–45.

Gigerenzer, G. (2022). *How to Stay Smart in a Smart World*. Cambridge, MA: MIT Press.

Gillingham, K., Houde, S. & van Benthem, A. (2019). Consumer myopia in vehicle purchases: Evidence from a natural experiment. NBER Working Paper No. w25845. www.nber.org/papers/w25845.

Glaeser, E. (2006). Paternalism and psychology. *University of Chicago Law Review*, 73(1), 133–156.

Goldin, J. (2015). Which way to nudge? Uncovering preferences in the behavioral age. *Yale Law Journal*, 125(1), 226–270.

Goldin, J. (2017). Libertarian quasi-paternalism. *Missouri Law Review*, 82(3), 669–682.

Graham, J., Wiener, J. B., & Robinson, L. A. (2019). Co-benefits, countervailing risks, and cost-benefit analysis. https://cdn1.sph.harvard.edu/wp-content/uploads/sites/1273/2019/09/Graham-Wiener-Robinson-2019.pdf.

Greenstone, M., Sunstein, C. R., & Ori, S. (2020). Fuel economy 2.0. *Harvard Environmental Law Review*, 44(1), 1–42.

Harremoës, P. (2003). Ethical aspects of scientific incertitude in environmental analysis and decision making. *Journal of Cleaner Production*, 11(7), 705–712.

Hayek, F. A. (1945). The use of knowledge in society. *American Economic Review*, 35(4), 519–530.

Hayek, F. A. (1948). *Individualism and Economic Order*. Chicago: University of Chicago Press.

Hayek, F. A. (1964). The theory of complex phenomena. In M. Bunge, ed., *The Critical Approach to Science and Philosophy*. New York: Free Press, pp. 332–359.

Hayek, F. A. (1967). *Studies in Philosophy, Politics and Economics*. University of Chicago Press.

Hayek, F. A. (1974). The pretence of knowledge. Nobel Lecture. www.nobelprize.org/prizes/economic-sciences/1974/hayek/lecture/.

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

Hayek, F. A. (1984). The origins and effects of our morals [Lecture delivered 1983]. In C. Nishiyama& K. R. Leube (Eds.), *The Essence of Hayek*. Hoover Institution Press.

Hayek, F. A. (1983). "The Rules of Morality Are Not the Conclusions of Our Reason." (unpublished manuscript).

Hayek, F. A. (1988). *The Fatal Conceit: The Errors of Socialism*. Chicago: University of Chicago Press.

Hayek, F. A. (2007). *The Collected Works of F. A. Hayek: The Road to Serfdom*, Vol. 2. Chicago: University of Chicago Press. (Original work published 1944).

Hayek, F. A. (2011). *The Collected Works of F. A. Hayek: The Constitution of Liberty*. Chicago: University of Chicago Press. (Original work published 1960).

Hayek, F. A. (2014a). *The Sensory Order: An Inquiry into the Foundations of Theoretical Psychology*. Chicago: University of Chicago Press. (Original work published 1952).

Hayek, F. A. (2014b). *The Collected Works of F. A. Hayek: The Market and Other Orders* (B. Caldwell, Ed.). Chicago: University of Chicago Press.

Hayek, F. A. (2015). *The Collected Works of F. A. Hayek: Hayek on Mill: The Mill-Taylor Friendship and Related Writings* (S. J. Peart, Ed.; Vol. 16). University of Chicago Press. (Original work published 1951).

Hayek, F. A. (2018). *New Studies in Philosophy, Politics, Economics and the History of Ideas*. University of Chicago Press. (Original work published 1978).

Hertwig, R. (2017). When to consider boosting: Some rules for policy-makers. *Behavioural Public Policy*, 1(2), 143–161.

Joel, S., Eastwick, P. W., & Finkel, E. J. (2017). Is romantic desire predictable? Machine learning applied to initial romantic attraction. *Psychological Science*, 28(10), 1478–1489.

Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). *Noise: A Flaw in Human Judgment*. New York: Little, Brown Spark.

Karplus, V. J., Paltsev, S., Babiker, M., & Reilly, J. M. (2013). Should a vehicle fuel economy standard be combined with an economy-wide greenhouse gas emissions constraint? Implications for energy and climate policy in the United States. *Energy Economics*, 36, 322–333.

Keynes, J. M. (1921). *A Treatise on Probability*. London: Macmillan.

Keynes, J. M. (1937). The general theory of employment. *The Quarterly Journal of Economics*, 51(2), 209–223.

Knight, F. H. (1921). *Risk, Uncertainty, and Profit*. Boston: Houghton Mifflin.

Knittel, C. (2019). Diary of a wimpy carbon tax. MIT CEEPR Working Paper 2019–013. http://ceepr.mit.edu/files/papers/2019-013.pdf.

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

Kotchen, M. J. (2018). Which social cost of carbon? A theoretical perspective. *Journal of the Association of Environmental and Resource Economists*, 5(3), 673–694.

Kuran, T. (1995). *Private Truths, Public Lies*. Cambridge, MA: Harvard University Press.

Kuran, T., & Sunstein, C. R. (1999). Availability cascades and risk regulation. *Stanford Law Review*, 51(4), 683–768.

Larrick, R. P. & Soll, J. B. (2008). The MPG illusion. *Science*, 320(5883), 1593–1594.

Mandeville, B. (1924). The fable of the bees (Vol. 2; F. B. Kaye, Ed.). Clarendon Press. (Original work published 1714).

Marchau, V. A., Walker, W., Bloemen, P., & Popper, S. (Eds.). (2019). *Decision Making Under Deep Uncertainty*. Cham: Springer.

Mill, H. T. (1868 [1851]). *The Enfranchisement of Women*. London: Trübner.

Mill, J. S. (1859a). *Dissertations and Discussions*. London: John W. Parker and Son.

Mill, J. S. (1859b). *On Liberty*. London: John W. Parker and Son.

Mill, J. S. (1869). *The subjection of women*. Longmans, Green, Reader, and Dyer.

Mill, J. S. (1873). *Autobiography*. London: Longmans, Green, Reader, and Dyer.

Mill, J. S. (1879). *Socialism*. Chicago: Belford, Clarke.

Mill, J. S. & Mill, H. T. (2026 [1859]), edited by Turner, P., Jacobs, E., McCabe, H., Osburg, L., Shefczyk, K., and Schmidt-Petri, C. *On Liberty*. New York: Hachette.

Milkman, K., & Thaler, R. (2016). 'Misbehaving': When psychology meets economics. *Knowledge at Wharton*. https://knowledge.wharton.upenn.edu/podcast/knowledge-at-wharton-podcast/misbehaving-psychology-meets-economics/.

Mullainathan, S., & Shafir, E. (2013). *Scarcity: Why Having Too Little Means So Much*. New York: Times Books.

National Highway Traffic Safety Administration (NHTSA) (2012). Final regulatory impact analysis: Corporate average fuel economy for MY 2017–MY 2025 passenger cars and light trucks. Retrieved from: www.nhtsa.gov/sites/nhtsa.dot.gov/files/fria_2017-2025.pdf.

Newell, R. G. & Siikamäki, J. (2015). Individual time preferences and energy efficiency. *American Economic Review*, 105(5), 196–200.

Nordhaus, W. (2015). *The Climate Casino: Risk, Uncertainty, and Economics for a Warming World*. New Haven: Yale University Press.

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

Packard, M. D., Bylund, P. L., & Clark, B. B. (2021). Keynes and Knight on uncertainty: peas in a pod or chalk and cheese? *Cambridge Journal of Economics*, 45(5), 1099–1125.

Pichert, D. & Katsikopoulos, K. V. (2008). Green defaults: Information presentation and pro-environmental behaviour. *Journal of Environmental Psychology*, 28(1), 63–73.

Potts, J. (2018). Governing the innovation commons. *Journal of Institutional Economics*, 14(6), 1025–1047.

Revised 2023 and Later Model Year Light-Duty Vehicle Greenhouse Gas Emissions Standards, Final Rule, 86 Fed. Reg. 74500 (Dec. 30, 2021) (to be codified at 40 C.F.R. Parts 86 and 600). www.govinfo.gov/content/pkg/FR-2021-12-30/pdf/2021-27854.pdf.

Rizzo, M. J. & Whitman, G. (2019). *Escaping Paternalism: Rationality, Behavioral Economics, and Public Policy.* Cambridge: Cambridge University Press.

Salganik, M., Dodds, P. S., & Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311 (5762), 854–856.

Salganik, M. J., Lundberg, I., Kindel, A. T. et al. (2020). Measuring the predictability of life outcomes with a scientific mass collaboration. *Proceedings of the National Academy of Sciences*, 117(15), 8398–8403. www.pnas.org/cgi/doi/10.1073/pnas.1915006117.

Sallee, J. M., West, S. E., & Fan, W. (2016). Do consumers recognize the value of fuel economy? Evidence from used car prices and gasoline price fluctuations. *Journal of Public Economics*, 135, 61–73.

Sarin, N. (2019). Making consumer finance work. *Columbia Law Review*, 119(6), 1519–1596.

Scheall, S., Butos, W., & McQuade, T. (2019). Social and scientific disorder as epistemic phenomena, or the consequences of government dietary guidelines. *Journal of Institutional Economics*, 15(3), 431–447.

Schubert, C. (2017). Exploring the (behavioural) political economy of nudging. *Journal of Institutional Economics*, 13(3), 499–522.

Smithson, M. (1989). *Ignorance and Uncertainty: Emerging Paradigms*. New York: Springer-Verlag.

Statement on AI Risk (2024). Center for AI Safety. www.safe.ai/work/statement-on-ai-risk.

Stewart, R. B. & Wiener, J. B. (2003). *Reconstructing Climate Policy: Beyond Kyoto*. La Vergne: AEI Press.

Sugden, R. (2019). The community of advantage. *Economic Affairs*, 39(3), 417–423.

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

Sunstein, C. R. (2019a). *Conformity*. New York: New York University Press.

Sunstein, C. R. (2019b). Rear visibility and some unresolved problems for economic analysis. *Journal of Benefit-Cost Analysis*, 10(3), 317–350.

Sunstein, C. R. (2020). Voluntary agreements. SSRN Working Paper. Retrieved from https://papers.ssrn.com/abstract=3543475.

Sunstein, C. R. (2021). *Averting Catastrophe*. New York: New York University Press.

Sunstein, C. R. (2024). *How to Become Famous*. Cambridge, MA: MIT Press.

Sunstein, C. R. (2027, forthcoming). *Who Knows Best? Mandates, Nudges, and Freedom*. New Haven, CT: Yale University Press.

Sunstein, C. R. & Reisch, L. A. (2014). Automatically green: Behavioral economics and environmental protection. *Harvard Environmental Law Review*, 38(1), 127–158.

Thaler, R. & Sunstein, C. R. (2021). *Nudge: The Final Edition*. New Haven: Yale University Press.

Ullmann-Margalit, E. (1978). Invisible-hand explanations. *Synthese*, 39(2), 263–291.

Ullmann-Margalit, E. (1997). The Invisible Hand and the Cunning of Reason. *Social Research*, 64(2), 181–198.

Vanberg, V. (2017). Editor's introduction. In V. Vanberg, ed., *The Collected Works of F. A. Hayek: The Sensory Order and Other Writings on the Foundations of Theoretical Psychology*, 1–112. Chicago: University of Chicago Press.

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

Acknowledgements

Special thanks to Peter Boettke and Bruce Caldwell for teaching me so much about Hayek – I know that I still have a lot to learn. Two anonymous reviewers offered valuable comments. I am most grateful to Hunt Allcott, Oren Bar-Gill, Tyler Cowen, Geoffrey Hodgson, and Lucia Reisch for valuable discussions. I am also grateful to Allcott for joint work that proved essential for my treatment here. Thanks too to Sarah Chalfant, my agent, for making this possible. Lexie Gagnon provided superb (and actually amazing) research assistance. Readers might be interested to know that I am finishing a new book on Hayek, to be published by MIT Press in 2028, and it will build on material here.

I am grateful for permission to draw on previous work here, including Hayekian Behavioral Economics, 7 Behavioral Public Policy 170 (2023); The Use of Algorithms in Society, 37 Review of Austrian Economics 399 (2024); John & Harriet: Still Mysterious, NY Review of Books (April 2, 2015).

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

Cambridge Elements

Austrian Economics

Peter Boettke

George Mason University

Peter Boettke is a Professor of Economics & Philosophy at George Mason University, the BB&T Professor for the Study of Capitalism, and the director of the F. A. Hayek Program for Advanced Study in Philosophy, Politics and Economics at the Mercatus Center at George Mason University.

About the Series

This series will primarily be focused on contemporary developments in the Austrian School of Economics and its relevance to the methodological and analytical debates at the frontier of social science and humanities research, and the continuing relevance of the Austrian School of Economics for the practical affairs of public policy throughout the world.

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press

Cambridge Elements

Austrian Economics

Elements in the Series

Public Debt as a Form of Public Finance: Overcoming a Category Mistake and Its Vices
Richard E. Wagner

Defense, Peace, and War Economics
Christopher J. Coyne

Cultural Considerations within Austrian Economics
Virgil Storr and Arielle John

The Origins and Consequences of Property Rights: Austrian, Public Choice, and Institutional Economics Perspectives
Colin Harris, Meina Cai, Llia Murtazashvili and Jennifer Murtazashvili

The Political Economy of Public Pensions
Eileen Norcross and Daniel J. Smith

The Political Economy of Terrorism, Counterterrorism, and the War on Terror
Anne R. Bradley, Christopher J. Coyne and Abigail R. Hall

Understanding Ludwig Lachmann's Economics
Virgil Henry Storr and Solomon M. Stein

James Buchanan and Peaceful Cooperation: From Public Finance to a Theory of Collective Action
Alain Marciano

The Socialist Calculation Debate: Theory, History, and Contemporary Relevance
Peter Boettke, Rosolino A. Candela and Tegan L. Truitt

Institutional Diversity and the Economic Calculation Debate: The Feasibility Issue Revisited
Paul Dragos Aligica and Adrian Miroiu

Clientelism
Aris Trantidis

Hayek for Our Time
Cass R. Sunstein

A full series listing is available at: www.cambridge.org/EAEC

https://doi.org/10.1017/9781047826167 Published online by Cambridge University Press