

**The Foundations of Corporate Trustworthiness:
How to Get Businesses and Organisations We Can Trust**

Matthew Thomas Clark

*Submitted in accordance with the requirements for the degree of Doctor of
Philosophy in Philosophy.*

The University of Leeds, School of Philosophy, Religion & History of Science, and
Inter-Disciplinary Ethics Applied.

December 2024

I confirm that the work submitted is my own and that appropriate credit has been given where the reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

The right of Matthew Thomas Clark to be identified as author of this work has been asserted by Matthew Thomas Clark in accordance with the Copyright, Designs and Patents Act 1988.

Acknowledgements

As Terry Pratchett (1993, p.42) puts it, 'It takes forty men with their feet on the ground to keep one man with his head in the air'. Many people have helped me, and I am thankful to them all (including those I forget to mention here).

My thinking has been helped immeasurably by discussions with attendees at conferences, especially on the core argument in Chapter 2, which has been clarified through discussion at the IDEA Centre and the Society for Applied Philosophy and Association for Social and Political Philosophy annual conferences, by engaging with the faculty at Leeds – especially those in the Grounds project – and my fellow PGRs.

I would never have considered doing a PhD without the encouragement of Dr Peter Dennis and Professor Jonathan Birch during my master's at the LSE. For their support I will always be thankful.

I can give a genuine and deep-felt thanks to my supervisors: Jamie, Robbie and Thomas. Thank you for pushing me when you needed to and supporting me when I needed it.

None of this would have been possible without the support of friends (especially those who never asked if I was done yet) and thanks specifically to Maddy Page.

Nothing I have done would have been possible without my family's support. I am always grateful for my Mum & Dad, Cara, Alys, Joseph and the rest of the family.

Finally, I would have given up on this – and much else – long ago without the encouragement of my partner, Joe. Thank you.

The Foundations of Corporate Trustworthiness - Abstract

Building trust is a key aim of businesses and policymakers. Yet crises of trust continue. The faltering dash for corporate trustworthiness raises philosophical questions which this thesis answers. What trustworthiness do we want from business? How, if at all, can something like a business be trustworthy?

I develop a new framework for understanding trustworthiness and trust.

“Trustworthiness” picks out features of agents that mitigate vulnerabilities we face as limited, social beings in specific ways which an agent brings about herself. These features take various forms, and so trustworthiness’s nature is radically pluralist.

“Trust” picks out various attitudes that attempt to track these features. Different forms of vulnerability dominate different interactions demanding different forms of trustworthiness and trust.

Our interactions with business are dominated by our reliance on them. I argue that trustworthiness which responds to others’ reliance requires a responsiveness to a novel practical obligation grounded in someone offering her reliability as a reason for action to another. To be trustworthy in this way an agent must understand herself as socially situated, reason appropriately about these obligations, and ensure that she acts on them.

I then develop an original model of organisational agency, based on a specific form of joint reliance obligations. This model shows that organisations are capable of the capacities required for reliance-responsive trustworthiness. I then argue that a business is any organisation operating in a market environment which relies on producing an economic surplus to exist and distributes that surplus to its members. I argue that the nature of “business” makes it difficult for them to be trustworthy, but that these problems can be overcome. Finally, I argue that this means we should work to increase the trustworthiness of businesses, but, where we cannot, we should look to alternative forms of regulation.

The Foundations of Corporate Trustworthiness

Contents

The Foundations of Corporate Trustworthiness - Abstract	3
Contents	5
Introduction	7
Section I: Trust, Businesses & Regulation	17
1 Why Trust in Business Matters	19
Section II: Trust and Vulnerability	29
2 The Nature of Trust and Social Vulnerabilities	31
3 Reliance and Business	63
4 Reliance-Related Obligations	79
5 Reliance-Responsive Trustworthiness	105
6 Trust and Relying	125
Section III: Trustworthy Organisations	155
7 Précis of Sections I & II	157
8 A Joint Obligation Account of Organisations	163
9 Organisations as Agents	191
10 Are Organisations Capable of Being Trustworthy?	205
11 A Trustworthy Organisation	219
12 What is a business?	241
13 A Trustworthy Business	263
14 Conclusion	277
References	283

Diagrams

Figure 1	253
Figure 2	255
Figure 3	257

Introduction

“My friends, remember this, that there are no bad herbs, and no bad men; there are only bad cultivators.”

Les Misérables, Victor Hugo (2013, p.111)¹

Philosophical research is formed from experience. The author’s ‘experience will normally have left him with particular doubts and uncertainties, and these, when pressed and probed far enough, will turn into philosophical doubts and philosophical uncertainties’ (Hampshire, 1992, p.3). Hampshire argues that when we are explicit about this, the reader can better grasp the reasons for which the argument followed its specific course.

The seeds of this thesis were sown in 2010, when I started at university. The financial crash had begun three-years earlier, but I had experienced little of it: I was a teenager in rural Yorkshire. Nevertheless, the effects of the Great Recession were formative because 2010 also marked the election in the UK of the coalition government of the Conservatives and Liberal Democrats. This led to a programme of “austerity”, which has run its course over my adult life. It has deteriorated the UK’s public services and led to the equivalent of between 190,000 and 300,000 deaths, largely of Britain’s most vulnerable people.² The justification for those choices was often framed as a necessary response to the misbehaviour of those businesses.

It seemed to me then (and still does) that we had created organisations whose behaviour was either reckless or immoral. We had let them grow powerful and allowed ourselves to become dependent on them. Yet when their misbehaviour came to a head, we did not know what to do.

In these conditions, philosophical interests took root, from which this thesis comes. I am concerned with what institutions and organisations are and how this alters what they can do. I worry about how they affect and shape the behaviour and

¹ I am aware of the quotation because it is the epigram of (*Les misérables*, 2019)

² See Berman and Hovland (2024) for the lower estimate and Walsh *et al.* (2022) for the higher estimate.

character of their members. I wonder whether we can improve how organisations behave and, if we can, what the limits to this would be.

I am especially interested in how this extends to businesses. Businesses have received less philosophical interest than their societal importance warrants. An importance which has only increased in the intervening decade and a half. The rise of “Big Tech” has produced some of the most concentrated corporate power since the break-up of the great American monopolies in the early 1900s. Arguably, the global response to the Great Recession catalysed this growth. The inflation of assets has been a consequence of the measures imposed after the financial crash, especially quantitative easing.³ This huge growth in assets has exacerbated inequality, distorted political power, and made us all far more vulnerable to corporate interests.

These concerns motivate this thesis and are why it takes the course it does.

I aim to provide an account of one way an organisation can “behave well”: by being trustworthy. I focus on businesses and trust because of a recent shift in policy and regulation. Digital innovation moves at a rate which makes it difficult for regulators to adapt. This, and other pressures, has increasingly led regulators to turn to regulation through “trust” and the production of “trustworthy businesses” or products. This thesis gives a philosophical foundation to this approach. If my arguments are correct, they suggest that while such regulation is possible, it requires significant social shifts in how businesses approach their activities.

I have tried to provide the most direct route to support that view, but getting there requires significant philosophical diversions to clarify the nature of trust and trustworthiness, and organisations and businesses. I only directly address it a couple of times.

I start by getting clear on why the trustworthiness of business should matter to us. I provide some further details on this regulatory turn. I also situate this problem in a broader set of philosophical debates. There is a way of thinking about what makes our current era distinctive, perhaps most thoroughly expressed by Max Weber’s

³ See Coppola (2019) whose argument is largely supported by evidence (including her own) provided to the House of Lords inquiry into ‘QE’ and summarised in Ch.2 of House of Lords Economics Affairs Committee (2021).

writings on bureaucracy (Weber, 2015). This stresses the reduction in interactions between individual people and the increase in interactions between individual people and institutions. This transformation is especially prevalent in business. This heightens the importance of understanding the institutions with which we increasingly interact. Philosophical research has made considerable progress in this, but its concern tends to be on the negative aspect of this debate, such as whether we can hold institutions responsible when they do wrong. This thesis contributes to the less explored positive aspect.

To know whether a business could be trustworthy, we need to answer two further questions to understand whether a business can be trustworthy. First, we need to understand what we mean when we talk about their “trustworthiness” and getting clear on this will involve getting clear on what trustworthiness might require of businesses and other entities. Second, we need to understand whether a business is the kind of thing that can meet these requirements.

I start by providing an account of what sort of trustworthiness we ought to want from businesses, and what such trustworthiness would require of a business. I begin by trying to understand what trust and trustworthiness (our “trust-concepts”) are. The current philosophy of trust literature provides little immediate help. Accounts of trust and trustworthiness are diverging, and we can often object to them with simple counterexamples. I follow Karen Jones and Thomas Simpson in thinking that this requires a change in our approach if we are to understand our trust-concepts, but I reject the approaches they have taken (Jones, 2012; Simpson, 2023, Ch.1).

My new account of our trust-concepts begins by asking what kinds of agents we are. We are limited in what we can do, we are social, and as a result we are dependent on each other. I argue that agents like us face specific types of vulnerabilities as part of our interactions. These include ways our interactions make us vulnerable to not successfully exercising our agency, and in having needs met which we have, and which we can only satisfy in interacting with others.

Gaining a clear perspective on what agents like us are like, I argue, is central to understanding our trust-concepts. This is because our trust-concepts allow us to navigate these social vulnerabilities. “Trustworthiness” picks out an important kind

of feature which other agents can have, which reduces these vulnerabilities we face in interacting with other agents in a particular “self-regulated” way. Trustworthy features matter to us because without them we face risks we can only reduce by not interacting or by expending significant resources attempting to control and direct others. “Trust” picks out various attitudes that “track” these features.

These vulnerabilities are so varied that there is no single, informative characterisation of trustworthiness which addresses them all. I suggest that instead we ought to focus our research into trustworthiness and trust by considering what feature most robustly reduces a kind of social vulnerability in a given class of interactions. (By “robustly”, I mean for the greatest range of agents, domains of action, and span of time.)

This means that to understand what form of trustworthiness we want from businesses, we need to understand what vulnerability we are most exposed to in our interactions with them. I argue that the power of many businesses and their position in contemporary society means that we are vulnerable because of our increasing reliance on them. Reliance makes us vulnerable to an action failing when the thing on which we are reliant does not happen. To clarify the nature of this vulnerability, I develop an account of reliance from the work of Matthew Noah Smith (2010). I distinguish “bare reliance” (a form of dependency relations which exist between events in the world) and “relying attitudes” (those mental states that represent them). It is our bare reliance which makes us vulnerable. You will only succeed in reading this thesis if your computer hard drive does not overheat. It does not matter what attitudes you may have about it.

If we are most vulnerable to business because of our reliance on it, then to understand their trustworthiness we need to know what feature an agent needs to reduce this vulnerability most robustly. I argue that this form of trustworthiness requires that an agent can appropriately recognise and respond to obligations which arise from others being reliant on us.

This raises the question of how such obligations arise. I argue that these obligations are a specific form of practical obligation. They come about because in certain interactions, we can offer our own reliability as a reason for action to someone else. I

argue that norms of inter-personal deliberation require that for us to suggest a reason for action to someone else, we must consider it a good reason to act in that way (at least, a good reason *for that agent* to act in that way). Our reliability, unlike other facts which ground reasons for action, is something over which we can often exercise control. This binds us to ensuring that our reliability is a good reason for action, when we invite others to become reliant on it.

I then show how appropriately reasoning about these reliance-related obligations allows us to explain many puzzling cases in the philosophy of trust. I argue that reliance-responsive trustworthiness amounts to an agent being able to appropriately recognise, reason about and act in accordance with her reliance-related obligations. So, this is what we require a business to be able to do to be trustworthy.

This does not allow us to answer whether a business can be trustworthy, because we still do not understand whether a business is the kind of thing that can be an agent which can appropriately recognise, reason about and act in accordance with its reliance-related obligations.

So, we turn to this problem. Eventually, I argue for a descriptive account of “business” as a kind of organisation, which requires a financial surplus to continue to exist and which redistributes this surplus to its members. (To be clear, a surplus is not identical to a profit, and certain non-profit enterprises are businesses on this account.)

This account requires us to be clear about what an “organisation” is. So, I first offer an account of organisations. This builds on Margaret Gilbert’s pioneering account of a “joint commitment”, but I attempt to avoid the unclear nature of the normativity which this account invokes (Gilbert, 1992). I start with the fact that we seek ways of answering practical questions, including practical questions about how some of us can come together to carry out some action which we could not complete as individuals. One way in which we can answer questions like “What should we do to X?” is by *organising*. I argue that what it takes for two (or more) parties to organise, is for each of them to offer their reliability to the other as a reason for action. This means that, following my earlier arguments, a reliance-responsive obligation arises. When we organise, we offer this reliability in a specific way. First, we offer it

mutually. That is, for the proposed action to be carried out each member is reliant on her own reliability and the reliability of each other member. It is the total set of reliances which is necessary to complete the proposed action. I argue that this means that each of the parties subject to the obligation uniquely share in it, because it is only fulfillable through their acting *together*. It is, in this sense, a *joint obligation*. Second, each party offers her reliability in a specific way. Each offers to reliably follow a *decision-making structure*. A decision-making structure is the range of role-requirements and specific procedures which allow members to determine what each one of them should do, if they are all to do something together. An organisation is then constituted by this joint obligation to follow a decision-making structure.

I then argue that we should understand such organisations as agents because they can appropriately be taken to act and to do so for their own reasons.

I suggest that although we do not commonly conceive of our involvement in organisations in this way, the model above captures a common form of collective agent. I then show how this model allows an organisational agent to have the features required to be reliance-responsive trustworthy: the ability to appropriately recognise, reason about and act in accordance with its reliance-related obligations to others. This shows how organisations can be trustworthy.

With an understanding of organisational agents, I then specify what further features such an organisational agent would have to have were it to be a business. To do this, I build on economic descriptions of firms to provide a descriptive account of the kind of entities businesses are. This emphasises the place that economic survival plays for a business, and the necessary role that economic reasons must therefore play for it.

The role of these reasons, however, poses a problem for businesses being trustworthy. For example, it will mean that businesses will often prioritise economic reasons over their reliance-related obligations, and in doing so fail to appropriately reason about them and so fail to be trustworthy. I argue, however, that there is no philosophical reason to think that such failures are necessary. Overcoming them so that a business can be trustworthy does, however, require that businesses drastically

re-value how they make others vulnerable by inviting others to rely on them. It requires that they ensure that this evaluative perspective is difficult to change, and it requires that investors and other business leaders accept this and use their influence over a business accordingly. If we can do this, we can have trustworthy businesses.

This allows us to return to the problem of regulation with which we started.

Regulators talk of trust and of trustworthy businesses. Are they justified in doing so?

I finally argue that they can be and that some likely are, but these trustworthy businesses are unlikely to be those large and powerful businesses whose trustworthiness would matter most. Many contemporary businesses fail to be trustworthy in the required way because they do not appropriately respond to their reliance-related obligations; whether this is because they are unaware of them or because they do not recognise or distort their importance. This leads me to conclude that we ought – as citizens, employees, business leaders and regulators – to reform our current businesses to realise their trustworthiness. If we will not do so, then we ought to step back from trust-focused regulation.

That is the shape of the argument to come. It offers some answers to what an organisation is, how they need to be if we want them to be trustworthy and so – in this limited sense - behave well. Overall, I hope that this thesis suggests that what Victor Hugo observed of humanity may well be true of the organisations we constitute. Their failures are ours, or at least their failures to act trustworthily are. We could change this. If we did, the world would likely go much better for far more people. Each of us could better plan and more successfully act in a world which trustworthiness had made more reliable for all of us. The thesis indicates how we must act and why we should do so if we are to be better cultivators.

*

Outline

Below, I offer a chapter-by-chapter outline of the thesis. I hope that this will help a reader to navigate it. As I am covering a range of topics, some readers may find this outline useful to home in on elements which interest them. The Sections stand largely independently, though it is easier to understand the account of organisations and their agency in Chapter 8 if you have read the account of reliance-related obligations in Chapter 4.

I start each chapter with a similar section-by-section outline. The chapter proper starts after these, and readers who find such detailed route-planning tedious should feel free to ignore them.

In Section I, I explain why we ought to be concerned about the trustworthiness of business.

Section II answers the question of what we mean by trustworthiness and what is required to be trustworthy. In Chapter 2, I put forward a new account of our trust-concepts, which provides us with an approach to understanding specific forms of trustworthiness which is followed in the rest of the Section. In Chapter 3, I explain how the size and power of businesses (and other organisations) makes us heavily reliant on them, and this in turn makes us vulnerable to our actions and plans failing because of them. In Chapter 4 & 5, I provide an account of what is required to be trustworthy with regards to this vulnerability. In Chapter 4, I argue that inviting reliance can bring about an obligation to act reliably, and that we can be robustly responsive to others reliance by appropriately responding to these obligations. In Chapter 5, I show how this offers the best available account of the kind of trustworthiness we need. In Chapter 6, I explain what the relevant form of trust is and how we can better understand it if we better understand relying attitudes.

Section III answers the question of whether a business is the kind of thing that can meet the requirements of trustworthiness. In Chapter 7, I summarise the arguments from Section I and II. I then derive the capacities an agent would need to be trustworthy in the required way. This provides the requirements that businesses would have to meet to be trustworthy in the way we ought to want them to be. In Chapter 8, I abstract from businesses to first provide an account of organisational agency grounded in a “joint obligation”, which arises when we offer mutual reliability. With this account of organisations in hand, I show in Chapter 9 why such organisations can appropriately be taken to be agents. In Chapter 10, I show how they are agents that can meet the requirements of trustworthiness. In Chapter 11, I provide a picture of what a trustworthy organisation would be like, in part by contrasting it with an organisation that I take to have failed to be trustworthy: the Metropolitan Police. I also argue that we cannot escape our reliance-related

obligations, which provide some duty to be trustworthy, and that we as individuals may have a broader moral duty to bring about trustworthy organisations. In Chapter 12, I offer an account of what a business is, showing that it is a form of organisation, but one which faces specific economic pressures. In Chapter 13, I finally show how these pressures cause problems for the trustworthiness of business, but ones which (at least from a philosophical perspective) are not insurmountable. They can be overcome, so long as businesses – and those who constitute them – are willing to change how we approach businesses as a society and in how we understand what a business does.

In the conclusion, I return to the problem of regulation which we first met here. I argue that as things stand many businesses will not be trustworthy. We could then turn away from trust-focused regulation. I argue that organisational duties to be trustworthy and our individual duties to create such organisations means that we ought not to accept this option, except in a limited number of specific situations. We should instead work to try to realise the capability for trustworthiness that our businesses can have.

Section I: Trust, Businesses & Regulation

1 | Why Trust in Business Matters

Chapter 1 aims to clarify the motivations behind the thesis and set clear research questions. §1 explores the social transformation from interactions between individuals to interactions between individuals and institutions. §2 lays out how philosophical research has responded to this change and notes the relative scarcity of research into how institutions can behave well. §3 argues that we can better understand a shift in contemporary business regulation if we better understand whether we ought to trust businesses. This sets a core research question: Should we trust businesses? Two questions follow from this: What do we want when we say we want to be able to trust businesses, and are businesses the sorts of entities which we can appropriately trust?

*

This chapter aims to clarify the motivations behind this thesis.

1.0 Individual-institutional interactions

Modern societies are distinctive because of a shift from interactions between individual people to interactions between individuals and institutions.¹ Institutions are structured entities in our social world that are ‘an establishment, organization, or association instituted for the promotion of some object.’² We have moved from interactions with a feudal lord, a money-lender and a candlestick-maker to interactions with government agencies, banks and Big Candle Incorporated. Some institutions are the state, universities, government agencies including health agencies and regulatory bodies, charities, churches and other religious bodies, and business corporations.

¹ This shift has been noted by several philosophers considering modernity and perhaps most prominently in the work of Max Weber. It has been raised in relation to trust and cultivating trust by O’Neill (2002 p.20). Several philosophers have pointed to the shift. For it increasing the need to understand collective responsibility, see especially Scheffler (1995). Kutz (2000) – particularly Chapter 1 – discusses the rise of what he terms “I-We” problems: individual victims attempting to blame (elusive) collective actors.

² Its everyday sense as in the OED, definition 7a. In philosophy and other fields (including economics), institution is more often taken to refer to a social institution; that is, some convention or regulative principle (see OED, definition 6a). I use “institution” as a general term for any social organisation because I use “organisation” in a more limited sense for that subset of institutions that have the features listed in Chapter 8. The more philosophically familiar sense of institution, I refer to as “social institutions”.

We now live in a heavily institutionalised society. This means that many everyday interactions are directly with, or mediated by, institutional structures. So, many of our relationships with each other are too. These shifts in how – and potentially what – we interact with make it important that we understand what kind of entity or kinds of entities institutions are. We also need to understand what behaviour we can reasonably expect from institutions, how we should behave towards them, what kind of relationship we can appropriately form with them, and the impact that these institutions and their structures have on us as individuals and on our interactions with each other.

2.0 Collectives in philosophy

These shifts prompt questions, to which we do not have clear answers. When a public health agency fails to adequately prepare for a pandemic, who is to blame for such failures and what ought we to do about that? Do certain institutional structures impair our capacity to make moral decisions and what effects on the moral capacities of collectives might considering moral questions as a group have? Can an institution behave well? Can it even be good? Should we trust them? And could they ever be trustworthy? The more we probe our interactions with institutions, the less clear things become.

Philosophy can contribute to clarifying our situation. It allows us to understand the nature of institutions, their capabilities and the appropriateness of different interactions between people and institutions. Philosophical research has made considerable progress in this. The main interest in this research so far has been in the negative side of our relationship with institutions: their misbehaviour. This is a focus which seems well justified. Consider the following examples.

*The Great Recession:*³ In the run up to 2008, financial organisations offered mortgage products to people, who were unsuitable credit-risks for them. The mortgage products were financially viable for the businesses because the associated debt was split and mixed with higher-grade debts to form complex derivatives. These derivatives were packaged and sold as stand-alone financial instruments. Even those

³ This draws on Jaffer, Morris and Vines (2014) and Keeley and Love (2010)

trading these derivatives knew little about them because they were opaque. Investors initially made strong returns, which led to more and more financial organisations investing heavily in them.

Eventually some of those people who could never afford to pay their mortgages, failed to pay.

Then, the value of mortgage-backed, liability-driven investments crashed. Financial organisations who were heavily exposed to these liabilities began to collapse. The international financial system teetered. The global cost of bailouts and support in the subsequent economic slump cost trillions of US dollars. Austerity policies instituted afterwards, which politicians largely justified with reference to the economic turmoil, were independently costly.

*Volkswagen Emission Scandal:*⁴ In 2011, U.S. regulators found that the German carmaker Volkswagen had set up in-car computers that switched on emission reducing technology during vehicle testing but did not do so when being driven on the road. Many Volkswagen cars being driven across the U.S.A. were releasing up to 40 times more emissions than they claimed to when sold.

In each case, the misbehaviour is intuitively ascribable to the collective itself. The *banks* did something wrong in taking systemic risks that almost collapsed the global economy and caused widespread misery. *Volkswagen* misled regulators and the public. This is just a glimpse of the wrongdoing that we intuitively blame on institutions, whose charge sheet runs to injustices and war crimes.

Are we right, though, to make such ascriptions? Is it not those individual members' actions rather than those of the collective that bring about the wrong? Each of their faults rather than that of the institution as a whole? Are we misled by loose talk?

Philosophers have taken an interest in these questions and provided answers that embrace or are sceptical of the possibility of *collective* wrongdoing. Philosophers

⁴ For a philosophical discussion of this and its relation to blameworthiness see Collins (2022). There are several good quality summaries of the scandal, for example, Topham *et al.* (2015).

attempt to explain the nature and aptness of holding collectives and their members morally responsible.⁵

Philosophers are also increasingly interested in the more positive potential for collectives. Philosophers have carried out first-order ethical theorising about how businesses and the people who work in them ought to behave, through the large amount of work in business and professional ethics. There is also a growing body of work detailing the potential duties of collectives and potentially even their rights.⁶

There has been less philosophical attention paid to how we ought to understand collectives' abilities to behave well. This raises a worry. If collective agents do exist, are they potentially 'moral idiots' or psychopaths incapable of moral reasoning and potentially incapable of standardly preventing unethical behaviour?⁷ Is this why we see so much poor behaviour from collectives?

This thesis starts work to philosophically investigate and defend the ability of collectives to behave well.

The course that such an investigation should take is open. It could be answered by considering the general form of a "virtue" or a "duty" and providing an account of how an institution would have to be if it were to be capable of having a virtue or a duty. Such discussions are useful, but by their nature they must provide a very

⁵ Philosophers pay attention to collective moral responsibility in waves, often in response to concerns about collectives and actions attributed to them by activists, the press and politicians. An initial wave in contemporary philosophy arose immediately after World War II in response to the actions of the Nazi government – particularly in the Holocaust – and in response to denazification and the novelty of the Nuremberg Tribunals. See for example: Jaspers (1946) or the sceptical view of Lewis (1948). A second wave arose in response to the actions of the United States of America in Vietnam – see the collected papers in French (1998a). More modern interest has arisen in response to environmental degradation and climate change and from interest in decolonialisation and historical responsibilities for imperialism. Much of the focus in these discussions has been on governments and on society as a whole, though interest was raised in businesses' moral responsibility following the Great Recession – see the special issue on 'Moral Responsibility and the Financial Crisis' of *Midwest Studies in Philosophy* 42:1 (2018). A further strand has arisen internally to philosophy in response to work in the philosophy of action and social ontology on collective intentionality and agency and its implications for collective moral responsibility. For arguments in favour see Miller and Mäkelä (2005); Gilbert (2006); List and Pettit (2013); Kutz (2000); Collins (2022) and for representative sceptical concerns see McKenna (2006) and Shoemaker (2019).

⁶ See for example Collins (2019) or Wringe (2016)

⁷ A worry raised by McKenna (2006), more generally by Hedahl (2013). These concerns do not affect this account of organisational trustworthiness. See Chs.10 and 11.

general account of how an institution must be. This can reduce their usefulness in providing guides to how we ought to act in specific, actual institutions.

An alternative approach is to focus on a specific “good” feature and explain how an institution would need to be if it were to be capable of having this feature. This allows for a more detailed picture, which in its greater clarity may provide us with better guidance on how to construct such institutions and be more readily applicable to actual institutions. It is, however, less useful in (though it may still provide insights into) constructing the general account of a “good” institution.

Either approach has its advantages, and we can be guided by the further aims we have in providing an account. I propose to pursue the second route, as I believe it will be helpful to sacrifice some comprehensiveness to gain some specificity.

I propose to focus on trust and businesses. In part, this is because business has been relatively neglected in the study of collectives. The state, for example, has been a far more common object of study. Where philosophers have engaged with businesses, they have tended to focus on a single form of business: the corporation.⁸ As we will see in Section III, there are good reasons to question this focus and useful work to be done in more clearly defining the nature of business.

3.0 Trust in business

Why care about trust in business? First, because we have good *prima facie* reasons to think that trust in business should concern us. Consider the intuitively acceptable responses to the behaviour described above by financial institutions and Volkswagen. It seems right to talk of investors being let down by financial organisation because they failed to protect their investments. Savers too were let down by banks which lost their savings. Consumers also seem able to appropriately talk of being let down or betrayed when the products that they thought they had bought were not the products the companies provided. In each case the appropriate

⁸ For example; Gilbert (2008); List and Pettit (2013); Lawford-Smith (2019); Pasternak (2021). A notable exception has long been Peter French, though his focus is again the corporation; see for example French (1979). I discuss some contemporary approaches to businesses and corporations as collectives in Chapters 8 and 12. The state (or the government) has also been a common focus for work on trust in collectives including in Mark Warren’s work and in Faulkner (2018) – and for counter-arguments that we ought not to trust governments Hardin (2013).

responses of those interacting with the business seem to be being let down or being betrayed. They are not mere disappointment; the response usually taken to be typical of something we have relied on failing to happen. They are the responses typical of broken trust.⁹ Our reactions suggest that we may trust businesses; so, we should want to understand whether or not this is appropriate or a mistake. It may be that we should trust business. Perhaps, this is simply another example of our widespread anthropomorphising of institutions.¹⁰ Reactive attitudes in general can be very sloppy. We may feel indignant at a world that treats us unfairly, though society does not seem like an appropriate target for our indignation or resentment.¹¹ The task of a philosopher is to explain when these reactions are appropriate and what justifies them. To answer this, we need to know if a business can be the kind of thing that is an appropriate object of our trust.

It is important to understand the appropriateness of trust in businesses not only to find out if our trusting responses are mistaken, but also because it underlies many of our social practices with businesses.¹²

Economic theory and practice often assume that trust is possible between business organisations and between people and businesses. Economic crises are often explained as a sudden breakdown of trust.¹³

The assumption that trust in business is appropriate is increasingly dominant in the regulation of business, in corporate governance and in activist practices with business.

⁹ The tracking of trust by reference to being let down and betrayal is common to nearly all philosophical discussions of trust and originates in Annette Baier's discussion (see Baier, 1986).

¹⁰ That our responses to corporations are to inappropriately anthropomorphise them is a common feature of accounts that are sceptical of collective responsibility, where the "markers" of blame (resentment and indignation) are often taken to be inappropriate in this way.

¹¹ This has been debated, but it is unclear that "society" itself is an appropriate target of moral responsibility; even if individuals may have a responsibility to change a system that produces bad outcomes. For an alternate view see Cooper (1998).

¹² And other organisations, particularly charities and political parties.

¹³ For a discussion of this see Jaffer, Morris and Vines (2014, esp. pp.4-14). I do not agree – as will become clear in Chapter 3 and in Chapter 5 – with these writers' definition of trust for business.

Businesses are interested in developing trust in themselves. Entire businesses (like the consultancy firm Edelman) exist to help them do just that. Trust and trust-building feature prominently in corporate strategies and leadership courses and are a constant focus of business journalism. The UK Corporate Governance Code, for example, states that to ‘succeed in the long-term ... companies ... need to build and maintain successful relationships with a wide range of shareholders. These relationships will be successful and enduring if they are based on respect, trust, and mutual benefit’ and calls for corporate entities to build trust (Financial Reporting Council, 2024).

Corporate regulation and activism now often assume that it could be appropriate to trust businesses and that we know something about how to bring this about.

Traditionally business regulation has worked by providing guidelines with which regulators have expected businesses to comply. Increasingly regulators alter corporate governance trying, instead, to structure businesses so that these structures “produce” (or at least increase the likelihood of) compliant behaviour. The aim of this transformation is for the regulator to be able to step back and for the business to still act appropriately.¹⁴ This change is often framed in terms of producing trust and in ensuring trustworthiness. The goal of reducing monitoring and maintained good behaviour do appear similar to what we might expect from

¹⁴ This is particularly common in the U.S.A. whose regulatory moves are often followed globally, and is becoming increasingly prominent in the U.K. – see Anderson and Waggoner (2014). Whatever your political outlook on regulation, I note that there is likely to be an assumption of trustworthiness somewhere. The dominance of this *laissez-faire* approach among regulators up to and after the Great Recession is noted by Jaffer, Knaut and Morris (2014) favouring a view that the market was self-regulating – as those authors point out this still represented a range of nesting trusts: an ‘intellectual trust’ in certain economic experts and orthodoxies from the regulator, and in them from politicians and the public (or, perhaps, the public trust is triply nested in the politicians, regulators and experts). Those who believe the market requires regulation are still likely to require some trust in businesses or at least some “nested” organisational trustworthiness (the legislature or regulatory body – also organisations – needs to be trusted).

Within regulatory studies, internal changes such as adopting certain systems of corporate governance may also be considered a form of regulation. This is how regulation is used, for example, by Braithwaite (2011). I understand regulation to require *external* pressure ensuring some desired behaviour. I take governance – if not adopted because of requirements imposed by someone or something outside of the organisation – to be a form of *self*-regulation. I discuss the nature of self-regulation in Chapter 2 and argue that organisations are capable of it in Chapters 9 and 10.

trust relationships. Activists also increasingly do this by campaigning for certain kinds of socially responsible governance, like campaigns for Environmental & Social Governance (ESG).

As this regulatory transformation continues, and we move towards trust-focused regulation, it may not be that you think that you trust business, but someone will be trusting them on your behalf. This move is only justified if businesses are the kinds of things that we can appropriately trust.

Our social practices around business are increasingly related to trust and trustworthiness. This may just be part of a social paranoia about diminishing trust and our current state of “crisis”. Even if it is, there is usefulness in research which can help us to understand if such an approach could ever be justified, and if it could what broad shape it would have to take. Philosophy can make this kind of contribution to understanding our choices for regulation. It requires us to understand what we want when we say we want to be able to trust businesses, and whether businesses are the sorts of entity which we can appropriately trust.

So, this thesis contributes to the debate on whether collectives can behave well. It does so by focusing on a neglected kind of institution in the literature: businesses. It focuses on a specific way they could behave well – being appropriate to be trusted – and in doing so, it allows us to pursue a second aim of whether the policy turn to ‘trust-based’ regulation could be justified or whether we need to reform these trust assuming practices.

4.0 Conclusion

The aim of this chapter has been to lay the groundwork for the rest of the thesis. I have explained that the thesis is motivated by a concern about a shift within our societies from interactions between individual people to interactions between people and institutions. I aim to build on the existing philosophical research into these questions by focusing on how we ought to think about institutions ability to behave well. We saw how this problem could be addressed at one of two levels of abstraction, and that choosing between them required us to consider whether we wanted to sacrifice generalisability or detail. I have chosen to sacrifice

generalisability, so that the thesis can contribute to our understanding of whether important kinds of business regulation are justified.

This focuses our research questions from whether an institution can behave well to “Should we trust business?” Answering this requires pursuing two lines of research:

(1) What do we want when we say we want to be able to trust businesses,
and

(2) are businesses the sorts of entities which we can appropriately trust?

These are the overarching questions of Section II and Section III, respectively. We can now turn to answering them.

Section II: Trust and Vulnerability

In Section I, we saw the increasing role of trust in our regulatory systems and its increasing importance to business. This raises the question of whether we ought to trust businesses. To answer this, we need to know what it is that we want when we say we want to trust them. Then we need to know whether businesses are the kinds of things that can be how we need them to be so we can appropriately trust them.

This second section of the thesis has the goal of answering the first of these. I argue that what we ought to want from business is a form of trustworthiness, which reduces the vulnerability we face because of how reliant we are on them. To get there, we first need to understand what “trust” and “trustworthiness” (our trust-concepts) are. I develop an approach which claims that our trust-concepts allow for us to better navigate our interactions with each other in ways that reduce a specific kind of vulnerability (*social* vulnerabilities) that we face in interacting with each other. I also argue that there is no substantive form of trustworthiness which reduces all these vulnerabilities. We should instead accept that there are many forms of trustworthiness and trust. We can identify the relevant form by identifying the vulnerability which dominates a specific interaction. For example, our emotional vulnerabilities dominate many of our interactions with a romantic partner. The trustworthiness we ought to want would be one that is responsive to such needs, and the form of trust most appropriate to these interactions will be one that tracks that form of trustworthiness.

I go on to argue that our interactions with business are dominated by their relative power and the reliance that this induces in us. Adopting my approach, what we want when we trust business ought, therefore, to be a specific form of trustworthiness which reduces our vulnerability to this reliance. I argue that this is best done by developing the ability appropriately to recognise, reason about and act in accordance with a special form of obligation (a reliance-related obligation), of which I develop an account.

This leaves us at the end of this section with a clear picture of what we ought to want from a business if we are to trust them. We can then turn to seeing whether they can be trustworthy in this way, and so be appropriately trusted, in Section III.

2 | The Nature of Trust and Social Vulnerabilities

Chapter 2 aims to argue for a new general approach to understanding trust: the social vulnerabilities approach. This argues that “trustworthiness” is any feature of an agent brought about in a self-regulated way which reduces a vulnerability that another agent faces in interacting with that agent. “Trust” picks out any attitudes that track these features. §1 explains the need for a new account. §3 discusses the vulnerabilities we face as agents who are limited, interdependent and – importantly – social, and introduces a kind of vulnerability – “social vulnerabilities” – which are prominent in the lives of agents like us and which I argue in §4 are central to understanding our trust-concepts. §5 provides a summary of the social vulnerabilities account. §6 argues that we want accounts of trust and trustworthiness that relate to the dominant social vulnerability in a set of interactions. This provides the approach taken to understand the kind of trustworthiness we want from business in the rest of the thesis.

*

We have seen how businesses are increasingly dealt with in a trust-focused way. To answer the question of whether they can be trusted, we first need to understand what sort of trust we should have in them. This chapter sets out a new way of approaching trust, which will allow us to answer this question in the rest of this section.

1.0 Approaching trustworthiness and trust

1.1 An atmosphere of trust

Trust is something that ‘we are so familiar with that we scarcely notice its presence and its variety’: ‘We inhabit a climate of trust as we inhabit an atmosphere and notice it as we notice air, only when it becomes scarce or polluted’ (Baier, 1986, pp.233-4). Today, I trusted my fellow commuters not to attack me, my colleague to share her notes, and a bartender to give me the correct change. I trusted my partner in ways both trivial (to do the laundry) and serious (to stay faithful).¹ It is better to say simply, “I trusted him.” Some trust feels barely noticeable. My trust in my

¹ Baier (1986, pp.233-4) notes that trust can exist between ‘intimates’, ‘strangers’, ‘declared enemies’, ‘burglars and police’ (‘not to carry lethal weapons’), ‘those we encounter in lonely library stacks’, ‘neighbouring strangers’ on ‘trains or planes’, ‘pilots, drivers, doctors’ and ‘day-care centres’.

partner can dominate my experience. What makes this rag-rug of situations occasions of trust?

Research into trust is almost as varied, though researchers try harder to regiment trust and pin it down. Social scientists and social scientifically minded philosophers often take trust to be whatever phenomena enables cooperation between rational agents.² There has been interest in trust between citizens and political institutions.³ Philosophers of trust have tended to be more interested in the kinds of trust we find in interpersonal relationships, such as between a child and her caregiver or between friends or lovers. Such trust has been talked of as being emotional or a belief, and of being normatively laden in distinct ways (coming from the participant stance, or with commitments or obligations).⁴

It would be useful at this point if there were an account of trust which we could simply take “off the peg” and apply to businesses. The wide range of accounts, their divergence and competing appeal when considered against different experiences of trusting, and their incompatibility with each other means that this is not an available option. Instead, we need to start from the fundamentals.

² As will become clear from the discussion in the main text, I think that there are limits to this “cooperative” view. For further limits to this view see Russell Hardin’s (2006) discussion.

³ The most influential account is Mark Warren (2017), who develops Russell Hardin’s encapsulated interest account and applies it to political settings. Warren largely assumes that political institutions can appropriately be trusted. I do not think that we can justifiably make this assumption, hence the extensive argumentation in Section III. There has been somewhat less direct interaction between the philosophy of trust and philosophers of political trust. Exceptions include Preston-Roedder (2017) and Govier (1992a). There has been greater crossover in discussions of the political effects of distrust, much of it influenced by Krishnamurthy (2015) and other papers introduced in Lenard (2015). I believe that political trust is also best approached in line with the arguments of this chapter, but I leave that for future work.

⁴ On non-cognitive and emotional accounts of trust see for example Jones (1996); Becker (1996) and Kirton (2020). I am happy to accept that some forms of trust may be affective, though I do not think that all are. On trust and the participant stance see Holton (1994); Holton’s suggestion that trust is linked to relying from the participant stance has also been adopted by Katherine Hawley and Hieronymi (2008). Holton and Hieronymi’s accounts are also among the most strongly cognitive: they both associate trust with a belief about the other agent. For trust as commitment-based see Hawley (2014); and for an obligation account see Nickel (2007). As will be clear from the main text – particularly §§6 and 7 – I do not think that all these accounts are in actual disagreement, and my aim here is not to demonstrate that any of them are wrong except in their claim to be *the* account of trust.

How should we approach understanding trust and trustworthiness? Philosophers of trust have usually started with trust and attempted to provide a unified account of what trust is.⁵ As Thomas Simpson points out, philosophers have tended to try to give accounts of the form ‘trust is this’ (indeed of the form trust is this and *only* this) (Simpson, 2012, p.550). This kind of approach has, however, faced significant problems.

1.2 Problems for unified accounts

Thomas Simpson has suggested that we could better grasp trust by moving away from unified accounts of it. I agree with Simpson that there are good – though no ‘knock-down’ – reasons to be sceptical of taking a unified approach. He points to the wide range of accounts from the very thin to the very thick all of which seem to capture some, but not all, of our ‘folk concept’ of “trust” (Simpson, 2012, pp.551-553). This is apparent as every unitary account so far developed has been readily challengeable by everyday counterexample.^{6,7} For example, the goodwill account of trust – according to which trust is to rely on another agent being disposed towards you with “goodwill” motives – suggest that a confidence trickster trusts her mark: she relies on the target of her con’s goodwill for it to be successful. Yet, talk of trust here seems misguided. An account where we understand trust as relying on another’s commitment to act in a certain way can rule out the confidence trickster’s trust, but it struggles with cases where individuals can be betrayed even though a commitment has been explicitly refused (e.g., it seems that a sailor clinging to a raft following the sinking of her ship may rightly feel betrayed when she is not picked up

⁵ The full scope of many accounts of trust is not clear. Baier, for example, talks of both ‘different forms of trust’ and ‘versions of trustworthiness’ suggesting an open, more pluralistic view, but then defines trust as ‘reliance on another’s good will’ which suggests this is *the* account of trust (Baier, 1986, pp.234-5). Nearly all philosophers of trust have chiefly been interested in interpersonal relationships and their limited discussion of trust in other contexts can make it uncertain whether they feel that they are providing an exhaustive account of trust. Those who do deal with trust in other contexts either attempt to expand interpersonal trust – see for example (Preston-Roedder, 2017) account of civic trust as goodwill trust – or suggest that they may be doing something revisionary (as Onora O’Neill, 2014, suggests that her account of trustworthiness for organisations would be).

⁶ For a good summary of these counterexamples see (Kirton, 2020, pp.583-590).

⁷ While I agree with Simpson (2012) that there are *prima facie* reasons to take a pluralistic approach to trust, I disagree with his genealogical excavation of an ‘Ur-trust’. See §1.3 below.

by the rescue vessel, because more senior ranking officers who are in a lifeboat are saved first – even though such protocols were known to the sailor).⁸

Current accounts are also diverging. The discussion is increasingly divided into cognitive and non-cognitive accounts – a point emphasised in Simpson (2023, pp.14-20).⁹ Where the philosophical investigation of trust most closely touches on empirical work in political science, there is a concern that theorists' conceptualisations are actually attempts to capture different phenomena (Bauer and Freitag, 2017). While common features do remain, it is unclear that the field is moving to a better understanding of what unites the various occasions of trust.

The range of counterexamples and growing divergence of accounts offer a *prima facie* reason to doubt the viability of a unitary approach as it has so far been pursued. The question is what to then do. Some have suggested that we ought to be more pluralistic about what trust is. I think that this is to start off towards the correct location, but current approaches have taken the wrong route.

1.3 Current alternative approaches

Current pluralistic accounts take a genealogical approach (Jones, 2012; Simpson, 2012, 2023, Ch.1). That is, we need to give a description of a situation which shows how the need for some concept would have arisen (given a specific set of conditions that are applicable to our actual circumstances), which will illuminate some otherwise disregarded aspect of the concept or (more ambitiously) show us a need we have and explain the concept that would fulfil that need.¹⁰

⁸ The account of trustworthiness requiring goodwill is that of Baier (1986). The confidence/trickster examples come from Holton (1994, p.65). The commitment account has been influentially proposed by Hawley (2014) and the sailor counterexample is taken from Kirton (2020, pp.587-590). I consider the con-artist and sailor examples – and show how my own account can explain them – in Chapter 6.

⁹ See also the excellent overview of the wide variety of different accounts provided by McLeod (2020).

¹⁰ This draws on Bernard Williams (2002, Ch.2). Simpson (2012) explicitly takes an approach indebted to Williams. Jones' (2012) approach appears to owe more to conceptual engineering (Haslanger, 2000), which is of the more ambitious type. Both align with Williams' methodological distinction. A genealogical argument is constructable from the account I present here, but it is not the approach I take.

Both agree that standard conceptual analysis has failed.¹¹ They then develop accounts which prioritise a form of trustworthiness (for Jones) or trust (for Simpson) rooted in reliance. In this, they follow the dominant tradition in philosophy of trust, which – stemming from Baier (1986) – is concerned with discriminating between relying and trusting. I disagree with this prioritisation, because I think it starts with an incorrect characterisation of the kinds of agents that we are. Once we have this clear – as I will try to do in Section 3 – the prioritisation is not obviously justified.

I instead want to develop an analysis which does justice to the diversity of trust and avoids prioritisation. That is what I argue for in this chapter.

1.3 Getting clear on my approach

Our initial, pre-theoretical responses to questions of what trust and trustworthiness are point in a pluralist direction. As we move between different domains of activity trust and trustworthiness seem to shift. When we consider businesses, it seems that the trust we put in them is different in kind, not just in degree, from the trust that we have in our interpersonal relationships. We may not, however, be willing to give up the idea that we do actually trust certain businesses, although our experience of trusting is so different. That is true of many other situations and potential trusted people, as I discussed at the start of this chapter.

Philosophers have tended to be dismissive of our commonsense understanding of “trust” and “trustworthiness”. It is, it is said, confused or picks out such a ragtag of phenomena as to be of little use. Karen Jones is representative when she says that ‘ordinary usage is quite messy and probably does group together a number of distinct, more or less closely related, trust/trustworthiness concepts’ (Jones, 2012, p.62). I disagree. I think that the reason philosophers have tended to think this is because the philosophy of trust is held back by a problem of attention, like that which affected ecology. There are the big beasts of ecology. Panda and polar bears fascinate, and their study and conservation absorb time, effort and resources. However, if we are to understand how ecosystems operate then we must also understand mosses, bugs and liverworts. The philosophy of trust has been taken in

¹¹ As will become clear, I do not.

by its “sexier” forms: the more moralised or virtuous forms of our trust-concepts. If we need to understand trust to understand society then we must understand trust in its mundane diversity, as our everyday understanding of them does.

I think that we have reason to doubt our ordinary usage is so confused. How we can utilise a concept does not necessarily demonstrate that we have a good grasp of what the concept is, but humanity does generally do quite a good job at trusting. We do not seem to be in a situation where – outside of theory – we are fundamentally mistaken or even particularly muddled about what trust is and when we should place it. I think we can and should accommodate the diversity of our folk concept. My aim here is to show that there is a conceptual unity to our folk concepts which is useful and can help us to develop trust research. Our folk concepts, as referred to in ordinary usage, are not some conceptual pick ‘n’ mix.

Taking our commonsense understanding seriously means that I will draw heavily on ordinary usage and on the intuitions that we have about our trust-concepts. The account that I present in §6 is, I think, an analysis of our overarching folk understanding which links together the divergent forms of trust and trustworthiness.

To get there, I will also adopt an alternative approach to that which has been prevalent. The most common primary object of analysis is “trust”. I think that this has been a mistake. We can get clearer by starting with “trustworthiness”. In doing so, I agree with Karen Jones that the strongest argument in favour of this approach is that if we have failed to make progress through one method, it is at least worth trying another (Jones, 2012, pp.61-62).¹²

These are two of the three core elements of the approach that I will adopt. The final one is to provide a fuller account of the kinds of agents which we are. It is especially important to consider the social nature that human beings have. This is lacking in current pluralistic approaches. This misguides their understanding of what “trust” and “trustworthiness” are. My account therefore starts with a story about the kinds

¹² As Jones notes, the trustworthiness-first approach has also been adopted by Potter (2002). It has been taken by Jones (2012) herself and Kelp and Simion (2023). A dual “trust” and “trustworthiness” approach has, I think, been taken by Hardin (2006) and Hawley (2019).

of agents we are. It is not, though, a genealogical story. It does not say why a need for something like “trustworthiness” or “trust” would arise. Instead, I aim to describe features of the world as it is and show that there is a useful and important kind picked out by the concept of “social vulnerability”. I think that the failure to consider this important kind has held back our understanding of our trust-concepts.

I do not provide an argument in favour of these three elements, nor can I demonstrate the methodological errors of other approaches. It is unnecessary to do so. I instead hope to demonstrate through the course of this thesis (and the specific analysis of reliance-responsive trustworthiness and trust that follows in the rest of Section II) the usefulness and plausibility of my account. For those who are unpersuaded by this demonstration, the account offered of reliance-responsive trustworthiness and trust in Chapters 5 and 6 can explain as many of the phenomena associated with trust as any other unified account. If you prefer to take this account as offering another unified account, then you can.

2.0 The Social Vulnerabilities Approach

I call this approach to understanding our trust-concepts the social vulnerabilities approach. I will start by explaining what social vulnerabilities are, which requires a fuller understanding of human agents as limited, dependent and social creatures. I then show how certain features which agents can have coincide with pre-theoretical, ‘folk’ judgements about what is required to be trustworthy in each situation. I argue that this allows us to analyse “trustworthiness”. An agent is trustworthy if and only if she has some feature that (i) mitigates the social vulnerability another agent faces in their interactions, and (ii) this feature is appropriately self-regulated. I argue that there is no overarching trustworthiness, and that what trustworthiness we require will therefore vary by context. “Trust” picks out a range of attitudes that respond to these trustworthy features. Some of these are cognitive, while others are the emotional fulfilments of basic psychological needs. Finally, I lay out how we ought to discriminate between forms of trustworthiness and trust, and how this can guide research by focusing it on specific contexts in which particular vulnerabilities dominate and considering what forms of trustworthiness best fit these individual contexts. I then adopt this approach in the rest of the thesis, developing it further in the rest of Section II.

3.0 *Social vulnerabilities and the value of trust*

To begin, we can ask the question ‘why should trust seem so important a concept in our practical lives?’ (Simpson, 2012, p.555). I am asking this not to clarify a conceptual need, but because it helps illuminate a key feature of trustworthiness and trust. I think that trust and trustworthiness are closely related to vulnerabilities that we face when we interact with other agents and that we ought to begin to understand these concepts and their associated phenomena by focusing on these vulnerabilities. The rest of this section will work through and argue for this approach.

3.1 *Vulnerabilities*

It is generally accepted that trust *exposes us to vulnerabilities* (Baier, 1986, 1991, p.112; Govier, 1992b, p.17, 1993, pp.104-5, 109; Holton, 1994, p.67; Pettit, 1995, p.204; Lahno, 2001, p.171; McGeer, 2002, p.34, 2008, p.242; Jones, 2004; Hieronymi, 2008, p.215; Smith, 2008, p.209; O’Neil, 2012, p.323; Domenicucci and Holton, 2017, p.151; McGeer and Pettit, 2017, §2.1; Preston-Roedder, 2017, p.5; Kirton, 2020, p.538; Nguyen, 2022b, p.234).¹³ What should we take this to mean?

We can first consider what vulnerability is and why it matters to us.¹⁴ A vulnerability is an openness to harm, injury or another negative outcome due to some feature of an object. A ceramic vase is vulnerable to breaking because of its brittleness. A crocodile is vulnerable to attack because of its soft underbelly. The Axis armies of World War II focused their manpower and materiel on the Eastern and Western European fronts, and so they were vulnerable to attack from the Mediterranean.

¹³ The nature of these vulnerabilities is, of course, debated. That is something that this account helps to explain: they are debatable because they vary. I disagree with McLeod (2020) that this is characteristically taken to be a vulnerability to betrayal. That is present in some accounts (e.g., Kirton’s), but is rejected by others (e.g., Domenicucci and Holton’s), and I am not certain it even fully captures Baier’s view (see Baier, 1986, p.239).

¹⁴ There is a significant literature on vulnerabilities, especially influenced by feminist philosophy – e.g., Gilson (2014) and Mackenzie, Rogers and Dodds (2014). I am not aware of a classification of vulnerabilities in line with the ‘social vulnerability’ that I outline below. I agree with the view that vulnerability is not bad but do think that we have clear interests in generally lessening the hurt we can experience because of them.

We can be exposed to vulnerabilities by nature, by choice (as the Axis armies did when they deployed insufficient forces in the Mediterranean) or by chance (as when a crocodile rolls down a hill and is left on its back).

We can protect ourselves against our vulnerabilities in two ways. We can look to *reduce* the vulnerability. By this, I mean that can do things to alter features so that the openness to harm, injury or other negative outcome is lessened. We would reduce the ceramic vase's vulnerability to breaking if we covered it in an epoxy resin which makes it practically unbreakable. We can also *mitigate* our vulnerabilities. When we do this, we act in ways that lessen the chances of harm from those vulnerabilities. We would mitigate the ceramic vase's vulnerability by placing it in a secure display cabinet.

3.2 Limited, dependent, social agents and their vulnerabilities

We are agents who are embodied as human beings. As embodied human agents, we face limitations. We have a certain physiological structure and our bodies shape and limit our experience and what we can do. Sometimes these limitations exist for all humans, like our inability to see ultraviolet light (though we could, were we embodied as birds). Sometimes these limitations are individual; given my height I cannot reach things on the top of our wardrobe, though my taller partner can. These physical limitations produce temporal limitations. We must allow time to eat, drink and sleep when deciding what to do. Our limited lifespans give us only so long to try to do everything we want to do.

We are also cognitively limited, in part because of the limited cognitive resources that we have. The physiological processes powering our minds are energy intensive and so we have developed ways to use these as efficiently as possible. We cannot, for example, constantly reconsider what we are going to do and why we are going to do it. Such a demand would be energy intensive and would likely lead to us never doing anything at all. So, we must settle on courses of action and follow through on them (for a related argument, see Bratman, 1999).

As well as these internal limitations, we also face external limitations. If the Nile were not to have flooded, then an ancient Egyptian farmer could not have grown her

crops. If a stranger were to attack you in the street, you would not have been able to walk home safely. External happenings enable and block our own actions and plans.

This embodiment and our associated limitations make us vulnerable. We are relatively easily injured. We are severely affected by extremes of heat or cold. We are vulnerable to famine and to disease. We are also vulnerable in the exercise of our agency because of these limitations. We can fail to do things we want to do, because of how we are embodied and because things in the world do not go the way we need them to if we are to be successful.¹⁵

These limitations also make us interdependent. You can want a tool that you do not have the skills to make yourself. You can imagine and want to build a great Temple to Philosophy. Yet, both are beyond your physical limitations. We are dependent on events and others' actions and behaviour for each of us to act ourselves. We can therefore need to interact with others – or, minimally, to avoid others acting to block our plans – to fulfil our goals.

This characterisation of humans as embodied agents who are therefore limited and interdependent would be accepted by Jones and by Simpson (see also, Baier, 1986, pp.236–7). It is, though, an incomplete portrait of what agents like us are like, and it means that they miss a distinctive kind of vulnerability that we face. What is missing is that we are embodied as *social* creatures.¹⁶

We are social creatures and, being social creatures, we need to interact with each other. There is growing evidence from medical research that loneliness and isolation have negative impacts on our functioning. We have worse mental and physical health if we do not interact with other individuals (for an overview see Park et al., 2020). This appears to be a physiological disposition, as well as being reinforced by pro-social norms.

¹⁵ I discuss this in greater detail with respect to reliance and planning agency in the next chapter.

¹⁶ This makes the prioritisation of reliance-related trustworthiness and trust which I discussed in §1.3 hard to maintain. It requires further argument than either Simpson or Jones give to justify prioritising our interdependence over our sociality. The alternative route seems just as justified. I do not think there is really a dilemma here, and I develop an informative approach to our trust-concepts that avoids the fork.

Our sociality means that we have to interact, and so we have an interest in mitigating a distinctive kind of vulnerability which arises in our interactions while simultaneously limiting our options for how to do so. This gives rise to our trust-concepts, which help us navigate this situation. I will introduce this distinctive *social* vulnerability in the next sub-section, how we mitigate it in the subsequent one, and its links to our trust-concepts in the next section.

3.3 Distinctly social vulnerabilities

Some of our vulnerabilities are like those of any other physical entity. A vase's vulnerability to smashing because ceramic is brittle is much like an elderly person's vulnerability to a fracture because her bone is brittle.

We are also vulnerable in ways other entities are not. We face specific *social* vulnerabilities. By social, I mean that these are vulnerabilities which are characteristically interpersonal.

The distinctive nature of a social vulnerability can be seen in the contrast between the following cases.

Fall 1: Out walking, you slip on black ice formed in a recent cold snap.

Fall 2: Out walking, you slip on black ice formed by a TikTok prankster.

Both falls rest on the same basic vulnerability: the hurt that comes to humans when they fall. Only *Fall 2* includes a social vulnerability in addition to that basic vulnerability. *Social vulnerabilities* can be defined as:

the openness of an agent, *A*, to a particular harm, injury or other negative outcome because of some feature of *A* when *A* interacts with another agent, *B*, where the harm, etc., to *A* comes about *only as part of that interaction*.

It is important to note that social vulnerabilities can impact non-social or basic vulnerabilities but remain distinctive. The harm of falling remains when you slip because of an influencer, but you are vulnerable in a distinct, additional way because you are interacting with another agent. (I further clarify 'only as part of that interaction' in §5.1.)

There are at least two important sub-classes of social vulnerability.¹⁷ Some social vulnerabilities arise because of our sociality. Most prominent among these are our socially satisfiable needs. These are the needs we have the satisfaction of which is constituted by our interaction with others. This makes us distinctively vulnerable in our interactions with others to their withdrawal from us, and so to the non-satisfaction of our need which harms us. We may have many such needs and enumerating them is the work of psychologists and clinicians as well as philosophers. Among them are a psychological need for group membership or identification and for certain forms of mutual love. These are satisfied only in others accepting us or loving us back.

The other important class arises because of how our limitations make us interdependent. This makes us vulnerable in our interactions to our own actions and plans (the exercise of our own agency) failing because of how others behave. We can call these our “agential vulnerabilities”. They can be expressed in the form ‘For *A* to *Y*, *B* needs to *X*’. We face agential vulnerabilities in our practical agency. We also face them in our epistemic agency, where the appropriate formation of an epistemic attitude is dependent on another.

Our lives teem with social vulnerabilities. Consider the following examples:

Stranger Danger: Anthony walks home through a park late at night. Beatrix follows him. As Anthony approaches an especially dark area of the park, he hears Beatrix’s footsteps quicken.

Partner’s Action: Ava and Bertrand are a married couple in an exclusive relationship. Bertrand is going out for a meal with a former partner.

Peatlands: You read a newspaper article written by an expert on British peatlands, in which the expert explains that their mismanagement for game shooting and destruction for agriculture, horticulture and whisky making has decreased biodiversity.

¹⁷ There may be more, such as distinctive moral social vulnerabilities (ways in which our moral personhood is open to damage, harm, etc. in our interactions with others). Developing this point would take us beyond the focus of this thesis.

Anthony faces a social vulnerability of being attacked by Beatrix. Ava runs the risk of Bertrand acting in ways which withdraw the satisfaction of her emotional needs. If the expert has lied, then you face the danger of mis-forming your beliefs. These are social vulnerabilities because they exist only because of our interaction with other agents. Our social vulnerabilities matter to us, and we are highly aware of them even if we would not recognise the theoretical coherence that I am giving them. “Social vulnerability” picks out an important kind of vulnerability that we face.

We can summarise the social vulnerabilities (SV) from our examples in the table below with the underlying basic vulnerability in brackets:

Table 1		
<i>Stranger Danger</i>	<i>Partner’s Action</i>	<i>Peatlands</i>
SV: to attack (e.g., physical injury)	SV: to withdrawal of satisfaction of emotional needs (e.g., heartbreak, distress, jealousy)	SV: to bad testimony (e.g., faulty belief formation)

To clarify, social vulnerabilities are not exclusive to any relationship. In the mess of real-life interactions, different social vulnerabilities are likely to arise mixed together. We can, however, usefully and non-arbitrarily distinguish these social vulnerabilities. I also think that we can identify and describe the social vulnerability (or vulnerabilities) that are most characteristic of particular relationships, as demonstrated in the examples above.

3.4 Reducing social vulnerabilities

Social vulnerabilities matter to us because they raise the chances of things going worse for us. We have an interest in navigating our interactions while lessening the exposure of our social vulnerabilities. This is part of a general interest we have in lessening the exposure of our vulnerabilities. We generally ought to reduce risk when making decisions. Vulnerability is closely tied to risk. When we consider potential courses of action or outcomes, we can think about how these will expose us to one or more vulnerabilities. A riskier option can expose us to a greater chance of

exposing a vulnerability or of a range of vulnerabilities becoming exposed. All else being equal, we try to minimise the chance and range of vulnerabilities that we must face. We therefore try to minimise risk: if given the option of walking along a clifftop path on a windy day or taking a path across the meadows away from the dangerous drop, we would normally pick the latter.

If we can navigate our interactions while lessening the exposure of our social vulnerabilities, then we will usually do better. Things will go instrumentally better because, as far as a person's goals are dependent on this interdependence, she will be better able to achieve them if she can better lessen the chances of failure her social interactions introduce. Things will not just go instrumentally better. The protection of some social vulnerabilities is inherently valuable: for example, the satisfaction of many socially satisfiable needs like belonging and loving and being loved back are *prima facie* inherently valuable.

We can try to reduce our social vulnerabilities. One way to do this may be to attempt to change our sociality. This is not easy to do once we acknowledge our embodiment as social creatures. We would need to change our physiology or be complex agents without physiological sociality. There may be no social vulnerabilities in a future populated with superhuman, independent bionic space-hogs and Artificial Intelligences. This Muskian fever dream though seems to be a future lacking some central parts of what makes life human. If what I argue in this chapter is correct, such a world would have no place for trust.

We can try instead to mitigate our social vulnerability. We can isolate ourselves or limit our goals to only those which are not dependent on others. This is perhaps why the Stoic will have no need for trust, as raised by (Baier, 1986, p.231). This splendid isolation is, though, difficult to achieve; even the hermit Simon Stylites, living on top of his pillar, needed someone to bring him food. Isolation is no way to mitigate our vulnerability because of our socially satisfiable needs. It is simply to accept the harm that comes from their never being satisfied.

Power is also an option for mitigating social vulnerability, at least of some kinds. Economic, political and social power may allow an individual to ensure that she is less vulnerable to others, by altering how they can behave towards her or increasing

her independence. Hierarchical power will though leave those less powerful still socially vulnerable. Equality of power does not seem to have the same mitigating effects.

For agents like us isolation is rarely an option. Our expansive goals and aims give us reason to pursue, rather than refuse, our interdependence. Our sociality pushes us to interact. This means that for us social vulnerabilities are real, and they matter. We maintain, however, an interest in their mitigation. That, I will show, is where our trust-concepts come in.

4.0 Social vulnerabilities and trustworthiness

Why think that these vulnerabilities are linked to trust and trustworthiness? My conjecture is that “trustworthiness” refers to a variety of features that agents can have, which mitigate the social vulnerabilities to which other agents are exposed by interacting with them. We can think of a trustworthiness_{physical harm} or a trustworthiness_{testimony}. We tend not to distinguish these forms of trustworthiness in our everyday speech, but in embracing this more pluralist view we can better illuminate what trustworthiness is and what it requires of agents. I will show that there is a conceptual unity to our trust-concepts despite the variety of phenomena and forms of trustworthiness and trust, which allow for more specific conceptualisations to be formed to better guide action.

I will attempt to show that trustworthiness picks out phenomena that mitigate certain social vulnerabilities by considering our earlier examples and seeing how where these vulnerabilities are mitigated in specific ways it seems appropriate to talk of trustworthiness.

4.1 A first hypothesis

Here is my first hypothesis about how “trustworthiness” is related to social vulnerability:

H1: An agent is trustworthy when she has a feature that mitigates others’ social vulnerabilities to her.

This is different from the ways of reducing social vulnerability discussed above. It requires something of the agent to whom we are vulnerable, rather than changes in or by the vulnerable person. We have reasons to think that *H1* is correct. First,

situations in which it is appropriate to talk of whether *A* ought to trust *B* or ask if *B* is trustworthy – that is “occasions of trust” – happen only when social vulnerabilities are present.

To be a *social* vulnerability the vulnerability must be *part of the interaction*. What this amounts to will vary depending on the nature of the social vulnerability. For socially satisfiable needs the withdrawal or end of the interaction often amounts to the vulnerable party being harmed. The boundaries of what counts as an interaction for our agential vulnerabilities may vary. Consider the following situation first raised by Annette Baier (1986, p.235):

Königsburg: Immanuel Kant went for walks at the same time every day.

Some of the citizens of Königsburg set their clocks by his walks.

Intuitively this is not an occasion of trust, it is one of merely reliance. I think that this is because while there is a vulnerability here, it is not properly a *social* vulnerability. This is due to the nature of what classes as an interaction, and the fact that Kant’s going for a walk is not properly thought of as an interaction. Some reasonable boundaries to what counts as an “interaction” seem fairly apparent. An interaction cannot be any event brought about by an agent which affects another agent. This is too expansive. If a playing child pulls a leaf off a hedge and drops it on the ground, then later I slip on it, we surely have not interacted. It would also seem, intuitively, to be part of that child being trustworthy that she does not drop the leaf. Yet, the prankster who targets me and I have interacted. Whatever it is that constitutes an interaction must take account of various considerations, including the actors’ reasonable intentions, and the reasonable interpretations of the action open to bystanders, which will be affected by moral and social norms that demarcate what is private and what is public.¹⁸ The vulnerability here is instead the same kind of vulnerability we face when we are reliant on an object or natural process. I will clarify the distinction in the next two chapters.

¹⁸ This means that distinct cultures may have different sets of actions that count as interactions; and so social vulnerabilities may vary between cultures. I take this to be an advantage of the account, which helps to explain why the appropriateness of trust varies highly across cultures. This may pose problems for organisations, as discussed in Chapter 11. I develop this point further in Chs.4, 5 and 6.

There seem ways that Kant's walks could become occasions of trust. They involve altering the situation into one which is more clearly an interaction. This would happen, for example, if Kant made a commitment to his fellow citizens to act as a timekeeper.

Where, intuitively, we are in an occasion of trust, there are social vulnerabilities. In the three cases outlined above, social vulnerabilities are present, and it seems appropriate to ask whether one of the agents should trust the other, and whether Beatrix, Bertrand or the expert are trustworthy. Where there are no social vulnerabilities, we do not have an occasion of trust. This is best explained by social vulnerability playing an essential role in the concept of "trustworthiness" and in "trust". This relation makes sense. Things go better for us if we can lessen the exposure of our social vulnerabilities. We need a way to pick out features of other agents which do this.

4.2 A problem with the first hypothesis

H1 appears necessary, as the preceding discussion illustrates. I cannot think of an intuitive occasion of trust which does not involve a mitigation of a social vulnerability. It cannot, however, be the whole story about what "trustworthiness" is. *H1* can be satisfied by features that are clearly not "trustworthy". Consider those laid out in Table 2.

Table 2		
<i>Stranger Danger</i>	<i>Partner's Action</i>	<i>Peatlands</i>
SV: to attack	SV: to withdrawal of satisfaction of emotional needs	SV: to bad testimony
Beatrix is disposed to react to threats of violence from others.	Bertrand is responsive to the legal commitments of marriage because he fears the cost of a divorce.	The expert values having a good reputation and does not want it damaging by a failure to report her research accurately.

It is inappropriate to call individuals with these features “trustworthy”. In fact, I agree with the stronger claim that these features are incompatible with trustworthiness. Imagine that Anthony attempts to trigger this disposition in Beatrix by threatening to hit her should she come nearer; or Ava reminding Bertrand as he leaves for the drinks that, “A divorce is a costly business.”; or the editor reminding the expert of the professional scandal should she misreport her findings. It is appropriate in each situation for them to ask, “Do you not trust me?” This is an incompatibility that has been noted before and which, I will show, is grounded in the nature of “trustworthiness”.¹⁹

In fact, trustworthiness seems incompatible with any form of exercise of authority, manipulation or coercion. Imagine Anthony attempts to use authority. Maybe, Anthony is an off-duty police officer, who could brandish his identity badge to Beatrix and warn her to keep her distance. Or Anthony could even attempt to incentivise Beatrix not to attack. He could offer her some money if she refrains, saying that he will pay her once he has reached the other side of the park and are back in a busy area. Any of these might help to mitigate Anthony’s vulnerability to being physically harmed. They all raise the question, “Do you not trust me?”

When we can ask whether another person has trusted us, it standardly follows that a question of our trustworthiness has also been raised.²⁰ The above features mitigate our social vulnerabilities, but they do not do so in a way that implies we are taken to be trustworthy. That is why they raise questions of trust.

This is because the cases in Table 2 bring about a mitigation of our social vulnerabilities in the wrong way. We can mitigate our social vulnerability through

¹⁹ It is most prominently noted by Hieronymi (2008) and Nguyen (2022b). I think that this incompatibility is because of the necessity of self-regulation, as discussed in the main text. This account is, however, compatible with certain attitudes which are forms of “trust” themselves being incompatible with such actions; though I think it unlikely that all forms of trust are. See the discussion of the variety of trust below. The need for self-regulation suggests that non-agents cannot be trusted, *contra* Nguyen. We may use “trust” for attitudes that are similar in their phenomenological effects, and trust on this account certainly reduces the need for worry.

²⁰ I say standardly because there are of course times where we may hope someone has trusted us, though we are unconcerned with our trustworthiness. For example, a confidence trickster may not have any intention of behaving trustworthily but nonetheless hope that her mark trusts her. Such cases are, however, parasitic on genuine cases of trust.

authority, manipulation or coercion or by establishing social structures (such as legal or social norms) and using their punitive aspects to ensure appropriate behaviour by non-conformers. These actions are a form of *regulation*, where what it is for an entity to regulate another entity is to bring it about that that object performs in line with some goal.²¹

Regulation serves our interest in mitigating our social vulnerability but is costly to set up and maintain. It requires us to redirect our limited energies towards regulating others and monitoring their compliance. A regulator needs to sit and watch to ensure that desired behaviours occur. That is either costly to her limited time and resources or requires her to set up systems of surveillance, which are then likely to simply push the need for regulation back a step. A regulator will need to impose penalties if someone misbehaves. Ava, for example, must be willing to follow through on her threat of divorce if she thinks that it will ensure Bertrand behaves appropriately.

H1 is insufficient and an important class of features that mitigate social vulnerability but are not trustworthy are those that enable regulation. It was, however, the case that intuitive cases of trustworthiness did mitigate social vulnerability. We therefore require a further condition which specifies how this reduction comes about.

4.3 A second hypothesis

Let us try to pick out features which mitigate social vulnerabilities but do not require our regulation. That suggests:

H2.a: An agent is trustworthy when she has a feature that mitigates social vulnerability that does not need to be regulated by others.

This does not seem right either. In *Stranger Danger*, Beatrix could have a phobia that makes her incapable of approaching others. That is why she goes for walks late at night. This feature mitigates Anthony's social vulnerability. She is, intuitively, not

²¹ This meets the standard uses of regulation. It aligns, for example, with Braithwaite's general definition 'as that large subset of governance that is about steering the flow of events, as opposed to providing and distributing' (Braithwaite, 2011, p.217). In contrast to regulatory studies, I standardly use "regulation" only for external efforts to bring about performance in line with some goal.

trustworthy. Similar features may be developed for the other cases. My conjecture is that this is because such a phobia has still not been brought about in the right way. Something is missing in the relationship between Beatrix and that feature which mitigates Anthony's social vulnerability to her. The next subsection will explain this missing relation.

4.4 *Self-regulation*

We can regulate others, but we can also have features that are self-regulated. Say a toddler throws a tantrum. There are two ways that it could be brought about that she calms down. You may pick her up and attempt to soothe her. Alternatively, you might suggest that she calms down. She might see that she should calm herself down, then modify her breathing to settle herself, or distract herself, or come to you for comfort. You may still play a role, but *she* has brought about the change in state in herself.

For consistency, we can think of this ability as a form of *self*-regulation. In standard regulation the regulating entity is different from the entity being regulated. In self-regulation they are the same. A key feature of self-regulation is that it does not require the action of another entity or agent for the appropriate behaviour to be brought about. So, a room's heating may be regulated if it requires that someone turn the radiator up or down to maintain a certain temperature. It will be self-regulating if a thermostat brings about the maintenance of this temperature.

We self-regulate by paying attention to reasons and by having appropriate mechanisms (whatever exactly these are) which allow us to modify our actions accordingly. This is a crucial element of human development, as in the toddler example.

An agent, *B*, can self-regulate if and only if:²²

²² Similar features for self-regulation have been noted (Debus, 2016 and McGeer and Pettit, 2002). It is also close to the reasons-responsiveness developed as part of self-ownership argued for by Fischer and Ravizza (1998, Ch.8). Self-regulation is a weaker requirement than moderate reasons-responsiveness, which Fischer & Ravizza take as essential to moral responsibility. I leave it open that an agent may be able to self-regulate so long as she takes something to be a reason (even if it is not actually a reason). This could be enough for trustworthiness (though of a very non-robust form); but in practice such aberrant reason-recognition will be unlikely to be recognised by others and unlikely to lead to trust.

- i. *B* can recognise a reason to bring about a feature in *B*,
- ii. The reason for *B* to bring about this feature is taken as a reason for *B* independently of others acting to impose that reason on her,
- iii. *B* can appropriately reason so that she aims to act in ways that bring about this feature in *B*, and
- iv. *B*'s reasoning and actions are involved in the actual causal sequence that brings about this feature in *B*.

We can see how (i) is necessary because unless an agent can recognise something as a reason, she will be unable to bring about an action. (iii) ensures that a feature is not brought about accidentally. It appears inappropriate to consider the toddler to have calmed herself, if she did not see how the actions that she considered taking were linked to calming herself. (iv) is required if the self-regulation is to be successfully brought about by that agent herself.²³ If the toddler was actually calmed because she becomes so exhausted her brain cannot maintain the emotional exertion, then something again seems to have gone awry. (ii) is necessary to rule out standard regulation. It does not require that others' actions cannot be sources of reasons for an individual, if, for example, those actions are not intended to bring about changes in others' behaviour.²⁴

This definition is somewhat stipulative. It is an attempt to capture a feature that we commonly recognise.

Self-regulation puts us in a specific relation to a variety of features that we develop in ourselves. It is absent, however, when in cases like Beatrix's phobia. A phobia is not something that can be self-regulated.

²³ I use "bring about" following Bratman (1999) to capture a range of potential causal mechanisms, whatever exactly they may be.

²⁴ It may be worried that this leaves open situations where an individual takes herself to have a reason related to others but does not actually have such a reason. For example, someone who thinks that the police will arrest her if she behaves in a certain way, although this will not happen. The second condition is meant to rule out situations where the stability of this reason for the agent relies on the actions of others. This situation would not count as self-regulation so long as were this person to find out that the police would take no action, the reason would cease to be an effective one to alter her activity.

This suggests a more specific hypothesis:

H2.b: An agent is trustworthy when she has a feature that mitigates social vulnerability that is self-regulated.

This seems better. The features summarised in Table 3 satisfy *H1* and *H2.b* and each align with commonsense intuitions about “trustworthiness”.

Table 3		
<i>Stranger Danger</i>	<i>Partner’s Action</i>	<i>Peatlands</i>
SV: to attack	SV: to withdrawal of satisfaction of emotional needs	SV: to bad testimony
Beatrix has a minimal commitment to legal or moral norms. ²⁵	Bertrand has emotional care towards Ava.	The expert believes in using a reliable methodology and is honest.

Self-regulation helps us to understand why trustworthiness matters so much to us. It is good to be able to pick out a kind of feature that other agents have that mitigates our social vulnerability. Features that are not self-regulated are costly if they require that we act to maintain them. We may care less about features that are not self-regulatable either because they are less predictable (a phobia may sometimes not activate or is less standardly available) or because it cannot be reliably encouraged or developed in others. In contrast, self-regulatable features will

²⁵ The establishment of enforceable laws makes us less socially vulnerable, as discussed above; but it is only when the rule of law is internalised by individuals that we can be said to live in a trustful society. This contradicts Thomas Simpson’s suggestion that this is insufficient for such a society, which appears overly demanding (Simpson, 2024). As I discuss in the main text, there are varying contexts to trust and trustworthiness, and we should be careful to limit the robustness of the trust that we require to the demands of that context.

This feature of trustworthiness may also help to explain how certain institutional arrangements do not support trust but undermine it (Pettit, 1995, p.65; O’Neill, 2002, and Jones, 2012, p.73). If an arrangement regulates behaviour, then it will standardly be taken as being incompatible with the regulated agent being trustworthy. This suggests that to support trustful societies, we ought to arrange institutions so that they foster internalised commitments, such as through schemes of professionalisation.

standardly lead to certain behaviours and can be highlighted to other agents. Features that both mitigate our social vulnerability and can be brought about and maintained by those to whom we are vulnerable are especially valuable.²⁶ Agents like us are well-served if we can locate such features; our lives in our social world will go so much better when we do. We have good reason to consider them features worthy of having. This is the role that “trustworthiness” and “trust” play for us and why they matter so much.

Our two hypotheses have helped us to pick out two conditions that underlie our commonsense or folk understanding of trustworthiness.

5.0 The Social Vulnerabilities Account

5.1 Trustworthiness

This suggests the following analysis of “trustworthiness”:

B is trustworthy (to *X*) for *A*, if and only if:

- i. *B* has a feature that will mitigate *A*’s social vulnerability to *B*’s *X*-ing;
- ii. *B*’s feature in (i) is appropriately self-regulated (as defined above).

This aligns closely with commonsense intuitions about occasions of trust and evaluations of trustworthiness. The definition also captures a wide range of the potential analyses of “trustworthiness” that have been offered in the literature. I think that this shows that the account has captured a common thread underlying them, and one which can help us to clarify their relationships.

This account is also pluralistic about the nature of trustworthiness. The features in Table 3 are only some of those that reduce the dominant social vulnerability in that situation. Beatrix could have a range of other features, such as a cultivated gentleness, so long as they mitigate Anthony’s social vulnerability to attack and are appropriately self-regulated. Trustworthiness – in its general and specific forms – is multiply realisable.

I think that both are advantages of this account. They help to show why we have such a diverse understanding of trust and trustworthiness. They allow us to better

²⁶ This does not mean that trustworthiness is only instrumentally valuable. Certain forms of it may be intrinsically valuable: for example, certain emotional forms.

clarify how we should both research our trust-concepts and to better direct us if we want to be trustworthy and to trust well. I will explain how in §6. First, I will complete the account by considering trust.²⁷

5.2 Trust

We can then consider the following analysis of “trust”:

Trust: An attitude is an attitude of “trust” if and only if it attempts to track the presence of features of an agent that will mitigate a given social vulnerability and is fitting only if those features are in fact present.

Several considerations favour this analysis. It closely aligns with attitudes which we would intuitively consider trust; and no case of trust that I can think of does not include some attempt to track a feature that reduces a social vulnerability. This flips the standard understanding of the relationship of trust and vulnerability. To trust is usually thought of as being to open oneself up to a vulnerability. On this account, all trust is an attempt to reduce vulnerabilities we inevitably face.

It also easily explains the broad range of experiences of trust, which we discussed at the start of this chapter. This is because it is again pluralistic about the nature of trust. Our experience of trust is highly diverse because “trust” picks out various attitudes which are phenomenologically distinct. Some are nothing more than a belief that an agent has some appropriate feature or will behave in a way that reduces our social vulnerability.²⁸ Some may be a complex range of attitudes, such

²⁷ As Katherine Hawley (2014, p.3) has forcibly argued, a desiderata of an account of trust is that it can explain distrust (and the absence of trust). Distrust and untrustworthiness are not my main focus here. The account can, though, be easily extended. An untrustworthy feature is one that is self-regulated, but which affects or exacerbates our social vulnerability. Distrust are attitudes which track such features. We may neither trust nor distrust when we do not face a social vulnerability, so evaluations of trustworthiness are irrelevant. This can help to further clarify some forms of distrust. When we are not accepted and this leaves our need for social acceptance unmet, that psychological need is an exposed vulnerability. If – as I argue below – the meeting of that psychological need is “trust” in this form, then its absence may be distrust.

²⁸ It also helps to clarify the range of ways that “trust” may interact with evidence. The standard view is that trust must go beyond evidence, but Simpson (2017) has argued – in my view correctly – that this is not always the case. Though I disagree with Simpson, and especially the development of his view in Simpson (2023) that evidence responsive trust is standard; it is standard only of certain forms of trust, and these forms of trust may not be dominant (even if they ought to be) in different situations. I do agree that, as Simpson says, those undertaking an expedition to the South Pole ought only to trust if they have very good

as relying on another and forming some appropriate evaluation of her. In some cases, the best attitude to attempt to track the appropriate feature may be an emotional one. Felt attachment, for example, may best track the actual satisfaction of that socially-satisfiable need (Kirton, 2020, pp.590*ff.*); and so felt attachment will be a form of trust which most easily and best explains how even babies can trust (Baier, 1986, pp.241-242). Likewise, my love for my partner may be a major form of the trust that I have in him.

This also helps us to understand why trust has seemed to be three-place – the standard view, again originating in Baier (1986, p.236) (see also Holton, 1994, p.67; Jones, 2012, p.62; Hawley, 2014, p.1) – two-place and one-place (Domenicucci and Holton, 2017; Faulkner, 2015). This is to say that trust is sometimes in an individual agent to do something (*A trusts B to X*), trust is just in an agent (*A trusts B*) or that to trust is simply to have an attitude without a specific object (*A trusts*). If the forms of trust are each different attitudes, then they simply have different appropriate objects and potential content. If mutual love can be a form of trust, then it would best be understood as a two-place form. If – as I will argue in Chapter 7 – a form of trust involves reliance, then it will best be understood as three-place.

It may be objected that this account of trust misses some distinctive commonalities of trust, among these are its contagiousness, the fact that it strangely interacts with evidence, the fact that we ought to be responsive to others' trust, the distinctive responses of betrayal and being let down, and that trust (as an attitude) standardly leads to action. How can we explain these common features with a variety of different attitudes?

I will respond to only one of these directly here: the relation of trust as an attitude and as an action. I will argue that we can explain the others for different forms of trust attitude in Chapter 6. The argument there will attempt to show that there is no reason to think that we must have a single account of trust to explain these features;

evidence that the person they are trusting will provide them with their airdropped provisions. My account would explain this based on the severity of the potential vulnerability and the specific form of trust. As I will argue in Chapter 6, reliance-responsive trust involves a judgement of reliance-responsive trustworthiness: the severity of the potential reliance determines the amount of evidence required to justify such a judgement.

they are shared by several types of attitudes. Though, I should note that I am less certain that all these are common to all forms of trust, but I do not pursue this line of argument here.

It seems characteristic of trusting someone that you will be highly likely to act on that trust, though I do not think that it is tenable that to trust is an action. My account explains why we will often have reason to act in line with our trust. If we must navigate our social vulnerabilities and if trust attitudes follow whatever features in agents best mitigate these, then we will generally have reason to act as we trust, and, indeed, to act as we most trust. This follows from the fact that when we act, we usually aim to do so in the way that – all else being equal – reduces our risk. This will standardly be the rational thing to do. Some forms of trust will simply form part of our reasoning about what is best to do. The emotional forms of trust may have more direct motivational effect.

We trust in many ways, and this helps explain why our experience of trust varies so readily and to such an extent, as I discussed at the start of this chapter.

6.0 What trustworthiness we should want

In this section, I will address related concerns that the social vulnerabilities account fails to explain some important aspect of how we trust, in ways that make it an unappealing analysis. The first concern is that my analysis does not seem to say why we ought to be trustworthy. My account does not, however, fail to capture this normativity. The social vulnerabilities account requires self-regulation, which requires reasons-responsiveness. So, it is necessary to be trustworthy to be appropriately responsive to relevant reasons. Those relevant reasons being reasons which appropriately reduce some social vulnerability. The account allows for the fact that what these reasons are can vary, but we nonetheless need reasons to be trustworthy. There is always then at least a weak reason why we ought to be trustworthy.

6.1 Reasons, robustness and reasonableness

Say that we are interested in reducing the risk that we will not be able to do something because another agent fails to act in the required way. If an agent has her own, self-interested reasons to act in the required way, so that her interests sufficiently overlap with ours then she has reasons which appropriately reduce

another's social vulnerability and is – according to this analysis – trustworthy. So is an agent who is appropriately responsive to her commitments to act when others are relying on these commitments. Both have reasons to be trustworthy.²⁹

This response might be taken to miss the point of the objection. It is not that there needs to be any responsiveness to reasons, but that there needs to be a responsiveness to a particular type of reason or a particular type of responsiveness. You might accept this analysis, but still object that it has not got at the aspects of trust about which we *really care*. When we talk about trust and trustworthiness and about building societies rich in trust, what we mean by “trustworthiness” is something more like a virtue. Compared to what the social vulnerabilities account picks out, these would be more moralised and more “robust” in that they would be triggered across a greater span of time, for a greater range of trusting agents, and in a greater range of contexts.³⁰

I accept that we are sometimes interested in these more robust forms of trustworthiness. My account explains this. Our trust-concepts matter because they help us to navigate our interactions while mitigating our social vulnerabilities. A responsiveness to certain kinds of reasons is going to better serve this interest than others will.

In many circumstances, it will be better if we can find and develop a form of trustworthiness which has a more robust source of normativity underlying it. In many circumstances, the kind of encapsulated interests trustworthiness discussed above will be insufficient. It will no longer reduce a social vulnerability, if the two

²⁹ These are the encapsulated interest account of Hardin (2006) and the commitment account of Hawley (2014) respectively.

³⁰ That trustworthiness requires a certain robustness is argued by McGeer and Pettit (2017), Potter (2002, Ch.1) and Jones (2012, pp.73-78). Though their versions recognise some of the importance of more everyday trustworthiness, they do not seem to think it is really trustworthiness. This is a view I argue against in the main text. McGeer and Pettit's view of trustworthiness is a responsiveness to another's manifest vulnerability. In many ways this aligns with the social vulnerabilities account. I disagree – for the reasons given in the main text – that there is a substantive overall disposition that we can cultivate which reduces all these vulnerabilities across all contexts or that we ought to try to cultivate such a disposition.

agents' self-interests start to diverge.³¹ A responsiveness to commitments that the other agent has undertaken to act in specific ways will more robustly reduce our social vulnerability. We have an interest in understanding what the most robust form of trustworthiness that reduces a particular social vulnerability will be.

At certain times we want robust forms of trustworthiness. One facet of being powerful is that others will be more socially vulnerable to you; in that others will be more vulnerable to you or that more others will be vulnerable to you. It will be more important that powerful agents – whether they are individual people or collective agents like businesses – are trustworthy and to be trustworthy they will have to be more robustly trustworthy. In such situations we will therefore be concerned with robust trustworthiness.

We are not though always interested in these robust forms. In passing interactions, such as those which are normal between individuals in many modern globalised societies, it will often be enough to reduce the social vulnerability that we face that we have overlapping interests.

We ought not to think of such interactions as a trust desert. To do so misdescribes our situation and may lead to mistaken treatments and prognoses for the trust in our societies. I have argued that we can find a value in our trust-concepts linked to how they help us navigate a world in which we face great vulnerability. There may be moralised versions of “trustworthiness” or “trust” which offer a great deal of robustness, but I think that the burden lies with those who moralise trust to show why it is the sort that we ought to think the best way of analysing our trust-concepts. Just as those who insist that only robust trustworthiness is worthy of the name, need to explain why we ought to demand so much of each other in our standard interactions. Less demanding forms of trustworthiness can still help us to interact better. To dismiss them is to miss how rich our lives are in trust and how variable the demands of trustworthiness. Being alive to this context-sensitivity can help us to be better trustees and trustors.

³¹ Compare the discussion of Hardin in Jones (2012) which is similar. I think that Hardin does, nonetheless, capture an important form of limited trustworthiness. It is sufficient, for example, in certain interactions, like many one-off market transactions.

6.2 *Context-specific trustworthiness*

The social vulnerabilities account also allows us to think more clearly about what trustworthiness requires of us, not only in terms of its robustness, but in the specific form of trustworthiness which may be most relevant in a given situation. This is because different social vulnerabilities will be dominant in specific interactions. Emotional needs will be preponderate in our intimate relationships; the risk of our own actions failing because of their dependence on others in instrumental interactions.

This means that to understand what sort of trustworthiness we need, we must think carefully about which social vulnerability is our concern. Different accounts of trustworthiness will address different underlying social vulnerabilities. If we are concerned with practical agential vulnerabilities, we should be concerned with forms of trustworthiness that address this vulnerability. These include Russell Hardin's encapsulated interests, Katherine Hawley's responsiveness to one's commitments on which others are relying and many other "reliance+" accounts.

We would be misguided, however, if we concerned ourselves with a feature that reduced a different underlying social vulnerability. Say our need for belonging, as described in Andrew Kirton's (2020) account of attachment security. To be successfully fulfilled, this socially satisfiable need requires a fundamentally different kind of feature in the trusted agent than one which reduces the risk of a relied-upon action failing.

The disposition to be an appropriate object of secure attachment and the disposition of cultivated reliability are different, and not necessarily complementary, character traits to develop in oneself. In fact, attempting to develop both robustly may be incompatible. We may, for example, be tempted to offer too much to those to whom we share an attachment in ways that increase our risk of letting them down.

The distinct types of vulnerability that need to be reduced and their potential incompatibility give us reason to focus on understanding the various types of trustworthiness and what specifically they demand.

This also means that we would be mistaken to attempt to achieve the overarching "trustworthiness" that the social vulnerabilities account suggests. This is that to be

trustworthy is to be disposed to respond to others' social vulnerability. This is true, but uninformative and unhelpful in understanding what trustworthiness demands of us.

This also clarifies the relation between accounts in the existing literature. Our approach ought to be to first identify the underlying social vulnerability, and then to consider what feature most robustly and appropriately reduces that given vulnerability. Accounts like Andrew Kirton's and Katherine Hawley's are not in direct disagreement. They instead target and explain different forms of trust and trustworthiness. In contrast, Hawley's and Russell Hardin's accounts are in direct disagreement: they both target a form of trustworthiness and trust which are concerned with our vulnerability to other's actions for our own actions' success. Hawley's account is the more robust (though, as I will argue in Chapter 5, not as robust as the form of reliance-responsive trustworthiness which I put forwards in this thesis). We can make more progress by being aware of this, and it is the approach that I adopt in this thesis.

Whether we require a more or less robust form of trustworthiness will vary by the situation that we find ourselves in. The social vulnerability account gives us a way to understand the unity of these varying demands, while providing a structure in which we can better elucidate the specific requirements of certain forms of trustworthiness.

6.3 Discerning trustors

The variety underlying our trust-concepts also means that when our trust is well-placed will vary hugely by context. We will need theories that help us to understand what vulnerabilities we face in different interactions and what is required if we are to have that vulnerability reduced. The social vulnerability account suggests that there is useful and important work for each of us to do to become more discerning trustors. I have argued that a single concept "trust" picks out a range of attitudes. We use the term that refers to that concept for all these attitudes too. We may, therefore, often confuse one form of trust for another. I think this is quite common. You can think of how your trust in a loved one can make you likely to think she is trustworthy in a range of ways that she is not. Part of becoming a good trustor will then be learning to carefully separate our different forms of trust and recognise

when and in whom they are appropriate. This helps to explain the contagion and strange interactions with evidence we find when we trust, but I leave its development for other work.

7.0 Conclusion

The Social Vulnerabilities Account shows that there is a unity to “trust” and “trustworthiness”, but a wide variety in the forms that both can take. They centrally concern the kinds of vulnerabilities that we face as limited, interdependent and social agents. They give us a way of navigating our interactions with each other while reducing these vulnerabilities.

In clarifying this, the account helps us to better understand the complexity in what trustworthiness demands and in building trust. It suggests that trust researchers have, frequently, been talking past each other: focusing their attention on different forms of trust or trustworthiness, but thinking they were looking at the same thing. It offers a way through this, which will allow us to better discriminate between the relevant forms of trustworthiness and trust we want in specific contexts. It shows why all forms of trustworthiness and trust can matter in our everyday lives. It better explains the diversity of our experience of trusting. The kind of trust that I have with a stranger or with someone I interact with only briefly, like a barista in a coffee shop, is barely noticed and normatively thin. If I were to be upset or betrayed were the barista to fail to make my coffee correctly then this would be an overreaction. This contrasts with the dominating feeling and normative richness that is part of my trust in my partner.

Most importantly for this thesis, the approach argued for in this chapter suggests that we need to take a specific approach to understanding what trustworthiness requires. First, we need to consider our interactions with businesses and identify the dominant social vulnerability. I will argue that the dominant social vulnerability we face comes from our great reliance on business in the next chapter. I will then argue for a clearer understanding of our notion of reliance: as a relation in the world and a specific type of attitude central to our planning. In the next two chapters, I consider what feature an agent would need to have to robustly reduce these vulnerabilities. I argue that this requires a responsiveness to obligations which arise when someone is reliant on us.

3 | Reliance and Business

Chapter 3 has two aims. First, to distinguish a relation in the world (bare reliance) from a mental state which we form about that relation (relying attitudes). Second, to show that being reliant makes us vulnerable to our own actions failing, and that this is the main vulnerability we face to businesses because of our reliance on them. §1 sets out the distinction between bare reliance and relying attitudes. §2 argues that it is standing in a relation of reliance (rather than having an attitude about it) that makes us vulnerable: it leaves us at risk of failing to act as we wish. §3 defines “reliability” and shows how it reduces this vulnerability. §4 summarises how this puts us in a ‘web of reliance’. §5 argues that we are reliant on businesses, and that this is the dominant feature of our interactions with them.

*

We want to know what sort of trust we should put in business. In line with the approach argued for in the previous chapter, we need to get clear on what social vulnerability we face to businesses to understand the kind of trustworthiness that they need. Only then can we understand what kind of trust we ought to have in them. Getting clear on this vulnerability is the aim of this chapter. I argue that our interactions with businesses are dominated by how reliant on them we are. This makes us vulnerable to our own actions failing if they fail to act as we are reliant on them to act. Faced with this vulnerability, we have two choices: we can either try to make businesses more reliable or attempt to cultivate a form of trustworthiness which reduces this vulnerability. The rest of the thesis explores how we could go about doing the latter of these.

As I argued in the previous chapter, some of the limitations that agents like us face are physical. These limitations mean that much human action requires further things to happen to be successful. We need events in and states of the world to happen in ways that enable the successful performance of our actions. We need other agents to at least refrain from interfering with our actions, if not to actively cooperate with us. We are in many ways dependent on things beyond our control. We therefore have an interest in understanding how our actions are affected by things outside of our control.

Reliance seems central to how we approach these aspects of the world which are beyond our control. It is how we conceive of what happens when our own actions require other things to happen. I aim to first get clear on what reliance is. This will help us to understand the kind of relations that we stand in to those things which affect the potential success of our actions. This will help to understand the exact nature of the vulnerability which we face because of our limitations.

Despite being taken as central to trust and trustworthiness by many, reliance has not been as prominent in discussions about trust as this close relation would lead one to expect. I will argue that it can be understood as both an attitude, which is central to our planning, and as a relation in the world which exists between the successful performance of some action and some further thing that happens. This helps us to understand that the key vulnerability we face is not related to our relying attitudes, but to things happening to which our actions stand in these relations.

1.0 Bare reliance and relying attitudes

We sometimes use reliance in a way that picks out a relation that the success of some event stands in to other events or states in the world and to the actions of others, which affect that success. We do this when we speak of a car relying on petrol to run or about how an ancient Egyptian farmer may have been reliant on the Nile to inundate her fields so that she could grow her crops.

When we use reliance in this way, we can call the relation that we pick out "bare reliance" (to differentiate it from the "relying attitudes" we will come to shortly).¹

Where we are concerned with bare reliance, I use the formulation "*a* is reliant on (*b* to) *X* (to *Y*)", where *a* can be either an agent or an object in the world (as is *b*), *X* is some process, event or action, and *Y* is some further action or happening by *a*.

Reliance is always in relation to some further action or event, where this is understood broadly so as to include activities like living. For example, the car is

¹ Smith (2010) makes the same distinction, and I agree with his arguments for it. Smith's terminology is 'external reliance' for what I am calling bare reliance and 'internal reliance' for what I am calling attitudinal reliance (p.136). I am avoiding his terminology because it could lead to confusion in other contexts than those which are our primary focus. For example, it may be that some psychological states are reliant on other psychological states or that there are other bare reliance relations (and so in Smith's terminology external reliance) internal to the agent. My more descriptive terms are clearer. In Chapter 6, I argue that a relying attitude is an intention-like planning attitudes (something that Smith leaves open).

reliant on petrol to run means that there is some relation of dependency between the car running successfully and petrol. To be completist we ought really to say that the car is reliant on petrol combusting appropriately or something similar, though that is rarely necessary.

It would be nice to be able to exactly determine what kind of relation a 'bare reliance' relation is, but doing so would take us far beyond what is required for this thesis. It is enough for our purposes that it is some kind of relation of dependency such that if the event on which something is reliant failed to happen, then that something would not happen. A car without any petrol in it would not run.²

We need to be able to appropriately track actual relations of bare reliance in the world and to incorporate them into our plans. If the ancient farmer holds that it is the sun, not the Nile's floodwaters, that annually nourish her fields then her plan to grow crops is likely to be unsuccessful. The farmer needs to plan on the Nile's inundation coming at a certain time to successfully grow and harvest her crops without losing them to either heat or floodwaters. To successfully plan to grow her crops she therefore needs to be able to form some attitude towards the Nile's flooding.

We can form a "relying attitude" in order to incorporate events and actions into our plans.³ Where we are concerned with such relying attitudes, I use the formulation "A relies on (*b* to) *X* (to *Y*)", where *A* is some agent, *b* is some agent or object, *X* is some

² There may be situations such as getting to a meeting, where you are reliant on a bus turning up, but it does not matter which of several possible buses it is. I understand this as your being reliant on the disjunction of them. I think that what distinguishes bare reliance from other kinds of dependency is its relation to agency. I believe that talk of cars relying on petrol, etc., are extensions of the core of "reliance". Reliance helps us to pick out the dependency relation between action and the world. Nothing I say though depends on establishing that, and so I do not argue for it here.

³ The focus in this chapter is on reliance as it is involved in practical reasoning. That is because of the overall interest of this thesis in how agents interact and how we can improve such interactions. We may also rely epistemically. I intend the discussion that follows to extend to our epistemic agency, but I do not directly focus on this.

process, event or action by *b*, and *Y* is some further action of *A*'s, which she takes to be reliant on *X*.⁴ The farmer can rely incorrectly on the sun, or correctly on the river.

Relying attitudes and bare reliance can come apart.⁵ I may rely on my alarm clock to wake me up (that is I may have a relying attitude towards the alarm clock). I would then have an attitude that represented the world as being one in which I am reliant on my alarm clock to wake up every morning. When I have an early train to catch, I will plan to set my alarm clock for an earlier time than usual. It could be that the alarm clock has never worked. I have therefore never been reliant on the alarm clock to wake up (there is no bare reliance relation between my waking up and the alarm clock). Instead, I may actually have been reliant on my neighbour slamming her car door as she leaves for work.

This divergence of bare reliance and our relying attitudes matters because mistakenly relying on something that you are not actually reliant on can seriously upset how successfully you manage to act. If my relied-upon alarm clock does not wake me up (as it has never actually done), then I will miss my train. We usually want a clear idea of what we are reliant on. Though, there are limits to this. There are clearly many things on which we are reliant, but about which we never bother to form relying attitudes. We fail to form such attitudes because we are unaware of our bare reliance, because the thing is commonplace or taken for granted in our reasoning, or because, as above, we are mistaken about what we are reliant on or its role in our plans.

This distinction between bare reliance and a relying attitude is all that matters for our current purposes. I will return to the question of what a relying attitude is in Chapter 6. Understanding this more clearly allows us to better understand the nature of one form of trust, and I will argue that many features commonly identified with trust can be explained because they are features of relying attitudes, which form part of a complex attitude of reliance-responsive trust.

⁴ Reliance is then a three-place relation, like much trust.

⁵ Smith (2010, p.148) also point to the fact that bare and attitudinal reliance can come apart.

2.0 Reliance and vulnerability

It standardly follows that if you are reliant on something then you are vulnerable to it. You are vulnerable to your own actions failing because the thing that you are reliant on does not happen. This is – from the previous chapter – a form of agential vulnerability.

An agent or object on which we are reliant always runs the risk of providing a point of failure for (at least part of) our plans. If you have relied on your partner cooking your dinner to read this chapter but instead your partner has gone to the cinema, then you will go hungry. We are open to this vulnerability, whether or not we form a relying attitude towards the agent, object or process. Plans are often derailed by events which we are unaware were important to their success.

The level of vulnerability can vary in several ways. The vulnerability to the failure of our actions which reliance produces will depend on (a) its importance to the success of our action; (b) how many options are open to us, and (c) how important our plan is. Vulnerability can vary along any axis. (See the related discussion of the variation in reliance in Simpson, 2012, p.557.)

We may be reliant on an agent or object in a way central to the success of a plan. The event which our action is reliant on may be pivotal to its success. The Nile is Egypt's major source of fresh water and fertile silt. The success of the farmer's plan to grow crops therefore pivots on the flooding of the Nile. We are more vulnerable to these pivotal reliances because without them the entire plan fails.

Our vulnerability will also be increased the fewer options one has as to what one can be reliant on. You are less vulnerable if you have multiple potential and available options on which to be reliant. You are more vulnerable if your plan is reliant on a unique potential way to be successful. The ancient farmer's vulnerability is also high by this measure. Though it may then seem that these are identical, I differentiate them because while it does seem that something that is unique will likely always be highly important, it does not seem that what is highly important must always be unique. We may have multiple options for some highly important aspect of our plan, and this raises our vulnerability to all their potential failures, although there are numerous options.

Finally, we may be reliant on an agent, object or process in a highly important plan. That is, the plan may be particularly valuable to us (or central to other valued plans). Again, the ancient farmer's vulnerability is high on this measure: her growing crops is essential to her livelihood and her survival. Something may be highly important – and even the only thing that can make a plan succeed – but we may just not care very much about the success of that plan.

The vulnerability here is to our action failing because of our reliance. It will often be the case that our action failing has further implications for our wider vulnerabilities (including our non-social vulnerabilities). When you are reliant on a bus to run to get you home and it fails to come and you are left outside in the cold and rain, then this will be a harm to you. These further vulnerabilities are things that make the success of what you are reliant on matter to you, rather than the direct vulnerability you face because of your reliance. It will (often) matter less when something we are reliant on fails if it only matters because it will mean that we must change our plans (as when you notice there is another bus which you can get). As they standardly alter the importance of the reliance, these further vulnerabilities do matter (indirectly) for how vulnerable we are to that which we are reliant on failing.

This vulnerability is a product of our bare reliance – of what we are actually reliant on – not of our relying attitudes. This does not, however, mean that our relying attitudes do not play a role. Relying on something might mean that we become reliant on it. I can rely on the bus turning up, and in acting in line with this and walking to the bus stop and away from the train station, I might make getting to my meeting on time reliant on the bus. Nevertheless, it is because this attitude has led to my being reliant on the bus that I am vulnerable. This means that we need to ensure that we are relying well. Whenever we rely, we run the risk of exposing ourselves to a vulnerability by making ourselves reliant.

Reliance means we are vulnerable to our actions failing. This is not necessarily a social vulnerability. It is not when we are reliant on an object or natural process. It is also not for every reliance we have on another agent. We will return to how reliance can expose us to a social vulnerability (that is, how we can be vulnerable in an interaction because of reliance) in the next chapter, and how this illustrates the

difference between situations in which only reliance is appropriate and those in which trust is appropriate in the next.

Finally, it is worth considering the relationship between reliance and power. It would be more than was needed to provide an analysis of *power* here. What is clear is that whatever power is, it seems to correlate positively with others' reliance (and plausibly, though less clearly, with a greater vulnerability of others to the powerful). That someone is heavily reliant on you can be part of what constitutes "having power over" them. The more powerful an individual or institution is, then the more people will be reliant on them.

We have seen that reliance picks out either a relation in the world (*bare reliance*) or an attitude about that relation (*relying attitude*). It is the former that make us vulnerable to the failure of our actions, but our attitudes can shape on what we are actually reliant.

We are interested in how we can lessen this vulnerability. In the next three chapters, I will turn to what trustworthy feature can do this. I first want to discuss another way we can do this: by fostering reliability.

3.0 Reliability

3.1 Reliability

You are vulnerable to your own actions failing because the thing that you are reliant on does not happen. This vulnerability means that it is important for you to rely well and to ensure there are potential agents and objects to rely on. To be clear on how we rely well, we need to be clear on what reliability is.

I want to show that reliability is a feature that some entity (or process) has if it is such that it enables another entity to perform successfully. That is, it is something which makes a bare reliance relationship successful. On this view, as reliance is a three-place relation, an agent or object is only reliable with reference to another agent and the purpose for which she is relying on that agent or object. Kant is said to have taken regular walks through the streets of Königsburg. So regular that his fellow citizens were able to set their clocks by them.⁶ Kant was a reliable timekeeper

⁶ This originates in Baier (1986, p.235). We will assume Baier is correct that this is just a case of reliance. I explain why later.

to his fellow citizens because he was punctual. He may not have been reliable as a pacesetter for the annual Königsberg walking race. He may well have been a dawdling, inconsistent walker. To judge something as reliable requires only that the relied upon object or agent is able to serve the purpose it plays in someone's plans.

This is, I think, the core sense of reliability. It captures some intuitive cases of reliability. We can contrast it with a view that reliability is to have a stable disposition.⁷ This is the view that a well is reliable when it regularly gives potable water. When put like this, the competing view seems plausible. However, it makes less sense if we compare other, less functionally defined objects. If reliability is just to 'regularly or with some regularity' bring about some 'tokens of [a] phenomenon', as Matthew Noah Smith has it, then Stromboli is a reliable volcano. That, however, seems an odd use of reliable. We seem to need – as my definition does – to determine some goal or end to which a volcano is being put. Stromboli is a reliable volcano, for those who want to see a volcanic eruption. The regularity view seems plausible for the well only because "well" has packed into it a clear function on which it is generally suitable to rely on it (providing potable water). Anything that does not seem to leave open the question "reliable for what?" That this is an open question favours the definition I am giving.

The regularity view also fails to explain how reliability can be one-off. I can rely on something just once, as when I sit on a bench in a train station I am passing through. That bench is reliable so long as it holds me for the duration of my sitting on it. If it breaks when the next person sits on it, that is irrelevant to its reliability for me. Standardly, however, we are concerned with the robustness of reliability. Things

⁷ This contrasts with the view of reliability that Matthew Noah Smith puts forward. On Smith's (2010, Ft.13) view a reliable object is one where 'tokens of that phenomenon are realized regularly or with some regularity'. This seems to capture a sense, but not the core sense of reliability. As I discuss in the main text. Likewise, he defines reliability for an agent as 'she will do what she sincerely says she will do as a matter of character'. This does not seem necessary for reliability. It is unclear why she must do what she sincerely says she will do, rather than that she will just do what I am reliant on her for her to do. Smith's focus in his characterisation of reliability on regularity picks out features of objects and agents that make them likely to have more robust reliability, but, as I argue in the main text here and in my discussion of stability below, he has over-emphasised this.

that perform regularly are likely to be more robustly reliable (see below), but this is secondary to the sense defended here.

Being able to fulfil the purpose, which it is relied on to do requires an object to have an actual capacity to do the relied upon thing. To be reliable as a source of nutrition for the fields, the Nile has to actually bring nutrition to the fields. If the Nile were to flood but were to carry away soil rather than deposit sediment it would no longer be reliable to our ancient farmer, even though it may still be a regular flooders. An actual capacity to do the relied upon thing for an agent, is usually a competency in the relied upon thing. If Kant had become unable to walk, he could not have been relied upon by the townsfolk.

In the case of agents, being reliable also requires some motivation to carry out the relied-upon action or behaviour.⁸ Whatever that motivation is, so long as it brings the action about. Kant may have gone for a walk every day because he wanted to, but it could equally have been because his creditors called at that time and he wanted to escape them, or because he had a compulsion to do so. Why Kant walked is unimportant to whether Kant fulfils the bare reliance relation that the citizens of Königsburg stand in to him, when they rely on him to set their clocks.

The view of reliability for agents that this suggests is then:

- (1) competency to perform the relied upon action, and
- (2) for an agent, sufficient motivation to perform the relied upon action.

3.2 Creating reliability, reducing vulnerability

As we frequently rely on things for extended periods of time and multiple people want to be able to rely on the same thing, we have good reasons to want more robust reliability. Reliability can be robust over time, over interactions, for a range of different reliers, and over different domains of action. We therefore often try to increase the reliability of objects, processes and agents, and in doing so we mitigate our vulnerability to our actions failing. We dedicate much of our interaction with the natural world to promoting this reliability.

⁸ We can be reliable in bringing about something over which we lack control. Such as the reliably funny dinner-guest who is reliable only because of her natural pomposity.

We want to make some events and processes more robustly reliable. We can do so by increasing their capacity to perform as we require. This allows us to mitigate our vulnerability. Our farmer on the Nile is reliant on the river's inundation to bring water and nutrients to her soil. She could remain reliant on the Nile's flooding, but this will not decrease her vulnerability to its failure to do so. She may, instead, mitigate her vulnerability by directly intervening on the object (the river) on which she is reliant. She can build irrigation ditches, making the rivers sustenance of her land more reliable over time by bringing its waters into her farm beyond their annual flood. The building of such ditches may increase the cultivatable area, increasing the robustness of the Nile's reliability in nourishing land for a larger group of farmers. An effect of diverting the water may be that flooding is diverted from built-up areas, increasing the domains of action over which individuals can consider the river reliable.

We also try to make agents more reliable, though this is more difficult. We can directly mitigate for agents being unreliable by attempting to increase either their competency or their motivations to behave in the relied upon way.

We can alter the incentives open to other agents, in attempts to increase their motivations to behave in the relied upon way. For example, the *Bürgerin* may worry that Kant will stop his walks on Baltic winter days. She may then begin to offer him *Glühwein* on such days to encourage him to continue. We also attempt to alter motivations, by producing frameworks that alter the incentives for certain types of actions, be these frameworks through regulation, social institutions or formal organisation.⁹ Such systems sanction or reward certain behaviours, changing the motivational structure for agents caught up in them. Simply interacting with others may alter incentive structures sufficiently to encourage reliability.

We can also attempt to alter motivations by introducing what we standardly think of as systems of professionalisation. Here, we may try to instil an ethos in an individual and to lead to changes in their motivations. We are most familiar with this as the

⁹ I follow Luhmann's (1997, p.8) view that reliance (and so trust), regulation/law and organisation all work to attempt to ease the complexity of social interactions. See also my discussion of organising in Chs.8 and 12, §5.1.

kind of training characteristic of the “professions”: education, engineering, and legal and medical professionals. Though, such systems are obviously not necessarily restricted to professional spheres or even to the world of work.

Standardisation and qualifications are a way of increasing reliability by increasing competency. They provide a benchmark of competency against which we can consider the potential reliability of certain agents. This helps to ensure that other agents can be confident that another agent is competent in the area they are considering relying on her for. This may be why qualifications and professional certification tend to be more important in areas where there is a large gap between most laypeople's knowledge of the field or where assessing if the relied-upon agent has such knowledge would be impractical. For example, doctors' knowledge of our health is far beyond most laypeople's, but it may also be difficult to ascertain for an employer simply from interviews. We also tend to provide such institutional reassurances where competence really matters to us. We licence motor vehicles, for example, allowing us to rely on each other on the roads. Needs for recertification can help to increase how stable this reliability is. It is a central element of how we produce reliable agents and increase reliance.

Robust reliability will often correlate with stable dispositions in objects, processes or events. Kant is reliable in large part because he is a creature of habit. Phenomena which are stable or which occur with regularity will often be easier to harness for our ends. Stability is, however, option-independent in a way that reliability is not. By option-independent, I mean that increased stability does not guarantee an increase in potential courses of action for an agent. Reliability seems to me to do so. A more reliable world is one in which we can do more. A more stable world is not.¹⁰

Damming the Nile may make it more stable, but it need not increase its reliability. We need to look beyond stable or regular dispositions to understand how to lessen our vulnerability because of reliance. We need to consider the ends to which those dispositions can be put.

¹⁰ A similar point is made by Simpson (2023, Ch.1).

Developing reliability is one way to reduce our vulnerability to our own actions failing should objects, processes or agents fail to act as we are reliant on them to act. On the account I developed in the last chapter, reliability and some forms of trustworthiness which respond to the same vulnerability will often be close. They will be differentiated only because reliability is not necessarily self-regulated, nor is all reliance necessarily a social vulnerability. As I argued in the last chapter, we can engage with other agents either in a social or interpersonal interaction, or as if they are part of the background of the world in which we must act. It is the latter that happens when someone makes herself reliant on Kant's regular walking to set her clocks. There are other instances of reliance that are of this form.

As demonstrated above – unlike with trustworthiness – it is acceptable to intervene to improve, and even to maintain, reliability. In fact, it is often necessary to achieve a robust reliability. The world gives us few robustly reliable objects, processes or agents. As I said in the last chapter, these kinds of mitigation have significant costs. Producing and maintaining systems that promote reliability requires the expenditure of significant resources.

4.0 A web of reliance

Reliance picks out both a relation in the world (bare reliance) and a mental state that represents those relations (relying attitudes). Our bare reliance makes us vulnerable to our own actions failing if they fail to act as we are reliant on them to act.

We stand in a web of reliance, linked to each other by a broad range of bare reliance relations which only increase as the world becomes more interconnected. Reliability is one way of reducing the vulnerability we therefore face. It is not the only way, and it is often a costly way. We have an interest in reducing this vulnerability in other ways. That can be done through a form of trustworthiness which reduces this vulnerability. That means we want a self-regulated feature that an agent can have which robustly lessens the vulnerability we face because of our reliance on her. Pinning down that form of trustworthiness will be the job of the next three chapters. Before we turn to that, I will discuss why our reliance on business is so important in our interactions with them.

5.0 Relying on business

That our interactions with businesses are dominated by reliance and its associated vulnerability to our own actions failing if they fail to act as we are reliant on them to act is to simply make explicit something obvious. As customers and clients, we interact with businesses mainly so that we can get something or have something done, which itself plays a part in some plan that we have. We buy products from them, because we want to use those products or for the pleasure of it. We buy services from them; things which we either cannot or will not do ourselves. A product or service is bad when it does not meet the requirements that we had for what we had intended to do with it. That is, we are reliant on what these businesses provide.

Businesses play a large role in the web of reliance in which we stand. While it is difficult to quantify the power that businesses have in the modern world, it is certainly the case that we are often significantly reliant on businesses. This includes further general forms of reliance, on their contribution to our economic growth and the opportunity that we can access. We need factories, shops and other businesses to stay open to provide employment, on which the majority of us are reliant. This spreads reliance further, as other businesses and organisations become reliant on the money produced and paid out by those businesses to carry out their own activities. Likewise, we are all reliant on businesses properly handling the waste and pollutants that they produce. Some of the largest businesses (whether nationally or globally) will exercise significant power, including politically where they provide funding or influence on politicians or where they engage in regulatory or even state capture.¹¹

Businesses are not, of course, the only organisations that we are reliant on in these ways, but they are a vital one. If we lived under a different economic system, things

¹¹ Businesses arguably also hold significant epistemic power, being a major node for information. I do not directly address epistemic trust in what follows. This is in line with the approach outlined in the last chapter. Nonetheless, a trustworthy business may also need to consider its epistemic duties. This epistemic risk is not, I think, the dominant one that we face from businesses.

could be different, but most of us do not. Many of us are reliant on businesses in multiple ways and so vulnerable to failing to act as we want if they do not behave reliably.

We expend significant resources on regulating business. A key outcome of such regulation is the mitigation of the kind of vulnerability we have been discussing. This is not standardly how regulation theorists understand the aims of regulation. “Stability” has played a much larger role than reliability in discussions of why and how we ought to regulate business. Stability is not obviously valuable unless it is because it contributes to reliability.¹²

So, we ought to look to see if we can reduce our vulnerability because we are reliant on business by attempting to make them reliable. As discussed in Chapter 1, this is not the route that many regulators have turned to. We may also question whether this approach has been successful. Our reliance on businesses is clear, and it forms the dominant vulnerability that we face to them. We have an interest therefore in seeing if there is a different, trustworthy way that this vulnerability can be reduced.¹³

6.0 Conclusion

We are reliant on businesses in many ways, as consumers, clients, and as members of the public. We are vulnerable to our own actions failing if businesses are not reliable. To be reliable a business must be competent to perform the relied upon action, and (if businesses are agents) have sufficient motivation to perform the relied upon action. Reliability is one way of reducing the vulnerabilities that we face when we are reliant.

¹² I do not intend to argue for this as a justification of regulation or all that makes it good – on which see Morgan and Yeung (2007) and Sparrow (2020) – but I do note fostering reliability seems to be a good-making feature of regulation.

¹³ I note the less clear plausibility of any other form of trust as valuable in our interactions with business. It may be that we can turn to businesses to satisfy our social needs or in ways that expose us to the other social vulnerabilities discussed in Chapter 2. This seems inappropriate. We should not find attachment to a business or have our emotional needs satisfied by one. The burden of argument rests on whoever might wish to support the view that such relationships are possible. Our reliance is, by contrast, quite clear.

The next chapter turns to considering how specific kinds of interaction can lead to us becoming reliant in a way that gives rise to an obligation to act in some way.

These interactions involve other agents offering their reliability to us, as a good reason for us to act. I will argue in Chapter 5 that appropriately reasoning about these interactions and their associated obligations is what is required if we are to reduce the vulnerabilities that reliance produces in a robustly *trustworthy* – rather than a merely reliable – way, and which will suitably reduce this vulnerability we face to businesses.

4 | Reliance-Related Obligations

Chapter 4 aims to establish the existence of a form of non-moral, practical “reliance-related obligation”. In the next chapter, I will show how appropriately reasoning about these allows one to be trustworthy in a way that reduces the vulnerability we have to our actions failing because of reliance. §1 argues that there are clearly reasons for a relied-upon person to act, which are related to another person’s reliance on her. §2 distinguishes between situations where such reasons seem present and where they do not, and I suggest that this is not due to relying attitudes or bare reliance but to whether such reliance has been invited or not (§3). §4 discusses how we offer each other reasons to do things, and I introduce two norms that enable this practice: endorsement and ownership. §5 argues that when these reasons are grounded in facts about us – as they are with our reliability – then we become obligated to be reliable. §6 discusses how this suggests we ought to reason about these special reliance-related obligations. §7 attempts to show that these obligations have a rational, rather than a moral, force.

*

We are trying to find out what kind of trustworthiness we want from business. In the last chapter, I argued that reliance makes us vulnerable to our own actions failing if that which we are reliant on fails to perform as we are reliant on it to perform. As I said, this need not be a social vulnerability. It is not when we are reliant on an object, and in certain circumstances when we are reliant on another agent.

In this chapter, I will argue that certain kinds of interaction do expose us to a social vulnerability which is grounded in our broader reliance-related vulnerability. I will argue that this happens when those on whom we are (potentially) reliant offer us their own reliability as a reason for action and so engage with us in a distinctively interpersonal way.

These sorts of interactions give rise to obligations on the part of the person who offers her reliability. I will argue that when we offer our reliability as a reason for action and we can bring it about that we are reliable, then we are bound by the fact that we endorse our reliability as a good reason and that we accept responsibility for

that reason's contribution to the other's practical reasoning to be reliable. The associated obligations are reliance-related obligations.

I argued in the last chapter that we are most vulnerable to businesses because we are reliant on them. In the next chapter, I will show how being appropriately responsive to these reliance-related obligations is a form of trustworthiness which robustly mitigates this vulnerability. We first need to establish the nature of these interactions and their associated obligations.

Before we start, there are a couple of key points to remember from the last chapter. We can rely on something (have a relying attitude about it) and be reliant on something (stand in a relation of reliance to it). These can come apart, and they often do. Reliability is a feature that objects or agents have if they successfully act in the way required for the other agent to be able to successfully do whatever they are reliant on them to do. It requires an agent to be sufficiently competent and motivated to act in the relied-upon way.

1.0 Can reliance be reason giving?

I want to argue that there is a special kind of obligation which arises in relation to our reliance.¹ We can start by noting something that is generally accepted: there are reasons about when we are reliant on another.² If you are reliant on your partner moving her car first so that you can get your car out, then you have reasons to do certain things related to your partner moving her car. You have reasons to care about her moving it, to pay attention to whether she does, and to act in ways that increase the likelihood that she will. You would act irrationally if, for example, you removed her car's sparkplugs.

¹ Other accounts of reasons in the vicinity of reliance include Scanlon (1998, Ch.7) right to rely, Heape (2020) grounding of promises in reliance, and Alonso (2009) reliance-related obligations. I discuss some detailed differences with Scanlon and Heape in this chapter. Scanlon's view is actually based in assurance, not reliance. My account looks for a particular kind of assurance about reliability and thinks there are specific consequences of this for our obligations. See the next footnote for my discussion of Heape. My account of how these obligations arise provided in this chapter is not considered by Alonso. He also thinks that mutual reliance is necessary to give rise to the obligation, which I disagree with. I discuss some further differences between our views in Chapter 8.

² I am assuming here that some consideration is a reason for action if and only if it is some fact that counts in favour of or against doing something: that the tea is bitter is a reason for me not to drink it; that the tea contains caffeine is a reason for me to drink it.

There are also reasons when others rely on us. Consider the following example:

Carshare: Arturo is talking to his neighbour Bella about his arduous commute. He explains that he wishes he knew someone with whom he could carpool.

“I drive past your office everyday” says Bella, “I could give you a lift if you needed one.”

Arturo accepts Bella’s invitation to carshare. Monday morning comes and Arturo arrives at Bella’s house ready for his lift.

Arturo’s reliance on Bella seems to give a reason to Bella to act in line with Arturo’s reliance. She has reasons to give him a lift, which stem from his reliance on her.

As Alexander Heape has pointed out, we can locate a reason related to reliance in cases like this.³ We can stipulate that there is no cost for Arturo if Bella refuses.

Arturo has enough time to easily make the bus, and he is an easy-going guy who Bella knows would not be offended or upset. Bella will let Arturo down in some way,

³ Heape (2020) fails to fully distinguish between bare reliance and relying attitudes. His argument appears to rest on relying attitudes, which I disagree with as the source of such reasons (see main text).

Heape thinks that there are weak reliance-related reasons. I disagree. Heape offers the following example. ‘I express some interest in joining a friend of mine for a few days at the lake. On this particular occasion, I did not mean to convey that I was intending to go this weekend, or that I wanted to, but things get mixed up. This could happen in a number of ways: through my careless choice of words, a bit of bad luck, or an honest misunderstanding. Either way, suppose my friend comes to rely on me to join him this particular weekend. He decides to clear his schedule, to invest in a pair of fishing rods, and so on. When the weekend comes, he duly turns up at my house, raring to go.’ (p.153). I think what kind of reason Heape has depends on exactly how the misunderstanding came about. If there has been a misunderstanding or ‘bad luck’ then there has been no real offering of a reason grounded in Heape’s reliability and so, on my account, no obligation comes about. In this case, there is not even a weak reason *related to the friend’s reliance* for Heape to act. This seems right. There may be some weak reason related to the inconvenience and harm caused to his friend. There is not a clear reason related just to the friend’s reliance – as we can see if we remove these confounding elements. If Heape failed to carefully choose his words, then he may have no reason if – again – it is an honest mistake. If, however, he has sloppily chosen words then I think that Heape does not have a *weak* reason: he is obligated to act in this way. This again seems right. If Heape is ambiguous in his discussion with his friend, then it seems wrong that he can just shrug this off. He seems to owe his friend something. I think that he owes his friend his reliability (see my discussion in §6).

Heape’s arguments are about the nature of promises. I do not intend anything I say in this thesis to be an argument about the nature of promises (see also Ch.5, §5.1).

but this is inevitable because – as we discussed in Chapter 3 – when we are reliant on others and they fail, this raises the chances that we will fail to act as we wish. There is some reason for Bella to act, which is related to Arturo's reliance or how it came about.

In fact, reliance can produce very strong reasons. Consider:

Running late: One morning Bella is running late. She realises that were she to not pick up Arturo, she could reach the office on time. She thinks about how Arturo is relying on her and realises that she ought to pick him up.

In *Running late*, Arturo's reliance on Bella means that she ought not to act because of her concern about her lateness (bar some cancelling concern – such as the fact that she is running late to perform life-saving surgery), rather than that her lateness is simply outweighed by Arturo's reliance on her. Some other concern – such as being a witness to a car crash – would seem to justify her not picking up Arturo. It seems Bella has a *pro tanto* exclusionary reason. Following Joseph Raz (2002), we can understand such reasons as obligations. We can call them *reliance-related obligations*. We need to figure out what gives rise to these obligations.

2.0 Reasons from reliance

What gives rise to the obligations we face when others rely on us? If we want to take seriously that such reasons might be grounded in reliance itself, then there are three candidates: relying attitudes, bare reliance and reliability.⁴ I do not think that either form of reliance is a good candidate as the source of the reasons we find in *Carshare* and *Running late*. I am not arguing that there are no reasons grounded in others' relying attitudes or bare reliance, but I do not think these are the sources of these kinds of obligations. Instead, I will argue that what matters is that we offer our reliability as a reason to act for someone else.

2.1 Relying attitudes

Relying attitudes are not plausibly a source of reasons for the person who is relied upon. Say that Arturo thinks that it is Bella who is giving him a lift. He forms a

⁴ Some, such as Scanlon (1998, Ch.7) have been dismissive of such a possibility. I do not think that Scanlon is right, as the argument in the main text seeks to establish.

corresponding attitude: “I am relying on Bella to give me a lift.” Christian, Bella’s flatmate, drives himself and Bella to work, which makes Arturo’s relying attitude misdirected and makes Arturo reliant on Christian. First, it is unclear that Bella has reasons to act in any way towards Arturo (keeping the considerations about a lack of other costs to Arturo brought in after *Carshare* in §1 in mind). Second, there is not clearly a reason for Christian to give Arturo a lift, and certainly not one which is related to Arturo’s relying attitude.

The web of reliance in which we all stand as limited, dependent and social creatures is complex. We often form misplaced and misleading relying attitudes. They are a poor candidate for the source of reasons for others’ reliance unless we are willing for a profligate theory, which allows such reasons to arise frequently and very demandingly.

2.2 *Bare reliance*

Bare reliance is also not a clear source of reasons for the person who is relied upon. It is unclear why the fact that the success of your action is dependent on my acting in a certain way should provide me with a reason to act in that way. That will likely be overdemanding. The robustly reliable person who always acts in line with everyone’s reliance on her will be unable to pursue her own life. As there is no guarantee that others will be reliant on us in coherent ways, she may also have to be superhuman to do all the things that others’ bare reliance demands.

There may be specific cases where being reliant alone does seem enough, but they seem special. There may be situations where someone has no other choice than to be reliant on us and where the consequences of her reliance failing are likely to be severe. I can explain this (see §6.3 and Ch.6, §2.1.5), as well as why the reasons are absent when we are not in such an extreme situation. There may be some individuals whose bare reliance on us gives us reason to act, e.g., our family and close friends. If your mother is reliant on you to get to the hospital, you ought to take her regardless of anything else. These sorts of reasons only appear in specific relationships, and it seems more likely that this is because of some feature of that relationship. For example, it may be that part of a familial relationship is treating your family members’ reasons as if they were your own (at least in certain fields, cf. Westlund, 2009).

Bare reliance is a poor candidate for the obligations we see in *Running late*.

3.0 Relying with and without invitation

This leaves us with reliability. I think that reliability is the best way to understand why there are these reasons around our reliance. First, note that while there are reliance-related reasons in cases like *Carshare*, reliance is not always accompanied by reasons for the relied-upon person. It has long been argued, for example, that Immanuel Kant has no reason to remain punctual on his walks just because others rely on him (Baier, 1986, p.235). This chapter will clarify why this is.

We should distinguish between different forms of reliance involving agents. Some of this reliance is the same as the kind of reliance we have on events which do not involve agents. We are reliant on many things, some of them events involving no agents and some events involving agents. You can be reliant on the weather for that perfect beach trip. You can also be reliant on the squabbling family under the parasol beside you resolving their issues without dampening your mood. In this case, others' actions are simply other things that are happening in the world.

Distinctively, other agents can play a more active role in our reliance on them. They can invite our reliance by signalling that they will act in a reliable way. This inviting of reliance makes it a distinctly interpersonal form of reliance. Objects cannot invite our reliance, as agents can. I think that to say we invite reliance is misleading. We actually offer others our reliability, and when we do, we can obligate ourselves because we have control over whether we are reliable and we are bound by certain norms that standardly bind us when we offer reasons.

4.0 Offering reliability

4.1 Offering practical reasons

Let us first look more generally at what happens when we suggest a reason to do something to another agent.⁵ Consider this:

⁵ There are affinities between this account and the assurance account offered by Hinchman (2017). Hinchman similarly thinks that the obligations underlying trust can arise from rational pressures related to advising, and that it requires an 'appropriate responsiveness to the advisee's context-sensitive practical needs' (p.60) (which I discuss below and in the next chapter's discussion of socially situatedness). Unlike his account, I think that obligations come about only when the reason offered is to do with our reliability. There are also affinities with the literature on advice, and as far as I am aware, nothing that follows contradicts that

Train Ride: Anthony is talking to his neighbour Beth about his arduous commute. He explains that he wishes he knew someone with whom he could carpool to his office.

“I don’t know why you’re taking all those buses” says Beth, “because there’s a direct train from near yours which stops a five-minute walk away from your office.”

This scenario is best interpreted as Beth offering reasons for Anthony to change his course of action. In this, they are engaging in interpersonal deliberation. We can see how *Carshare* is analogous to *Train Ride*. In *Carshare* when Arturo raises his arduous commute, he implicitly raises a practical issue that he is facing which needs a solution. When Bella raises her own potential reliability in acting in a certain way as a response, she offers him a reason to act in a different way. Offering our reliability is the same as offering any other reason for action in interpersonal deliberation, except that those reasons are grounded in facts about us.

4.2 Practical reasons about oneself

In *Carshare* the reasons that are being offered include some reasons which are grounded in facts about the person offering them. Bella is not just responding that there is some *X* which would enable Arturo to *Y*, but, also, that *X* is something about her reliability (and so about her competence and motivation). Taking a lift with Bella is a good reason for Arturo only if – as we saw in the last chapter – Bella is competent to drive him and appropriately motivated to do so.

In inter-agential reliance, reliability is a fact about the relied-upon agent. As discussed in the previous chapter, it amounts to the fact that the agent will act or behave in some specific way because she is competent and motivated to do so. An agent can be reliable because she can control some activity in which she engages, as when an after-dinner speaker can reliably make a dinner party laugh. Alternatively,

literature – in fact, much of it complements it. I avoid the term advice because of two implications. First, that it connotes an epistemic inequality between advisor and advisee. Second, that it can imply a highly structured and sought out practice. It implies conversations, etc. As will become clear from the main text, I think we engage in a related “hunt for reasons” far more often than this would suggest, and that we do not necessarily assume an epistemic inequality. Though, I do think that this general search is a search for advice and is a social practice.

an agent can be reliable although she lacks agential control, as when the Reverend Spooner made dinner party guests laugh by misplacing his syllables.

Offering your reliability transforms the nature of the underlying reliance. The reliant person still has the basic vulnerability that we saw in the previous chapter. She remains ultimately vulnerable to her action failing because of the actions of the person on whom she is reliant. It is also clear that the two parties are engaged in an interaction: an offer of a reason. This introduces a particular social vulnerability, a vulnerability to that offer being a poor one.⁶

4.3 What happens when we offer reasons

I think that something special happens when we offer reasons about our reliability, because of some standard features of our offering reasons and the fact that we can control (some aspects of) our reliability. This is how the obligations we see in *Running late* arise. Bella has offered her reliability to Arturo. These obligations come about from the conjunction of two norms that operate when we are engaged in interpersonal deliberation about what to do. *Endorsement* means that when we offer a reason for action to someone, we ought to endorse it as a good reason (or at least as a good reason for the person we are suggesting it to). Standardly, we do not have control over whether it turns out to be a good reason. In those cases where we offer our reliability and it is a feature of our reliability which we can control, then we do. It is within our power to ensure that this is a good reason: we just have to be reliable. *Ownership* means that when we offer a reason to someone, we accept responsibility for that reason's part in the (reasonable) deliberation that leads to the other person's action.⁷ Taken with endorsement, this means we are responsible for being reliable when we have offered someone our reliability as a reason to act. When the reasons we offer are not grounded in facts about us, then these norms have only

⁶ I will argue that this allows us to understand why trust is inappropriate in the case of the citizens of Königsburg setting their clocks to Kant's walk (see Chapter 6). I am not suggesting this is the only way that a social vulnerability could be defended which is grounded in our reliance. Relying from the participant stance (Holton, 1994; Hieronymi, 2008) or taking on a commitment (Hawley, 2014) may introduce social vulnerabilities (though the former seems to overproduce occasions of trust – it is unclear that we are not relying on Kant from the participant stance, though settling that would require settling exactly what the participant stance is – cf. Hieronymi, 2022). My approach offers an especially neat way of explaining the distinction between standardly relying and that relying where questions of trust arise.

⁷ For further discussion of this reasonableness condition, see §6.1.

a limited effect. When they are grounded in facts about us over which we have control, then, I will argue, we become obligated. I will turn to establishing that in the next section. First, let me introduce these two norms.

4.4 Endorsement

I believe that the following norm operates when we discuss what to do:

Endorsement: (Standardly) if *A* suggests a reason for *B* to *X*, *A* endorses that reason as a good reason (for *B*).

Endorsement is a standard feature of our discussions. In *Train Ride*, if Anthony asks Beth “Oh, do you think that would be a good idea?”, she ought not to say that she does not. It would somehow be to mistake the rules of the game. Interpersonal practical deliberation has the goal of contributing to decisions about what someone should do. It would struggle to function without something like endorsement in place. If Beth could raise any potential reason in her discussion with Anthony, regardless of whether it was a good reason, then the discussion could not proceed without Anthony constantly asking Beth whether the reason is, in her view, a good one or not. It is a default assumption of such interpersonal deliberation that when we raise a reason to do something, we consider it to be a good reason for our interlocutor to do that thing.

I think that the existence of endorsement as operative in our discussions is confirmed by our need to take steps to cancel the implication that the reasons which we are offering are good ones. We must make use of provisos like “of course, you wouldn’t want to do this” or “hypothetically speaking”. That we must cancel such implications means that something must give rise to the implication. The operation of a norm of endorsement is a simple explanation of what does.

4.5 Ownership

Imagine that the conversation in *Train Ride* continues:

Train Ride: “I don’t know why you’re taking all those buses” says Beth,
“because there’s a direct train from near yours which stops a five-minute
walk away from your office.”

“So, you’re saying I should take the train?”, asks Anthony.

“I didn’t say you should.”

There are many interpretations of what Beth has said by saying that she has not said that Anthony should take the train. Some of these are unobjectionable. Maybe Beth and Anthony are philosophers, and she thinks he means an all-things-considered ought. Others are more objectionable and rooted in some sort of irrationality. For example, if the reason Beth is denying this is because she cannot recall having offered the reason.

I think that there is another interpretation; a good one which draws out a further norm in our interpersonal deliberation. On this interpretation, what is objectionable is that Beth refuses to take responsibility for the reasons that she has offered in the deliberation. This is, I think, because a further norm operates in interpersonal deliberation: the norm of ownership.⁸

Ownership: When B offers A a reason to X, B accepts responsibility for contributing to A’s deliberation about X-ing.

We act in line with ownership. When we suggest a reason to do something to someone, we accept that they may come back to us if the reason turns out to have been a bad one. A simple explanation is that we standardly take ownership to be operative.

Further evidence for ownership again seems present in the fact that we must seek to cancel the implication that it is in force. We say things like “I’m not advising you to do this, and...” or “don’t come running to me when...” or “I’ve always thought this a silly reason, but some people say...”. These sorts of phrases work to put a distance between the person offering the reason and any deliberative use to which it may be

⁸ Ownership is a relatively widely used term. Prominent versions include as part of the normative grounding for Gilbert (2008) joint commitments. I reject this sense of ownership in Chapter 8. Fischer and Ravizza (1998) also prominently use it to describe part of becoming responsible by recognising a moderately reasons-responsive mechanism as your own. That is also not the sense in which I am using it. Though it may follow from taking ownership of our own reasons-responsive mechanism that we recognise ownership when we exchange reasons with others.

put. We can try to engage in interpersonal deliberation while washing our hands of its outcomes.⁹

These are two norms in play when we engage in interpersonal deliberation: endorsement and ownership. Endorsement means that standardly when we offer another a reason for action, we ought to consider it a good reason. Ownership means that when we offer a reason for action, we must accept responsibility for the fact that it is a contribution to the practical deliberation of our interlocutor. We can now turn to how these norms give rise to the obligations we find in *Running late*.

5.0 Reliance-related obligations

This section will lay out what reliance-related obligations are. It will explain how they come about through the conjunction of endorsement and ownership, when the reason we are offering is grounded in a fact about our own reliability and where we have the control to ensure that reliability.

5.1 Standard operation of endorsement and ownership

We can start by considering how usually when we offer another a reason for action which is not grounded in some fact about us there are limits on what endorsement and ownership can demand.

Ownership means that when we offer a reason for action, we must accept responsibility for the fact that it is a contribution to the practical deliberation of our interlocutor. In *Train Ride*, this means that when Anthony tries to get the train and discovers that they are nearly always delayed or cancelled, Beth should accept some accountability for the fact that she suggested Anthony take the train.

This will be limited, however, by the fact that Beth cannot bring it about that trains will run on time. This constrains what Anthony can demand of Beth and what we should understand her endorsement to amount to.

Endorsement means that generally when we offer reasons for action to others, we accept that we ought to consider them good reasons. Usually, the limit of that

⁹ I think that this is possible when we offer our own reliability. We can signal – as I discuss below – that we intend for neither ownership nor endorsement to be implied. We can therefore ensure that a situation continues as one where others can rely on us, but only as if we are part of the background web of reliance.

endorsement will be some epistemic fact about the person offering the reason. That is, that she took that reason to be a good one. We cannot do more than this. We can only ask that Beth really did believe that the trains were a better option.

Standardly, therefore, endorsement and ownership have little effect on what we can reasonably demand of each other.

5.2 “*You said you would be reliable*”

When we offer our reliability as a reason, we can do something different because our reliability is a reason grounded in facts about us (§4.2).

It need not be. The Reverend Spooner may have been unable to reliably flip the start of his words in a humorous way. That is, he may have lacked the ability to self-regulate or control his reliability. In cases like this, we are in no different a situation to that just described, and endorsement and ownership are similarly constrained in their effects.

Sometimes, however, those facts about us are things that we can ensure are realised. If Bella is to be a reliable giver of lifts to Arturo, then this is something that Bella can ensure happens. Assuming she can drive (and I will say more on when we should offer reliability below), she can bring it about that she will do so and do so in a way which meets Arturo’s reliance.

That the reasons we offer when we offer our reliability are reasons grounded in facts about us changes the effects of endorsement and ownership.

When we offer a reason for action and accept responsibility for the fact that it is a contribution to the reasonable practical deliberation of our interlocutor, we can be held to account for that contribution should the action undertaken because of our contribution fail. We saw how that would work in *Train Ride*. The same applies when Bella offers Arturo a lift. It is as if she has said that she will be reliable. She can be held to account by Arturo for that reason for action (her reliability in giving a lift).

We have also endorsed that reason as a good one, and so we can be held to account more specifically: for it not being a good reason. When Bella says that her reliability in giving Arturo a lift is a reason for action, she endorses it as a good reason. Unlike in *Train Ride*, however, Bella can ensure that that reason is actually a good reason.

Taken together, ownership and endorsement mean that Bella is accountable to Arturo for her offer of reliability being a good reason for him. Arturo can hold Bella to account if she is unreliable and – if she is in fact unreliable but could have been reliable – she is also accountable for why she failed to do this.

The other side of the coin is that Bella ought to recognise that she has taken on responsibility for her reliability being a good reason for Arturo to act as he is. She can ensure that it is, and so the best way for her to behave accountably is for her to ensure that she is reliable in giving him a lift. If she does not, she violates endorsement and ownership. This is the source of the reliance-related obligations in *Running late*. We are obligated – because of ownership and endorsement – to be reliable when we have offered another person our reliability as a reason for action.

In the next section, I provide more detail on when and how reliance-related obligations come about, how we can avoid them, and how we ought to reason about them. In the last section, I turn to the question of what the nature of reliance-related obligations is. I think that they are a distinctive, non-moral, practical obligation, rooted in norms of interpersonal practical deliberation. I recognise, however, the difficulty of establishing this. For the overall thesis' argument, the acceptance of this is not vital. There is an alternative moral grounding, and if you prefer this it will result in the same practical guidance for the trustworthy business; all we will disagree on is whether they are considering putative or genuine reasons in their practical reasoning.

6.0 Reasoning about reliance-related obligations

6.1 How we offer our reliability

I have argued that by offering our own reliability as a reason for another to act, we can become obligated to be reliable. We saw in *Carshare* how this happens explicitly. Bella directly offered her own reliability as a reason for action to Arturo. Often this is what we do; we explicitly offer our own reliability as something on which someone else can rely.

This is not the only way in which we can offer another our reliability as a reason. Some cases of this are simple. We can tacitly offer our own reliability as when a friend asks if I am driving past her house, and I start walking towards the car. These

cases still express the offer we are making and so function the same as explicit offers.

There are social and personal practical norms which produce a reasonableness standard for reliance. “Reasonableness” here is a common standard across members of a society, rather than what it is subjectively reasonable for any individual to do. This condition limits what one person may reasonably rely on another person to do. These limit what we can reasonably be thought to have offered to reliably do. We are familiar with reasonable reliance, and I will offer some examples here, but I do not intend for this list to be exhaustive. What is reasonable will often depend on norms specific to a given society and so a general overview may not be possible.

A reasonable situation would be one where a colleague regularly turns up at the same spot each day, and you go and eat your lunch together. Over time, this may form a tacit offer of her reliability as a good reason for you to continue to turn up at that spot at lunchtime. This will generate a reliance-related obligation which she owes you.

There are, however, obscure and eccentric reliances which can be produced. The *Financial Times* crossword setter may regularly include a specific word in a grid in the first crossword of every month. I might decide to rely on this word appearing on the 2nd January 2025 to decide if I should propose to my partner. The word may fail to appear, and as a result I may suddenly view our relationship as star-crossed. It would be unreasonable for me to have interpreted the setter’s behaviour as an offer of her reliability with regards to decisions about my love life (though things would be different if I had a special understanding with the setter).

The reasonableness condition prevents us from becoming obligated to do things which we could not reasonably have thought we were offering to do. An important set of parameters for reasonableness are norms around privacy (which may be social or moral).

Privacy norms demarcate a sphere of activity within which others may be reliant on us and rely on us, but within which – depending on our relationship with them – it would be inappropriate to consider us obligated to act in this way unless we have explicitly invited such reliance. Say that you are the only one of your neighbours who

reliably puts the correct bin out every week. It seems right that your neighbours might rely on you to work out which bin to put out. They ought not, however, to interpret your putting your bin out as your offering your reliability to show them which bin to put out. This should not be taken as an offer because putting your bins out is a private matter (usually, in modern, Western culture at least). What actions are private is not a matter of whether they occur in public or not. This suggests a further value to privacy, which may underlie such norms. It offers us a direct refuge from the “strains of involvement”.¹⁰

Besides reasonableness and privacy, there is a final major way in which reliance-related obligations are limited. I will call this ‘due care’ following T.M. Scanlon’s terminology for a similar idea, though I do not think due care is a moral principle.¹¹ On my account, due care works by cancelling the implication that ownership and endorsement are in force. By taking such steps, we can avoid a reliance-related obligation. Consider the following:

Hurried Bella: Bella offers to start giving lifts to her colleague Arturo. Arturo agrees and they start to travel together. One morning Bella is running late and realises that were she to not pick up Arturo, she could make it to the office on time. She decides it is better to be punctual. Several hours later, Arturo arrives at the office. He charges over to Bella and says, “How could you? I was relying on you!”

Belinda: Belinda offers to start giving lifts to her colleague Albert and says “I’ll run you through the next few weeks, but I’m no taxi service and we can’t keep this up forever. You’ll need to find an alternative asap.”

¹⁰ Used originally, with a different context, by P.F. Strawson (1974, ‘Freedom and Resentment’, p.9).

¹¹ The terminology is Scanlon’s (1998). He constructs a contractualist moral *Principle of Due Care* which states: ‘One must exercise due care not to lead others to form reasonable but false expectations about what one will do when one has good reason to believe that they would suffer a significant loss as a result of relying on these expectations.’ Scanlon is concerned not with the value of reliability, but with the value of assurance. That explains the different focus of our treatments of these problems. My version of “due care” differs from Scanlon as I do not think it is a moral principle; and I think we must take due care to avoid others being reliant on us rather than forming expectations. I think expectations face similar problems as sources of reasons as I raised for relying attitudes in §2.1. (For the relationship between expectations and reliance see Ch.6).

Albert agrees, but a couple of months later Belinda is still driving him. One morning Belinda is running late and realises that were she not to pick up Albert, she could make it to the office on time. She decides it is better to be punctual. Several hours later, Albert arrives at the office. He charges over to Belinda and says, “How could you? I was relying on you!”

Arturo’s response is appropriate, Albert’s is not. The reason for this difference is that Belinda took appropriate steps to signal to Albert that she was not offering her reliability as a reason for him to pursue a course of action and so was not obligated to realise her reliability. Bella did not. Belinda exercised a form of due care; Bella did not.

Due care operates within the parameters set by reasonableness and privacy and allows us to clearly mark when we are not offering our reliability. It need not be as explicitly expressed as Belinda expresses it, but it does standardly require some action which expresses our unwillingness to be bound by ownership or endorsement of our reliability as a reason.

Due care may also allow us to override norms of reasonableness and privacy, and to make it clear that we are in fact offering our reliability. We do this to friends and loved ones, offering reliability within an intimate relationship in ways that would standardly be considered private. Specific relationships – for example, those between members of a hobbyists’ club or a specific sub-culture – alter what is reasonable. Alternatively, we may simply explicitly offer our reliability.

When we offer our reliability and so when we gain a reliance-related obligation is limited by reasonableness, privacy and by our own effort in using due care.

6.2 A default background of deliberation

If we offer reliability explicitly or in a situation which is clearly a discussion about what someone should do, it is obvious that we are deliberating and that relevant norms would be in force. Anyone who has in passing raised an issue in a meeting and found herself lumbered with the associated task will be aware of how common this is.

I think that we are in a deliberative situation in a wider range of situations. In fact, we are frequently in such situations, even as we go about our day to day lives. Absent

due care or a contravention of the reasonableness or privacy standards, we usually do take others' actions as offers of reasons for our own action. It is reasonable that, as humans, we are standardly looking to act, and we are standardly looking for good reasons to act in particular ways to achieve whatever it is that we want to do. I think that we recognise this about each other. So, we are standardly aware that others interpret – and should interpret – what we are doing as a potential offer of reliability. Many of our interactions can be understood as involving (implicitly) interactions of this sort. When you overhear someone complaining about being lost and you try a friendly smile, you are attempting to offer your reliability. When you arrive in a strange city with a dead phone and you desperately look around a near abandoned train station for someone to give you directions, you are looking for someone who is making such an offer.

It is enough for our purposes, if you accept that we can – and often do – understand ourselves as engaging in such a practice. That we understand ourselves to be acting in a default background of interpersonal deliberation may be part of the explanation for why, in some societies, we value stringent privacy norms and take such care to avoid even gestures that offer our reliability. We can understand our actions as part of a web of inter-reliance and our interactions as ways of offering our own reliability to each other. This means that ownership and endorsement are often operative, and so reliance-related obligations come about in a variety of circumstances.

6.3 Reasoning about reliance-related obligations

The aim of this sub-section is to provide more detail on how we ought to reason about reliance-related obligations. This will be important in thinking about how to be trustworthy in the rest of the thesis.

As we are offering our reliability, we will also need to consider those facts about ourselves which ground our reliability to be sure if it is something that we can endorse as a good reason for someone else. Given the discussion of reliability offered in the previous chapter, we will need to be sure about what we are competent and motivated to do. Bella should be sure that she can drive and that she will drive Arturo, before she suggests that her reliability is a good reason for him. To fail to do this is to endorse as a reason something which is not actually a good reason.

Reliance-related obligations are just *pro tanto* exclusionary reasons. I cannot fully detail what reasons may cancel them, but, as the preceding discussion has shown, that includes many social norms, especially those around privacy. It will also include moral reasons, which one should always act on rather than reliance-related obligations. As we saw, Bella can let Arturo down if she witnesses a car crash and has to help in the aftermath. For this reason, reasonableness and privacy provide the broad outlines within which we may offer our reliability and so gain reliance-related obligations. We must be aware of these and possible cultural and societal variation in them.

Further principles by which we ought to reason about reliance-related obligations derive from endorsement.¹² Recall that endorsement means that when we offer someone a reason, we understand that we should take it to be a good reason. It is endorsement that provides the push to be reliable, when we recognise that we are accountable to another for having offered our reliability as a reason.

Whether or not something is a good reason for someone to do something can change. We should be aware of whether that reason is a good reason for that person. To do so, we have to allow for the fact that that person's circumstances and character may shift, changing what a good reason is for her.¹³ It follows that we ought to be ready to update endorsement. Failures to do so cause problems. Think of the relative who you see occasionally at family events and who insists on continuing to give you advice, based on things you said back when you were a teenager. The reasons offered are not good reasons; and the problem is that the relative has not changed her view on what a good reason for you would be.

If we ought to be aware of changing circumstances for whether we continue to endorse a reason, then the force of our reliance-related obligations will shift alongside such changes. We ought to stop endorsing a reason as a good one, if the person to whom we are endorsing that reason has a change of circumstance. As our

¹² I consider some ways that a robustly trustworthy person ought to reason about these obligations in the next chapter.

¹³ There are some similarities between the account offered here and the classification of good advice and 'informed subjectivism' (Fleming, 2016) and 'appropriate responsiveness to the advisee's context-sensitive practical needs' in Hinchman (2017, p.60).

endorsement shifts, so does the requirement that we ensure our reliability is good, because we are obligated to realise our reliability in part because of our endorsement.

Consider the following variation on *Carshare*. Bella offers Arturo a lift, and in doing so endorses her own reliability as a reason for action to Arturo. It may be, however, that Arturo declines to accept this as a reason for action or accepts it as a good reason, but finds it outweighed by other reasons which means he does not accept the lift. Perhaps, all things considered, Arturo realises he would miss the time he gets to spend reading novels on his arduous commute. Bella ought to realise that her reliability is *not* a good reason for Arturo in these circumstances. She should therefore stop considering the reason to be a good one, and she should stop endorsing it to Arturo. In certain circumstances, this may involve her telling Arturo that this is the case. If Arturo has simply decided to do something else and gone off and done it, then this would not be necessary.

To summarise, whether or not our reliability remains a good reason for the other person provides a way to determine whether we remain obligated to realise our reliability.

A further principle by which we can reason about which reliance-related obligations we should act on relates to the level of vulnerability they face because of our reliability.

When we are considering the contributions which we have made to why someone has done what they have done, we can also recognise that the role our contribution's success plays in what that person is doing can vary in its importance for what they are doing. It will vary in the same ways as our reliance-related vulnerability varies – (a) its importance to the success of our action; (b) how many options are open to us, and (c) how important our plan is – because these determine the importance of a reliance. Understanding our contribution's role will be part of what it is to recognise why your reliability is a good reason for the other, because to understand what constitutes a good reason for someone you have to understand its relative importance to them. (If you are trying to determine which of “the sun shining today” and “the thesis being due tomorrow” is a good reason for me, you need to think

about what my overall plans are, and how these reasons would contribute to or undermine those plans.)

Say you could offer your reliability to two people who are having birthday parties on the same night. You ought to decide by considering how important your reliability is as a reason for each of them.

The importance of that reliance within a plan will vary largely by how necessary the relied upon event or action is for the reliant individual to successfully act as she intends. One of the birthday parties may be of your close friend and the other of a colleague. Your attendance is likely more important for your friend's plan to come about successfully.

One of the parties may have 50 attendees and the other only three. Not attending the larger party may lead the reliant individual to have to alter her plan somewhat, but this will be only a minor difficulty for her, because there are readily available alternatives for her to be reliant on. That is not the case for the person having the smaller celebration.

Finally, this will partially depend on how much they care about their plan. You may know that one of your potential hosts cares hugely about her birthday, while the other is holding a party on sufferance. The importance of another person's plan will partially depend on her own values, though certain kinds of plans (such as plans to continue to live or which are necessary to secure the means to living well, such as employment) will be important for any reasonable individual.

When deciding which reliance-related obligations we ought to take on and which we ought to prioritise, we should be aware of the relative importance of our reliability to the other person.

This may seem to suggest that there are circumstances in which we have to offer others our reliability. This may seem the case if, for example, we pass someone injured on an otherwise deserted street. I do think that we cannot avoid offering our reliability here, but I cannot justify this "internally", so to speak. That is, from within the justification of reliance-related obligations I have offered. To do so, I would need something stronger than endorsement. Something like:

Endorsement plus: (Standardly) if *A* believes there is a good reason for *B* to *X*, *A* ought to endorse that reason as a good reason (for *B*).

Endorsement (§4.4) is much weaker than *Endorsement plus*. *Endorsement* simply says that any reason we do offer, we ought to take to be a good reason. *Endorsement plus* would mean that if we are aware that we believe that there is a good reason for someone to do something, then we ought to tell them. This seems problematic. For example, it sits uncomfortably with respect for another's autonomy. It also does not seem necessary to get interpersonal deliberation off the ground. Fortunately, I do not need to defend it.

I do though think that there are situations in which we cannot avoid offering our reliability. These arise because there are more important principles which make it inappropriate for us to exercise due care. As I have said, due care gives us a way of altering to whom and for what we offer our reliability within other norms, which set the bounds of reasonableness. My discussion so far has focused on the negative side of these norms; that is, what they rule out by classing as unreasonable. There are things that it is unreasonable to not do. Certain moral and legal norms require action, for example, including those which require us to help others in need.

As such norms require us to act, we ought to abide by them. As we ought not to fail to act as they require, we cannot reasonably signal to another that we will not act as they require. So, they override any ability that we have to exercise due care. In effect, they force our hand. Others can reasonably take us to have offered our reliability to act as they require.¹⁴

When considering how to exercise due care and which of our reliance-related obligations to follow, we should think about how good a reason our reliability is, why it matters and so is a good reason, and whether there are further requirements we face which stop us from failing to offer our reliability or which contradict it.

¹⁴ This clarifies the situation described in *Stranger Danger* in Chapter 2. We have a moral (and legal) duty not to harm others. We therefore reasonably take others to have offered their reliability to do so.

7.0 *Non-moral, practical obligations*

For the arguments of the next chapter and the rest of this thesis, all that matters is that you accept the existence of reliance-related obligations, as I have argued for them.

I want to defend a further claim. That these reliance-related obligations are distinctive non-moral, practical obligations. I want to defend this claim because I think it provides a more complete explanation of what reliance-related obligations are like. I am concerned about cases like this:

Bunny and Clive: Bunny and Clive are two bank-robbers. Their outfit operates with Clive standing lookout at the entrance to the bank vault, while Bunny loads the loot into a bag. Clive has done this on countless previous jobs.

Bunny and Clive are robbing the National Bank. Clive is standing lookout when a security guard turns the corner. They clock one another and Clive prepares to knock the guard out with a cosh. He considers fleeing instead.

Intuitively, Clive will wrong Bunny if he does so. I think that this wrong is based on a reliance-related obligation. Potential moral explanations for reliance-related obligations would, I think, struggle to explain *Bunny and Clive*.¹⁵ If we look for moral principles that underlie such obligations, we face a problem. We surely cannot be under a moral obligation to harm innocent people, so that we can perform some further moral wrong.¹⁶

¹⁵ For example, those that tie it to moral principles such as T.M. Scanlon's (1998, Ch.7) view that it is *assurance* rather than reliance that binds in such circumstances. On Scanlon's view if assurance is itself a moral principle, it cannot give rise to an obligation to carry out an immoral act. I concur with other views that see a demoralized normative basis for trust (Hinchman, 2017; Bennett, 2021). My own justification is different but shares the features of these accounts in locating it in alternative rational pressures. I take the source of these reasons to be our engagement in a social practice of offering reasons, which requires the norms of endorsement and ownership to function and so requires participants in it to endorse these norms.

¹⁶ This is, of course, debated. I follow views in discussion of immoral promises, on which promises cannot be morally binding if they are promises to do an immoral action.

The best moral explanation here would, I think, have to say that Bunny and Clive are mistaken about their reliance-related obligation. If that is right, then Clive has only putative reasons, which would stop him from considering fleeing. It would also follow that we are mistaken to think that he wrongs Bunny.

If that is the explanation that you prefer, that is fine. The actual actions required to be appropriately responsive to reliance-related obligations (and so, given what I say in the next chapter, of being trustworthy) will be the same. We will have a minor disagreement on exactly how someone's practical reasoning should go to get them to that action.

There are available moral explanations which can be fitted into my account. It may simply be that ownership is a form of moral responsibility for reasons that we offer, or a moral norm related to advising.

I think that Clive has an objective reason not to flee and really would wrong Bunny. He would also be untrustworthy in doing so (though, someone who is trustworthy in the robust way we will be interested in would not be in Clive's situation). I am going to try to briefly defend why.

First, I note that the normative grounding of ownership or endorsement will be the grounding of our reliance-related obligations. If ownership or endorsement plausibly have a practical rational force, then so would reliance-related obligations.

I think that endorsement is plausibly a pragmatic principle that we accept to get deliberation off the ground. As I have suggested, without endorsement we seem to simply be able to offer any reasons we fancy, in a way which would severely curtail our ability to discuss what we plan to do.

Ownership has, I think, a similar status. It enables us to get interpersonal deliberation going. If you did not have ownership, individuals could simply renounce the reasons that they have offered. This removes any ability for individuals to exchange reasons, while ensuring that they conform with things like endorsement. The exchange of reasons would – without ownership – be unworkable. If we wish to engage in it, then we must accept ownership much as we must accept endorsement.

If ownership and endorsement are both simply principles we must accept to engage in interpersonal deliberation, then the obligations derived from them would have this distinctive normative grounding. It may be worried here that this does not give us enough. It seems that individuals can escape the force of such norms by refusing to engage in interpersonal deliberation, and it is unclear that by doing so they do anything wrong. It is important to be clear on the limits of this objection. If they engage in interpersonal deliberation and do not accept the principles that enable it, then what they do is something wrong. They fail to play by the rules of the game. It is true, however, that some individuals may simply choose not to play the game. I am not convinced that individuals who choose not to play the game have done something wrong. As I have said before, it is difficult to avoid social situations and once we are in a social situation, it seems that interpersonal deliberation will often come into play. It may though be that there is nothing wrong with the person who does not wish to get involved in that deliberation, and who in order to do so never offers a reason (related to her reliability or not).

You may not agree. I think the explanatory advantage favours my route, and that there is value in being able to explain the wrong of immoral trustworthiness. We can therefore capture organisations like the Mafia. It will change little if you disagree. In the next chapter, I will argue that robust trustworthiness requires appropriately reasoning about our reliance-related obligations, and – as we have seen – this requires acting in line with moral principles and reasoning about our reliance-related obligations accordingly. So, we will all end up in the same place practically.

8.0 Conclusion

In this chapter, I have explained the reasons that we find to act as others rely on us to act, to come from whether we have offered that person our reliability as a reason for action. When we do, we are bound by the fact that we endorse our reliability as a good reason (endorsement) and that we accept responsibility for that reason's contribution to the other's practical reasoning (ownership). This gives rise to reliance-related obligations to be reliable.

I have argued that we frequently recognise that we are in situations where we can offer our reliability as a reason to another, and that these are limited by considerations of what is reasonable and especially consideration of what is

reasonably private. I have argued that we can further avoid offering reliability – or offer it in non-standard ways – by exercising due care.

I have argued that when reasoning about these reliance-related obligations we need to pay attention to whether our reliability remains a good reason for the other person, how much it matters to them and so how good a reason it could be, and whether there are relevant overriding moral considerations which should either stop us from exercising due care or stop us from offering our reliability.

Finally, I have tried to motivate the view that reliance-related obligations are distinctive non-moral, practical obligations. Though, not much rests on this being accepted, so long as you recognise their existence.

In the next chapter, I will argue that we can be trustworthy by carefully reasoning about our reliance-related obligations. I will argue that this is the kind of trustworthiness we want if we are to reduce the vulnerability that we have to businesses because of our reliance on them.

5 | Reliance-Responsive Trustworthiness

This chapter aims to determine what the form of trustworthiness is that would reduce the vulnerability we face to businesses because of our reliance on them. §1 lays out some desiderata this account aims to explain. §2 rejects three existing forms of trustworthiness. §3 explains how a responsiveness to the reliance-related obligations achieves our desiderata better than competing accounts. §4 lays out more detail on how a trustworthy individual would be. §5 considers some final advantages.

*

We now have everything in place to explain the form of trustworthiness which would robustly lessen the vulnerability that we have to businesses because of our reliance on them. I have argued that we can offer our reliability as a good reason for action to another person. As a result of norms which we accept as part of our interpersonal practical deliberation, this obligates us to ensuring that our reliability is a good reason. That is, we have reliance-related obligations. That is what we established in the last chapter.

In this chapter, I will show how an appropriate responsiveness to these obligations best explains the form of trustworthiness which we need to robustly reduce the vulnerability we face when we are reliant on someone. This is the kind of trustworthiness which we require of business because this is the kind of vulnerability which dominates our interactions with them. I suggest that this can be limited and one-off, but it also has a more robust form if it focuses on appropriately responding to and incurring reliance-related obligations. Understanding reliance-responsive trustworthiness in this way can help us to address some key problems raised in the philosophy of trust, with which other accounts struggle. To demonstrate this, I need to introduce the associated form of trust, which I do in Chapter 6. This form of trust is a relying attitude plus a judgement of trustworthiness.

This chapter should be read in conjunction with the arguments in that chapter. Together these chapters complete the account of trust offered in Section II and show the viability of the approach argued for in Chapter 2. This chapter focuses on

justifying reliance-responsive trustworthiness within the social vulnerabilities approach that I have adopted. The next chapter also shows how this specific account can solve many problem cases in the philosophy of trust and so is preferable to existing accounts. If you disagree with the social vulnerabilities approach, hopefully you will agree with the advantages of adopting this specific account and the independent motivation it provides for the reliance-related obligation approach to trustworthiness.

It is the form of reliance-responsive trustworthiness set out in the current chapter which forms the focus of the discussion in Section III: where we look at how organisations and businesses can be trustworthy in this way.

1.0 What we need this account of trustworthiness to explain

1.1 Reliance and vulnerability

I have already discussed the kind of vulnerability to which our reliance exposes us.

When we are reliant on others, their behaviour determines whether our own actions succeed or fail. If you are reliant on your partner getting back from work on time so that you can leave your child with her and see a friend, then her turning up an hour late directly impacts the success of your plans. We are harmed when this happens because our agency is undermined. Our limited, interdependent embodiment means that such vulnerability is difficult to avoid. It is being reliant (a relation in the world) which matters for whether we are exposed to this vulnerability. This is the case regardless of whether we form a relying attitude (conceive of that relation), though relying attitudes can produce bare reliances, when we come to be reliant on someone because we think we are.

This vulnerability can vary in three ways. We can be more or less vulnerable because the action we are reliant on is more or less central to some plan of ours, because there are more or fewer options open to us, or because the action we are reliant on features in a more or less important plan. This reliance-based vulnerability dominates certain interactions, including those that we have with businesses (as I argued in Chapter 3). Following the approach that I argued for in Chapter 2, we want a form of trustworthiness which reduces this vulnerability.

1.2 Maximal responsiveness is overly demanding

In Chapter 2, I argued that forms of trustworthiness could be more or less robust.

That is, they could mitigate a kind of vulnerability for a greater or lesser range of agents, domains of action, and time. Businesses actions often affect a great range of agents, span a range of domains of action, and businesses can continue for significant periods of time. We therefore do not just want a form of trustworthiness which mitigates the vulnerability we gain through reliance; we also want this feature to be robust.

An obvious first candidate would then be that an agent simply needs to have some feature which would maximally reduce the chances that someone's actions fail because of her reliance on that agent. We could look at a form of trustworthiness which required a simple responsiveness to others being reliant on us. That would be something like the following:

Maximal Responsiveness: *B* is disposed to behave reliably whenever she notes that *A, C...N* is reliant on her.

Maximal responsiveness is, however, an undesirable form of trustworthiness because it is overly demanding, in a way which we should not require trustworthiness to be.¹ Even if we limit the relationship to just two agents, the demands could quickly become excessive; especially as we can become reliant in quite eccentric ways. In a more complex world, someone who is maximally responsive to others' being reliant on her would be likely to have to give up on any attempt to pursue her own life. Being maximally responsive may also be practically impossible. There is no guarantee that others will become reliant on us in coherent ways and so meeting the demands of maximal responsiveness could require acting in incompatible ways.

¹ The problems that affect maximal responsiveness also affect a conception of trustworthiness as something like "a robust disposition to respond to manifest reliance" (McGeer, 2008; McGeer and Pettit, 2017). Where being 'manifest' amounts to the reliance being something of which the relied-on person is aware, then it is just maximum responsiveness. If we understand such 'manifesting' to have to occur through a trusting or relying attitude, then it faces the problems that relying attitudes caused as a source of reasons, which I discussed in the previous chapter.

So, we need a robust form of trustworthiness which reduces the vulnerability we face to each other because we are reliant. We need to find an account which balances this robustness against demandingness, and which provides us with a reasonably demanding, robust trustworthiness.

2.0 Current accounts of trustworthiness

We can begin by looking at some existing accounts of trustworthiness, which build on behaving reliably. These may be able to provide us with a form of reliance-responsive trustworthiness which is robust, but not overly demanding, and which reduces this vulnerability to our actions failing because of our reliance.

There are three currently dominant accounts of reliance-responsive trustworthiness in the literature. One is Russell Hardin's (2006) encapsulated interest account, one is that of Karen Jones (2012) in her more recent work, and the other is that of Katherine Hawley (2014; 2019).² All three involve a responsiveness to the kind of vulnerability with which we are concerned. All three offer a way to discriminate between reliance as we have it to external events and reliance as it occurs in interactions.

2.1 Trustworthiness as encapsulated interests

The encapsulated interest account suggests that trust is possible where two agents' individual interests overlap, so that they each have a reason to act in a certain way (Hardin, 2006). This can happen coincidentally (for example, you may be trusted to give a colleague a lift because he wants to go home and because you are looking to car-share to reduce your carbon footprint) or because an agent becomes motivated to 'encapsulate' the interests of another agent into her own (perhaps out of concern for another agent, because of fears about their own reputation, or – usually – to maintain the relationship with the agent).

² I note that while it has become less prominent in the philosophical literature concerned directly with trust; Hardin's account remains highly influential in other spheres, especially in political theory and science where it has formed the basis for the approach taken by Warren (2017). Its lack of appeal is not then universal, and it seems recognisably trust-like to others. I think that this also supports the observations made in favour of my approach in Chapter 2: proclivities for specific situations and subject-matter may distract (mainly moral) philosophers from other forms of trust.

The encapsulated interest account does reduce the vulnerability in which we are interested, but it does not do so in a particularly robust way. As Karen Jones (2012) has argued, the encapsulated interest notion of trustworthiness seems to allow individuals to be trustworthy even though they can behave in intuitively untrustworthy ways. For example, two individuals may be involved in a mutually beneficial relationship. One supplies the other with clothes at a low price; and the other buys them at a sufficient price to give the supplier a profit and to profit herself. The supplier may be in great demand because of her low prices compared to competitors. The buyer may therefore have an interest in increasing the amount she pays the supplier. That way she can maintain their relationship. On the encapsulated interests account this is sufficient for the buyer to be trustworthy for the supplier in their interaction.

Trustworthiness on the encapsulated interest account fails to be robust enough in a situation like this: a new supplier enters the market who is even cheaper than the old supplier. The buyer therefore no longer has any reason to wish to “encapsulate” the interests of the old supplier. The buyer can then cease to do so. Intuitively, this kind of switch is not trustworthy. If the supplier had trusted the buyer, her trust would have been broken. So, the encapsulated interest has failed to describe when we ought to trust and the form of trustworthiness it describes is not robust over changes in circumstance.³

The encapsulated interest account fails to explain why changes in robustness in different contexts may be acceptable. We want a form of trustworthiness which is

³ From presenting a version this argument at workshops and conferences, a significant number of people in the audience have disagreed that the buyer is untrustworthy in this case. Though this has not been the view in the literature, (e.g., Jones, 2012, p. 70). The approach I have been pursuing here offers a couple of ways to explain this; these are not open to the encapsulated interest account, and I think the ability to explain the clashing intuitions here is a further benefit of my account. It may be that the intuition kicks in only if we have a particular “robustness” of trustworthiness in mind. Alternatively, my account may allow for why this appears trustworthy in this context: it may be that we can appropriately signal in competitive business transactions the limits of our reliability, and so we can limit the scope of our reliance-related obligation, and so the behaviour is appropriately trustworthy. We would need to further describe the case for this to be clear. My account allows for the ambiguity here, whereas the encapsulated interest account does not: on that account, the buyer is trustworthy. Thank you to various audiences including those at the *Society for Applied Philosophy* and *Association for Social and Political Philosophy* annual conferences for pushing me to think further about this issue.

less changeable and so provides a more robust reduction in our vulnerability because of our reliance on each other.

2.2 “Counting-on” each other

Jones has offered her own account of a reliance-responsive trustworthiness. She considers trustworthiness to be the following:

‘*B* is trustworthy with respect to *A* in domain of interaction *D*, if and only if she is *competent* with respect to that domain, and she would take the fact that *A* is counting on her *were A* to do so in this domain, to be a *compelling* reason for acting as counted on’

“Counting-on” trustworthiness may appear to avoid the problems of maximal responsiveness. It is not any bare reliance to which we must be responsive, only those in which someone could be “counting-on” us. The problem is that it is unclear that counting-on trustworthiness does not just collapse into maximal responsiveness.

Jones’s account also leaves the notion of being “counted-on” under-described. It seems to be a normativity-laden reliance. If this is meant to be a bare reliance, then it appears to just be a form of maximal responsiveness. In fact, it may be even more demanding as it incorporates not only actual but potential bare reliances.

Things do not go better if we interpret “counting-on” as some sort of relying attitude. As we have seen, relying attitudes can be easily misdirected and misplaced. A responsiveness to these attitudes would often be overly demanding. (We saw examples of this in Ch.4, §2.1.) Its results seem even more counterintuitive if we consider trustworthiness.

First, there are all the cases where someone could mistakenly rely on us. This would make the account even more demanding than if counting-on is a form of bare reliance.

Further odd results arise in what the account seems to require of trustworthy people. For example, you may bet that your friend will be late to a dinner. You then appear to be “counting-on” her to be late. Your friend is capable of being

late and cares about you, but it seems odd to suggest that part of your friend being trustworthy would be that she should turn up late to dinner. Yet that is what the counting-on account seems to suggest.

There is a further step in Jones' (2012, pp.73-77) account, which suggests that to be 'richly' trustworthy we ought also to signal our potential trustworthiness.⁴ According to Jones, the richly trustworthy individual should 'signal' to other individuals the fact that they can be "counted on" to do or not do certain actions. That is, she must try to make others aware of what she will or will not do, if she is to be trustworthy. This also seems odd in the scenario just described. Your friend surely does not fail in any way to be trustworthy, even if she has failed to signal to you that she could not be "counted on" to act as she would need to act, so that you could win your bets.⁵

The focus on "counting-on" is, I think, a remnant of a trust-first approach. It still attempts to embed trustworthiness in something like trust. It gets us close, but it fails to give a form of trustworthiness which gives us exactly what we want. It is likely overly demanding were it fully unpacked and as it stands produces counterintuitive results. We have other candidates for normatively-laden reliance – commitments and reliance-related obligations – let us turn to them.

2.3 The Commitment Account

Katherine Hawley's influential commitment-responsive account offers a promising form of trustworthiness for our purposes (see Hawley 2014, 2019). For Hawley, the normative force of trust comes from promise-like commitments between two individuals. Trustworthiness is a property of a trusted agent, who appropriately incurs, reasons about and keeps those commitments. Commitments can be kept for

⁴ There are parallels between her idea of "signalling" and my own understanding of offering reliability which I outlined in the previous chapter and discuss further below. I disagree with Jones however that full trustworthiness necessarily requires taking steps to ensure that others know that you are trustworthy, and so come to trust you, as I discuss in the main text.

⁵ My own account handles this case by saying that your friend has never had a reliance-related obligation to you to turn up late to dinner, because she has never offered you her reliability to do so. It is not that she needs to have signalled she would not, because this would not be something that it was reasonable to assume she had meant to signal (except if there were some special understanding between you).

a range of reasons. You might agree to give a colleague a lift, out of goodwill towards the colleague: perhaps a concern for ensuring they have got home safely and easily. As we will lack our own reasons for fulfilling all of the commitments we might acquire, Hawley suggests that to be robustly trustworthy one has to develop a disposition to keep trust-commitments. As such commitments are likely to be multiple and perhaps contradictory, she also suggests that a robustly trustworthy individual should take steps to ensure that they only take on commitments they can fulfil (Hawley, 2019, Ch.4).⁶

These features of Hawley's account are promising. If the disposition to fulfil these commitments ensures that the trustworthy individual acts as others are reliant on them to act, then it reduces the kind of vulnerability in which we are interested. The requirement that we take on only those commitments which we can fulfil helps it to overcome the problem of demandingness.

I think that my obligation-based account is a stronger alternative to one based on commitments. I will first introduce what this requires, and then I will show how it can better handle some complex cases, explaining what trustworthiness requires while satisfying the desiderata for an account set out above.

3.0 Responsiveness to reliance-related obligations as trustworthiness

In this section I outline how reliance-induced obligations can help us to understand trustworthiness and what it requires of us.

Recall that to be reliable requires only that an agent is:

- (1) competent to perform the relied upon action, and
- (2) sufficiently motivated to perform the relied upon action.

In one-off interactions, this may be enough to reduce our vulnerability because of our reliance. In fact, there will be a form of trustworthiness which is very close to reliability. We can call this "particular" trustworthiness because it allows one to be trustworthy but in a limited way. That is, only for a limited scope of agents, domains or time. We may still be interested in such trustworthiness, as I discussed in Chapter

⁶ Hawley's suggestion mirrors my own. The difference between our accounts is the nature of the normative grounding of trustworthiness. My account does not require uptake.

2. This may be enough when you arrive at a train station late at night and look for directions.

Particular Reliance-Responsive Trustworthiness:

B is trustworthy to *X* for *A* (at *t*) if and only if:

PTY1: *B* is competent to *X*, and

PTY2: *B* is sufficiently motivated to *X* in a way that fulfils her reliance-related obligation to *A* to *X*.

You look desperately around the station concourse and a stranger catches your eye and smiles. You, correctly, interpret this as an offer of her reliability to try to help you. She has a reliance-related obligation to you to be reliable and – so long as she will act on it, whatever her own reasons (it may be that she just wants to avoid being approached by another, less salubrious lost traveller) – she will be trustworthy for you for that interaction. We would be wrong to dismiss such cases of trustworthiness.

We, though, are looking for something more robust, something like:

Robust Reliance-Responsive Trustworthiness: *B* is robustly trustworthy to *x*, *y*, ..., *n* for *A*, *B*, ..., *N*, if and only if:⁷

TWY1. *B* is competent to fulfil those reliance-related obligations she produces,

TWY2. *B* is disposed to fulfil her reliance-related obligations, where she lacks further reasons for acting in line with them,

TWY3. *B* ensures that she does not offer reliability where she would not be able to fulfil the resulting obligations (in other words, she adheres to due care).

⁷ It may be more natural to express this as “*B* is trustworthy”, but it would be a mistake to think that this is a form of one- or two-place trustworthiness; it is simply that the scope of the domains of action and potentially trusting individuals has been extended.

The competence requirement rests on the fact that any reliability which an agent cannot actually see through would not mitigate *A*'s vulnerability to their failure to *X*. If you cannot drive, you cannot be trustworthy as a lift-giver to your colleagues.

The second condition provides the robustness we require to ensure that our vulnerability from reliance is not regularly harmed. As with the commitment account, this robust form of trustworthiness allows for multiple potential motivations to be trustworthy. You can appropriately take on a reliance-related obligation so long as you will fulfil it and all that this requires is that you are competent and have a sufficient motive so that you are reliable. It does not matter if you are disposed to offer your colleagues a lift because you care for their getting home safely, or because you want to reduce the individual emissions footprint out of concern for climate change, or even because you want an excuse to leave early and know that one of them has a good reason for going before the end of the event. Many of the motivations suggested in motive-based theories are particularly likely to dispose an agent in the right way (Horsburgh, 1960; Baier, 1986; Jones, 1996; Preston-Roedder, 2017).

For the robustly trustworthy person, however, these are unlikely to always be sufficient. A person of goodwill may, in a small community where individuals are close, share values and pursue similar ends, be able to always find herself relied on to only have to act in ways which her goodwill disposes her to do. Such simple situations are not how most of us live. In a complex and inter-connected society, we are all likely to face reliance-induced obligations that we lack independent reasons to fulfil. If we are to fulfil them nonetheless then they must independently motivate us.

The third condition requires that an agent offers her reliability appropriately. She can do this by exercising due care, as I laid out in the previous chapter. This requires that she does not just cultivate a disposition to fulfil all those reliance-induced obligations that she acquires. A robustly trustworthy agent must also take steps to ensure that she can fulfil all the reliance-induced obligations that she acquires. In other words, she should only offer the reliability that she can actually fulfil. This avoids reliance-responsive trustworthiness becoming overly demanding.

This is common to the accounts of trustworthiness which we just considered from Jones and Hawley. Understanding trustworthiness as reliance-responsive trustworthiness has an advantage over those accounts. It gives us a clear and principled guide to how we ought to reason about the varying demands of trustworthiness, which is absent in those accounts, and allowing us to take a principled route between robustness and demandingness. My account provides a principled way to prioritise and reduce our reliance-related obligations. While it is not apparent how the trustworthy person ought to pick between her various commitments or respond to various levels of someone “counting-on” her.

I will first explain how robust reliance-responsive trustworthiness requires a trustworthy agent to think and act, which demonstrate how reliance-responsive trustworthiness offer a more robust form of trustworthiness than the commitment account. In Chapter 6, I then show how it can solve some key problem cases from the philosophy of trust.

4.0 How to reason as a robustly trustworthy person

To be trustworthy in the way described by robust reliance-responsive trustworthiness requires an individual to reason in a particular way. It will require you to appropriately reason about your reliance-related obligations. I discussed this in detail in §6 of the previous chapter. Here, I will focus on the specific ways of thinking that will help someone meet the conditions laid out above.

4.1 Understanding oneself and your social situation

First, a robustly reliance-responsive trustworthy individual will need to have a certain self-understanding as an agent who is situated in a wider web of reliance and is a potential source of reasons for others. We can call this a socially situated perspective, which we can adopt about ourselves and our actions. An individual who understands herself as socially situated will be aware of how her actions relate to the actions of others, that others will be responsive to how she acts in choosing what to do, and that how she acts will play a role in what others can or cannot do. We are familiar with viewing ourselves as socially situated and we can develop this perspective. We teach it to children, for example, when we point out how their

lateness may inconvenience others.⁸ The robustly trustworthy individual will modify her reasoning and actions in light of her social situation.

She will also need a good understanding of what she is capable of and what her motivations are. She will need this so that she can ensure that she is taking on appropriate reliance-related obligations. Whether your reliability is a good reason for someone else depends – in part – on whether you can actually be reliable. The robustly trustworthy person will have to be sure of this. This aligns with things that sit uncomfortably with trustworthiness. Overpromising seems, intuitively, to undermine trustworthiness. Often the cause of overpromising is a misalignment between one's actual capabilities and what one takes oneself to be able to do (a misalignment that may come about through good intentions, like fear of letting someone down or a keenness to help others). The robustly trustworthy person will avoid this.

4.2 Understanding the force of reliance-related obligations

The robustly trustworthy person will recognise her reliance-related obligations as giving her *pro tanto* exclusionary reasons. This will ensure that she acts so as to fulfil those reliance-related obligations where she lacks other motivating reasons. It will also ensure that she does not act because of some other excluded reasons, where she recognises that she has a reliance-related obligation.

She will also recognise that these can be cancelled by more important reasons including some moral, legal and social norms. This means that in order to appropriately reason about how to offer her reliability, the robustly trustworthy individual will need to be sure of the moral, legal and social norms which are in place and constrain her offers. She will know there are things that she ought not to offer to do and things that she cannot avoid offering to do.⁹

⁸ This is, I think, what Jones (2012) gets right in her discussion of being aware that others are "counting-on" us. My account maintains this advantage but is not as unclear nor as demanding as that of Jones.

⁹ To be clear, a more limited form of trustworthiness may be possible where we offer our reliability in – say – an immoral way. Given that a robustly trustworthy person will only take on those reliance-related obligations that she can fulfil, she should avoid immoral obligations. That is because if she is reasoning appropriately, she will realise that she should not fulfil

This is an advantage of understanding trustworthiness as responsiveness to reliance-related obligations. The commitment account faces a challenge in that it allows for a form of trustworthiness where we simply avoid taking on any commitments (Hawley, 2019). Some inveterately immoral individuals may be able to fulfil my account of trustworthiness in this way. If you cannot act in a reliably moral way, then you should make it clear that you will not. That will though be rare. Usually, it will not be reasonable for us to signal our moral unreliability and so we will not be able to avoid offering it. So, we will gain a reliance-related obligation, and one with the added force of the associated moral principle. This provides a minimum bar, which a robustly trustworthy individual will have to meet. This minimum bar is not open to the commitment account because one can act in a moral way without taking on a commitment. Someone aiming to be robustly reliance-responsive trustworthy meanwhile will not be able to avoid accepting at least these minimal set of reliance-related obligations related to moral demands.

4.3 Exercising appropriate due care

Within this background of reasonableness, the robustly trustworthy individual will also take care to exercise due care. This will be helped by her clear sense of what she can and cannot reliably do. She will be careful to clearly show when she is not offering her reliability. This again aligns with features that are intuitively incompatible with trustworthiness. Someone who equivocates about what she will do fits our pre-theoretical picture of someone who is not trustworthy. The friend who never says for sure that she will not come but never quite says she will is not trustworthy. The robustly trustworthy person will clearly express what she can or will not do. She will also recognise that where she has been unclear – and in doing so has offered her reliability – that she ought to perform the required action.

Generally, the robustly trustworthy individual will not have to signal her reliability. This is because the norms that underlie reliance-related obligations do not force us to tell everyone when we are reliable, they just constrain what we can do once we tell people about our reliability. This contradicts the proposal Jones outlines, which I

those obligations – but in this way she will be untrustworthy, and so this will reduce her overall trustworthiness.

discussed in §2.2. I think that the reliance-responsive account can help us get clear on what is attractive and what is wrong about Jones' account.

What the idea that signalling in this way is part of the kind of trustworthiness that we are interested in gets right is that the robustly trustworthy person will recognise that her own actions frequently do offer her reliability to others. As I have argued, we are often in situations where our actions are interpreted like this. The robustly trustworthy person will respond to this.

What the idea that signalling in this way is part of the kind of trustworthiness that we are interested in gets wrong is that the robustly trustworthy person should try to *ensure* that others know she can be trustworthy. I think this helps us to better manage our reliance-responsive trustworthiness in a way which better guarantees the overall robustness of our trustworthiness.¹⁰ In practice, signalling will be difficult to calibrate and control. It may be difficult to constrain the demands of trustworthiness if we have to widely broadcast our trustworthiness. This will make it difficult to ensure that we are only taking on those demands that we can meet, and this will undermine our overall trustworthiness. There will be some situations – like those where we have strong positive moral duties – which look like signalling in this way. On my account, these come about because we can reasonably be taken as a default to have offered our reliability, and so difficult to signal that we will not.

4.4 Offering is what matters

Consider the following:

Ditzy Colleague: At a work Christmas party, you are talking to two other colleagues. You offer your reliability by saying that you will give them both a lift home. One colleague is paying you little attention. She misunderstands you; thinking that it is the other colleague who will be driving. You are unsure she has understood you. At the end of the party, she comes up to you both and asks if you are ready to leave.

¹⁰ I note that this is an objection to Jones on her own terms. She suggests signalling allows us to best meet the “conceptual need” we have for trustworthiness, which comes about because of our reliance. If I am right that this actually decreases how robustly this is done, then the reliance-responsive account would also better satisfy this conceptual need.

I think that the intuitively clear “trustworthy” thing to do here is to give your colleague a lift. I do not think that this is clear on the commitment account. Standardly, we think of commitments as requiring uptake to be binding and the kinds of commitments that Hawley (2019) is thinking about seem to be of this kind. In this example, there does not seem to have been any uptake of the commitment. Your ditzy colleague had no real idea of what you were offering. So, it is unclear that the commitment is binding on you and so that you ought to give her a lift.

Reliance-responsive trustworthiness can explain this, and in doing so it allows for a more robust form of trustworthiness than the commitment account; one which better mitigates our vulnerabilities through our reliance. It is more robust because it includes cases like this, which the commitment account does not clearly do.

What matters for whether we have a reliance-related obligation is whether we have offered another our reliability as a reason, and so whether we have endorsed it and recognise that we are accountable for that reason. You have done this, in our example, and so you have a reliance-related obligation.

The reliance in this case has come about in a specific way: because you have offered your reliability and so, regardless of whether she has done so in a clear-sighted way, your colleague has become reliant on you.¹¹ Your offer involved your endorsement of your reliability as a good reason for action. That remains a good reason for action for her because she is still reliant on you. It is irrelevant to the fact that you are obligated whether she understood the situation or not. That obligation comes from your endorsement of your own reliability, and the responsibility you therefore took on to realise that reliability. You should still be reliable in this situation, which means you should still give her a lift.

It is easy to offer reliability and can be difficult to signal that we are not doing so. The robustly trustworthy individual, who understands herself as socially situated,

¹¹ To be clear, the difference here is in the fact that a commitment requires uptake to be binding, while a reliance-related obligation only requires that you have offered a reason and that it remains a good reason for the person to whom you offered it. It may be possible to interpret the commitment here as open, but this seems to require quite a lot of bending of what a commitment can require. It would, for example, require that you can accept a commitment despite being unsure who it is from.

will be aware of this. It will likely mean that the best way to reason about our reliance-related obligations will be to pay attention to how others are reliant on us. This will be because we do quite frequently offer our reliability. The robustly trustworthy individual will be more likely to meet the demands created by reliance-related obligations if she assumes herself to have done so unless she has clearly and unequivocally signalled otherwise.

4.4 Deciding between reliance-related obligations

When we face different reliance-related obligations, we can prioritise them by how good a reason our reliability is for the individual to whom we offer the reason. As I discussed in the last chapter, a large part of this will be considering how vulnerable the individual is as a result of being let down. Part of being robustly trustworthy and fostering a socially situated perspective will then require gaining an understanding of how those we interact with could be harmed if we are not reliable.

I think this again carries some intuitive support. It is easier to be trustworthy for those we “get” and part of this is that we understand how our own actions and theirs interact, we know their values and how important their various plans are, and so we can appropriately offer our reliability in line with this close understanding that we have of them.

The robustly trustworthy individual will be aware that this is how she will have to reason when she gains reliance-related obligations. As she will be concerned with ensuring that she can fulfil all such obligations, she will take these issues of prioritisation into account when considering to whom she ought to offer her reliability.

5.0 Some further advantages of the account

There are three further advantages to adopting the picture of reliance-responsive trustworthiness which I have just laid out, beyond the greater guidance and clarity in reasoning and action which it provides. These are that it allows us to remain neutral in the discussion of promises and trustworthiness, clarifies how power increases the need to be trustworthy, and avoids over moralising trustworthiness and trust.

5.1 Promises, commitment and trustworthiness

Promises and trust may or may not be related. On the one hand it seems like we may need to trust if we are to accept a promise.¹² On the other hand, we often use promises where trust is absent.

It is common to elicit a promise where you find it difficult to trust. Your blabbermouth friend may be promise-bound to keep a secret. Your suspicions about your partner are soothed once you have extracted a promise from them. Promises can help us to secure some of the advantages of trusting, but when we are struggling to trust. This suggests that we are at least aware in some way of two distinctive features another agent can have: that she will keep a promise; and that she will be trustworthy. It also suggests that we naturally think of these as distinct to some degree.

It may be that this is simply an awareness that promises have further institutional support (like opening up the other individual to specific responses of blame from bystanders) and so are more likely to motivate the agent to act (see Scanlon, 1998). There are though also significant social costs associated with being considered untrustworthy, of which we are also deeply aware (Horsburgh, 1960; Thomas, 1989, 1990; Govier, 1992a, 1992b; D'Cruz, 2020).¹³

This account allows us to stay neutral on whether promises and trust have similar normative roots. It allows for the fact that promises and trust will often overlap; promising always counts as an offer of reliability. This may be because they are similarly grounded, or it may just be that they are usually coincident with each other. This account allows us to leave this open. That gives us a further reason to prefer it to the commitment account, which requires the same sort of responsiveness both to keep promises and to be trustworthy.¹⁴

¹² The claim is controversial because some, including Onora O'Neill (2002), take trust to be a necessary condition for accepting a promise. Some have even taken trust to be the basis of promising (Friedrich and Southwood, 2010; Bruno, 2020; Shaver, 2020). This account can remain neutral on this.

¹³ See also, M. Clark, 'How Trust and Distrust Reshape Our Social World' (MS)

¹⁴ Nickel (2007) makes a similar point. Nickel's account differs from that offered here in that the obligations he takes to underlie trust are not the reliance-induced obligations that I argue

5.2 Power and trustworthiness

Power shapes reliance. The more power an agent has, the less reliant they are on other agents to achieve their goals as far as she can bring them about herself.

Powerful agents may still be reliant on others, but they may also have means other than trust to ensure that this reliance works out. They may be able to compel or use authority (as discussed in Chapter 2). The less power an agent has, the more they will be reliant on others. This makes the less powerful remarkably vulnerable to the powerful. The powerful will also feature more often in crucial elements of plans and in important plans than the less powerful. This suggests that the demands of trustworthiness will be more stringent for the powerful, and that they therefore may be harder to trust. This helps to explain directly how trust and power interrelate and shape each other, a point which has been heavily developed – but without a mechanism underlying it – by Nancy Nyquist Potter (2002).

5.3 Demoralizing trustworthiness

We can understand reliance-related obligations as non-moral obligations. If we do, and we understand trustworthiness as a responsiveness to these obligations it allows us to demoralise trust (as I discussed in Chapter 2) and better understand the relationship between trustworthiness and moral demands.

Moralized notions of trust and trustworthiness fly in the face of experience (Bennett, 2021). Mobsters still trust each other. Murderers can be trustworthy towards one another. Members of a cartel really do trust each other to not slash prices and undercut each other. We should be able to capture this trust and trustworthiness, as this account does (see also Baier, 1986, pp.231–2).

We should also be able to understand what has gone wrong with such trustworthiness. This account allows this. It shows how these limited forms of trustworthiness come about because of errors in the agents reasoning about the relative importance of moral reasons and reliance-related obligations.

There is a genuine wrong that happens here. Mobsters can really let one another down. It is just that this is a less serious wrong, than the immoral actions which

for. His account, in my view, over moralises trust and trustworthiness. It removes the kind of trust that is possible in immoral situations, as I signalled towards in the last chapter and discuss further below.

would have been required to be trustworthy in these cases. The moral obligations of a mafiosa should stop her considering her reliance-related obligations, and so it is sometimes not bad to fail to be trustworthy.

This account of trustworthiness allows for trust and trustworthy behaviour in immoral situations and to do immoral things. Someone who has reasoned appropriately about their reliance-related obligations – as a robustly trustworthy individual must – would recognise that they ought not to take on such an obligation and that the moral considerations ought to cancel it out. Understanding trustworthiness in this way makes its value clearer.

6.0 Conclusion

I have presented a form of trustworthiness, which is not overly demanding, and which robustly reduces our vulnerability to our reliance on each other. I have argued that this account is better on these two counts than the existing “counting-on” and commitment accounts. This form of trustworthiness is:

Robust Reliance-Responsive Trustworthiness: *B* is robustly trustworthy to *x*, *y*, ..., *n* for *A*, *B*, ..., *N*, if and only if:

TWY1. *B* is competent to fulfil those reliance-related obligations she produces,

TWY2. *B* is disposed to fulfil her reliance-related obligations, where she lacks further reasons for acting in line with them,

TWY3. *B* ensures that she does not offer reliability where she would not be able to fulfil the resulting obligations (in other words, she adheres to due care).

The robustly trustworthy individual on this account views herself in a socially situated way. She will see herself as part of a web of reliance, understand how her actions expose others to vulnerabilities, and understand how other individuals will interpret what she does as reasons for their own actions. She takes care to only take on those reliance-related obligations which she can fulfil, and which it would be appropriate for her to fulfil. It gives us a more robust form of trustworthiness than the others in the literature which respond to our reliance-related vulnerabilities, but

without slipping into over demandingness. With this picture of trustworthiness in hand, we can now see how it helps to explain many confusing cases of trust.

6 | Trust and Relying

This chapter has two aims. The first is to demonstrate how understanding trustworthiness, as I have just suggested we should, can help us to solve some difficult problems in the philosophy of trust. §1 lays out what I take the relevant form of trust to be. §2 shows how taken together this reliance-responsive account can solve various problem cases. The second aim of the chapter is to demonstrate that this account of the attitude of trust can explain many of the features of trust which are taken to support its being a particular type of attitude. §§3-8 outline how this form of trust can explain the key features that trust is taken to have. This also aims to support the view that our trusting attitudes may take various forms, supporting the pluralistic approach I have developed in this Section.

*

We have so far focused on trustworthiness. We can now turn to trust.

1.0 A reliance-responsive form of trust

For your colleague to come to trust you to give you a lift, let us assume that he needs to rely on you to do so and to think that you have a reliance-induced obligation to do so which you will fulfil.

This would give us something like the following:

Reliance-Responsive Trust: *A trusts B to X, if:*

T1. *A relies on B to X, and*

T2. *A judges B to be “trustworthy” with regards to X-ing, which amounts to ascribing a reliance-related obligation to B and judging that B will fulfil it across a given set of circumstances.*

I will argue for both conditions in the second half of the chapter, by showing how they help to explain several features of trust. For now, we can understand “relying” as some sort of attitude which features in a plan and represents a bare reliance relation. A judgement of “trustworthiness” will amount to a judgement that the other agent is, in some sense, obligated to act in this way and that she will see this obligation through. This amounts to a judgement in line with the conditions just given. There is an obligation. There is an understanding that the trustworthy

individual will reliably do it. This amounts to an understanding that they are competent and sufficiently motivated to do it. Obviously, I do not mean to suggest that the reliance-induced obligation is necessarily conceived of by your colleague as one. He need not explicitly think “You ought to give me a lift, because you have invited me to rely on you for a lift and having offered that reliability, you now have a special obligation to realise such reliability.” We do not understand this in these theoretical terms, of course, but they paraphrase what we do standardly think.

Before I turn to defending this form of trust, let us consider how it can help us to understand some complicated cases which have challenged other accounts.

2.0 Applying the account

2.1 Key problems for trust

Several problems that are at the heart of the philosophical trust literature can be explained with my reliance-related account. The first of these explores the difference between reliance and trust and originates with Annette Baier. The later examples are Richard Holton’s confidence trickster and Katherine Hawley’s gift-givers and cowboys.

2.1.1 Kant

Kant: Kant took a walk at the same time every day. He was said to be so punctual that you could set your clocks to him. Why would it seem inappropriate to talk of the townspeople trusting Kant rather than relying on him (Baier, 1986, p.235)?

The reliance-induced account easily deals with why trust in Kant is inappropriate, as we have already seen. By walking punctually each day, Kant does not offer his reliability to others, in a way that could give rise to a reliance-related obligation. It is unreasonable to understand someone behaving reliably in this way as offering you their reliability as a reason to act. A citizen of Königsberg, who relied on Kant even if they faced significant loss from doing so (for example, by using him as a timepiece for taking a life-saving medicine) would not, therefore, be justified in thinking that Kant had offered his reliability. So, the citizen would be mistaken in thinking a reliance-induced obligation existed, and so wrong to judge Kant trustworthy in this regard. This would mean they were wrong to have judged Kant as trustworthy and

so had failed to form an appropriate trusting attitude. They ought only to have relied on Kant.

2.1.2 *Con-artist*

Con-artist: A confidence trickster relies on a particular response from the person he is trying to trick, normally called “the mark”.¹ The mark clearly can trust the confidence trickster, which is what the con-artist relies on. The con-artist also relies on the mark, but can the con-artist trust the mark?

We intuitively want to say no. There seems something inappropriate about talking about the con-artist trusting the mark. It would be inappropriate were the con-artist to feel let down or betrayed – the standard marker of trust – if the mark wises up and backs off from dealing with him. Yet, say the mark transfers a small amount of money into the con-man’s bank account. Perhaps this is part of the trick, the small amount of money resulting in a payment which will be “completed” when a second larger amount of money is paid. That small amount of money seems to offer the mark’s reliability and invite the con-artist’s reliance. This might suggest that the mark behaving differently now would be for her to fail to fulfil that reliance-related obligation. This all seems to suggest that my account means that the con-artist does trust the mark, and even worse, that to be trustworthy the mark should act as the con-artist requires. My account would then be less intuitively plausible than many others.

Fortunately, that is not what my account suggests. The con-artist deliberately entices the mark to offer a form of reliability which the con-artist will twist. The mark, for example, understands herself as offering to reliably put in her share of an investment. That is what she offers, and if she were obligated to do anything it would be that. If she then finds out that this is not what she is doing, then she has not offered her reliability to enrich the con-artist at her expense and she cannot be bound to seeing it through. The con-artist should not have ascribed a reliance-

¹ (Holton, 1994, p.65). The case was developed to object to the goodwill account.

related obligation to her, and so it would have been inappropriate for the con-artist to have trusted her.

2.1.3 *Champagne Fridays*

*Champagne Fridays:*² Imagine that your colleagues begin to buy you champagne every Friday. You come to rely on that bottle of champagne to start your weekend. One day, the champagne buying stops.

You should not have trusted your colleagues to buy you champagne. Again, those characteristic responses of being let down or betrayed seem massively inappropriate. This inappropriateness comes about, on this account, because we ought not to ascribe a reliance-induced obligation in situations of gift giving. This is because gift giving, and other examples of generosity, are not ways in which we offer our reliability. To view gift giving like this seems to have been to misunderstand it. If you have a friend who always gets good gifts for you, it seems right that you can think she is reliable at getting you such a gift, but not that she has obligated herself to always be so. Gifting is incompatible with such an obligation.

This can, of course, change. Your friend might make a big deal about how good she is at buying gifts; how her present will always make up for any of the misses your other friends buy you. That is part of how we standardly can alter the range of ways we are obligated to be reliable. I think it can seem appropriate in cases like this to come to trust you friend, and this is because she has accepted a reliance-related obligation.

2.1.4 *Cowboys*

Katherine Hawley raises an objection which she takes to stand to all obligation accounts of trust and trustworthiness: her cowboy objection (Hawley, 2014, pp.18-19).

Cowboys: We are in the Wild West, during the Frontier period. There is some semblance of law and order in the towns, but out in the desert it is well known that anything goes. Two cowboys come across each other in the desert and each wonders if the other will draw his gun and shoot.

² Hawley (2014, p.3) uses the example to distinguish non-trust and distrust.

Hawley thinks that this is a situation of non-trust. That is the agents have no reason to trust or to distrust each other. ‘Before we exchange words, you may try to predict whether or not I will let you live. But no matter how confident you are in that prediction, it’s not appropriate for you either to trust or distrust me in this respect.’ In that desert, there is no commitment not to kill each other and so there is no trust (Hawley, 2014, p.19.).

Hawley suggests that an obligation-based account would struggle with this example. For that to be the case, we must adopt a certain view of obligations. She seems to be thinking that an obligation-account of trust would involve the ascription of some general duty to a specific agent. She points, for example, to the fact that an obligation not to kill another would still apply in the deserts of the Wild West. That suggests that she thinks *Cowboys* rule out an obligation account, because any obligation account would have to include the obligation not to kill.³ If an obligation account of trust had to appeal to general duties, then I agree with Hawley that it would face problems explaining the situation between the two cowboys.

I do not see why an obligation account of trust would have to take this form. Reliance-related obligations are not of this form, and I do not think that *Cowboys* offers a counterexample to my account. I can still explain why this is a situation where trust and distrust are inappropriate.

In that desert, where it is known that anyone can shoot anyone, neither cowboy can clearly offer his reliability to the other and so neither cowboy can come to have a reliance-related obligation towards the other. This means that no one should ever ascribe a reliance-related obligation to another, and so there can be no trust. Distrust is also inappropriate, because this also requires such an ascription.⁴ So, we are in a situation of non-trust. Obligation accounts can then deal, as well as commitment accounts, with such situations.

³ As Nickel (2007) would seem to. I do not think that Nickel’s successfully argues for a moralised account of trust, as his account fails to explain situations like this.

⁴ See my discussion of distrust in the main text.

2.1.5 *Trust with explicit denials*

Andrew Kirton has provided a powerful counterexample to the commitment account. Kirton provides the following example (Kirton, 2020, pp.587-88):

Sailor: “‘in Abram Kardiner’s psychotherapy of the navy veteran who had been rescued at sea after his ship was sunk, the veteran became most upset when revealing how he had felt let down by his own side: ‘The patient became rather excited and began to swear profusely; his anger aroused clearly by incidents connected with his rescue. They had been in the water for a period of about twelve hours when a torpedo-boat destroyer picked them up. Of course the officers in the lifeboats were taken off first. The eight or nine men clinging to the raft the patient was on had to wait in the water for six or seven hours longer until help came.’

The officers had been rescued first, even though they were already relatively safe in the lifeboats, while the enlisted men hanging onto the raft were passed over, and some of them drowned as they awaited rescue. Though Kardiner accepted this procedure as part of the normal military order, the patient was horrified at the realization that he was expendable to his own people. The rescuer’s disregard for the man’s life was more traumatic to him than were the enemy attack, the physical pain of submersion in the cold water, the terror of death, and the loss of the other men who shared his ordeal.” (Herman 2001, 55)

In this case, the procedure to rescue officers first was an express commitment (or a consequence of one) from the veteran’s side, that he understood. Yet regardless of what had been explicitly communicated, the veteran trusted his side to rescue him. The feeling of betrayal he experienced when his side in fact followed through on their commitment, was vivid and traumatising.’

I agree with Kirton that the example poses a serious challenge to the commitment account. It may seem to do the same to my account. The navy, in being explicit about the order of rescue, may seem to have taken the appropriate steps to avoid a

reliance-related obligation to have saved the sailors on the raft first. They have explicitly refused to offer their reliability in doing so.

I have two potential lines of defence against Kirton's example. The first attempts to break my account out of trouble by going around the example. In doing so, it may also offer a route out for the commitment account. The second faces Kirton's example head-on and in a way not open to the commitment account. The first relies on the Social Vulnerability Account that I outlined in Chapter 2; it simply says that Kirton's problem is not a problem for the kind of trustworthiness in which we are interested. The second suggests a way that the reliance-responsive trustworthiness account I have outlined in Chapter 5 could accommodate the problem. I prefer the second defence, but I can fall back on the first.

Let us try to avoid the counterexample first. I have claimed only that I am providing an account of one form of trustworthiness and trust. I accept that there are others. As Kirton goes on to persuasively argue the sailor's betrayal may well stem from a broken sense of attachment to the navy. He feels betrayed because he has been abandoned by those to whom he felt he belonged; no matter how clear they were about how they would treat him.

So, my first potential defence is to step around the problem. Kirton is right, that is the way in which this sailor was betrayed. It is not, however, the form of trust and trustworthiness in which we are interested. As my view is a pluralist one, I can accept that there are other forms of trust that may arise in other contexts and that they can be broken without necessarily affecting the forms of trust in which we are interested. There was no reliance-responsive trust in this situation or at least ought not to have been. The navy did everything correctly to signal that they were not offering their reliability (and for that matter, that they would not be committed). The sailor's betrayal has a different source: the sailor's attachment to the navy.

This is because we ought not to have such an attachment to the businesses with which we interact. The satisfaction of such a social need is not appropriately met in such interactions. They are therefore not our concern. So, we can move on.

This response is satisfactory. It is open to the defender of the commitment account as well. We can both escape. We should then favour my obligation over the commitment account for the reasons given in the last chapter.

I think that I can also meet this potential objection head on. I argued in Chapter 4 that where we have a positive moral duty, we cannot reasonably avoid offering our reliability for it. Here, there is a duty to save the sailors and the officers lives. This means that the navy cannot reasonably signal through due care that it would not be reliable in these situations. They cannot avoid gaining a reliance-related obligation.

I also argued in Chapter 4 that to reason about which reliance-related obligations we ought to prioritise we ought to consider how the other's vulnerability to us varies. The sailors were highly vulnerable on all three axes that I have considered. Their only hope of rescue is the navy. Their rescue is central to their survival, and their survival is (presumably) their most important plan. The officers are in the same situation along two of these axes, but their rescue plays less of a pivotal role in their survival. The sailors have been clinging to a raft. The officers are sitting in a lifeboat. The sailors are more vulnerable through their reliance than the officers.

The navy has then failed to be trustworthy on my account as well. They have inappropriately reasoned about what reliance-related obligations they have and then failed to recognise how they ought to prioritise them. The sailor did trust them in this, and that trust has been broken.

The reliance-responsive obligation account can then respond to this objection in a way that the commitment account cannot. Perhaps, both responses are correct. As I said, in real life the various forms of trustworthiness and trust will often be mixed together. The navy may have failed to be trustworthy on my account, but what really matters – and what really mattered to the sailor who describes his experience – was that attachment he had to them, which they broke through their actions.

The reliance-responsive account of trustworthiness and trust can handle the core cases which have puzzled and blocked other accounts. We have, therefore, good reason to consider it a good account of what trustworthiness and trust are, independently of its ability to offer a robust and appropriately demanding way of reducing our vulnerability to others because of our reliance on them. If you

disagreed with the arguments of Chapter 2, this may give you enough reason to accept the account offered in these final two chapters as *the* account of trust. In the rest of this chapter, I want to say more about the nature of trust to demonstrate why you should accept the overall approach argued for in this thesis.

3.0 What we need from an account of trust

An account of trust should illuminate trust in certain ways. An account of trust should be able to explain some core features of trust. Trust seems to respond unusually to evidence (Lahno, 2001, 2004; Jones, 2019). For example, we may continue to trust a friend's word that they have not done something, despite mounting evidence to the contrary. Trust also seems to be contagious: when we see others trusting an individual, we are more willing to trust them (Pettit, 2004).

In the social sciences and among some philosophers, trust is taken to be a necessary condition for cooperation.⁵ Social scientists infer trust from cooperative activity in “trust games” (games where two players must each make a dyssynchronous “move”, but where a higher pay-off is available when the players cooperate). Most philosophers of trust focus on its normative aspect: the presence of trust seems to mean that both the trusting and trusted party ought to behave in certain ways. This kind of trust is also inferred, normally by the presence of being let down or betrayed when those in trusting relationships behave in ways they ought not to.⁶ We need to understand how trust makes such responses appropriate, in ways that mere reliance (for example) does not (Baier, 1986, p.234, 1991; Holton, 1994, pp.2-3; Pettit, 1995, p.205; Hawley, 2014, p.2).

An account of trust should also be able to explain distrust.⁷ Distrust is different from the absence of trust or non-trusting. There are many things that we do not trust in, but which we do not therefore distrust. I do not trust my neighbour to paint the

⁵ For an overview of the social sciences view see Schilke, Reimann and Cook (2021). Philosophers who focus chiefly on the cooperative effects of trust include Simpson (2012); Nguyen (2022b); Lahno (2017) and the discussion in Hardin (2006, Ch.3).

⁶ This began with Baier (1986), and is present throughout most of the philosophical discussion on trust.

⁷ The point is made by Hawley (2014), and the observations about distrust are taken from that paper.

outside of my house. This does not mean that I distrust her. Distrust involves something more active than simply not-trusting. We should be able to explain this.

In the rest of this chapter, I will show that understanding trust as a judgement of reliance-responsive trustworthiness and a relying attitude can explain these features.

4.0 Trust and its interactions with evidence

It has been claimed that trust responds strangely to evidence.⁸ We can differentiate two separate ways in which trust interacts strangely with evidence. It can go beyond the evidence that we have, as often happens when we first trust someone. Once we trust, it can also seem to interact strangely with new evidence we acquire: not going beyond the evidence but rather lagging behind.

4.1 Judgements and going beyond the evidence

We can rationally make judgements which are not strictly responsive to evidence.⁹

The way that we come to trust in many cases seems to suggest that it includes a judgement of this sort.

The level of evidence required for trust seems to vary. Consider the following example from Thomas Simpson (2017, pp.177-8):

Antarctic Resupply. Lief is preparing to trek to the South Pole alone, pulling his food and equipment with him on a sledge. He is aiming to break the record for the fastest unsupported journey. The weight of his sledge would jeopardize the attempt if it had provisions for the return journey as well. So, Katherine—an old Antarctic hand, who runs an adventure support company—promises that she will be contactable via satellite phone. When called, she will drop by parachute a package of supplies at the Pole, and they arrange a contingency plan if communications should fail. If Katherine does not follow the plan, it is all but certain that Lief will die. Lief is very keen to survive the expedition. He sets out south.

⁸ The odd interaction of trust to evidence is highlighted by Hieronymi (2008, Ft.24).

⁹ See, for example, the arguments in McGeer and Pettit (2002) with which I largely agree.

Simpson argues that Lief only ought to trust Katherine if his total evidence supports the fact that Katherine will drop the package.¹⁰ I agree with Simpson that such a case will have a high threshold for trust. I think that the best explanation for this is that the constraints on how far a judgement can diverge from rationality vary by the overall risk that making such a judgement would produce. Lief's reliance here risks his death; he ought to be highly certain that Katherine will not let him down. This does not, however, require that Lief believe that Katherine is trustworthy, and it allows for the fact that, standardly, the threshold for rationally making a judgement of trustworthiness may be quite low; and so many of our judgements of trustworthiness go beyond the evidence available to us.

4.2 Relying and lagging behind the evidence

Trust seems to lag behind the evidence as well. For example, you may continue to trust the word of a friend, despite mounting evidence that they have done what they claim not to have done. (We must put aside, for now, the confusion that norms of friendship and loyalty may introduce to such a case.) If we accept that trust is an attitude formed of a judgement and relying, then we can explain these by noting features of such attitudes.

A common explanation for this evidence-resistance is that trust involves an affective element. This is often put down to and explained by trust being an affective attitude (Lahno, 2001, 2004; Jones, 2019). The effects of emotion, it has been pointed out, tend to focus rather than block attention on an object (cf. LeDoux, 2012; Brady, 2013). For example, it is not that fear stops me seeing how small the spider is, it is that fear causes me to concentrate on those aspects of the spider that seem dangerous. This can lead to further effects on the beliefs and other mental states that we form about the spider. A general form of the common argument would run like this:

¹⁰ I agree with Simpson that we can have second-order reasons which affect the amount of evidence required for trust to be appropriate, which vary according to the situation in which we are placing this form of trust; but I do not agree with his claim (made more strongly in Simpson, 2023) that this means that trust standardly follows the evidence. The latter does not follow from the former. It means that the threshold for well-placed trust will standardly follow the evidence; but we may be poor judges of the appropriate threshold. Equally, other forms of trust will not be evidence responsive like this.

1. Trust has limiting effects on the attention of the trusting party, leading her to ignore counterevidence.
2. Emotions have the effect of focusing attention on certain aspects of an object (e.g., fear may lead one to focus on the legs and fangs of a spider).
3. The attention directing effects of emotion have indirect cognitive effects, by leading us to focus on certain pieces of information and leaning us towards certain interpretations.
4. These indirect cognitive effects change our evaluations of objects, situations and agents.
5. 2-4 best explain how a trustor responds strangely to evidence.
6. So, trusting must have an emotional component.

The idea here is that we have a limited amount of attention. Affective trust directs our focus towards the trustworthy elements of the trusted party, using up attention that could be considering evidence that would lead us to reconsider our trust.

I think that we have good reason to question (5). There is an alternative explanation which can account for non-affective forms of trust just as well. This is that relying attitudes are parts of plans and plans interact strangely with evidence, and that this means that some forms of trust may also be evidence-resistant but that this does not show that they must contain an emotional component.

4.2.1 Relying and planning

The idea here is that relying is a planning attitude.¹¹ Plans interact strangely with evidence and as relying attitudes are parts of plans so do they. Plans interact strangely with evidence because they provide a 'default' for action. When we plan, we settle a course of action and in doing so we commit ourselves to see it through. This commitment is defeasible, but we face practical rational pressures to stick to our plans.

¹¹ The concept of planning here draws on Michael Bratman's view of planning agency. See particularly his Bratman (1999) and from that especially Chapter 3.

This is because of the benefits that planning in this way can bring to agents like us. That we have limited cognitive resources is one of the features of our limited agency. Plans providing a 'default' help us to save significant cognitive resources compared to if we had to constantly deliberate about what to do. By reducing our 'cognitive load', plans allow us to make better use of our time and to pursue a range of projects.¹² We only gain this benefit if we do not have to re-plan.

As they form this default, plans resist reconsideration. We tend to stick to plans, particularly once we have set the wheels in motion, unless things start to fall apart. We face practical pressures to continue to do so. These practical pressures will not necessarily align with the rational pressures we face to conform our plans to our available evidence. Where they do not align, practical pressures can provide counterweights to those rational pressures, which mean that we ought to stick to our plans. This does not mean that incoherence between beliefs and relying attitudes would not be criticisable, but it does mean that they are likely to lag and interact in odd ways. Contradictory evidence must push against the dead weight of our plan. It must be sufficient to make us willing to re-plan. I want to show that this strange interaction between planning and evidence can explain the lagging behind evidence which trust exhibits, once we understand that relying attitudes are planning attitudes.

Arguing that they are requires a bit of a detour. It is one that is worth it, I think, because it contributes to the plausibility of the pluralist approach for which I have argued. The multiple realisability of many features of trust has not been considered enough, and this has led people to commit to unitary accounts although this is not justifiable.

4.2.2 Relying attitudes as beliefs

Attitudinal relying seems to mediate between our plans, which allow us to achieve our ends, the limits of our agency, and our beliefs about how the world actually is.¹³ The ancient farmer's overall plan is to grow wheat, she cannot cause the ground to

¹² This point about planning is also made by Frost-Arnold (2014, pp.1960-1).

¹³ Smith (2010, p.136) has made a similar point on attitudinal reliance mediating the agent's theory of the world, her 'conception of [her] agency' and her 'ends as she understands them'.

be nutritious enough to grow her crop herself, and she believes that the Nile's flooding will. Is the relying attitude then best understood as simply a belief incorporated into a plan?

We can consider some uncontroversial features of beliefs, as laid out by Michael Bratman.¹⁴ Beliefs are commonly thought to be context independent. That is, an agent will hold a belief, regardless of the context they find themselves in. Your belief that a goldfish cannot cook dinner does not cease because you now want your dinner to be prepared. Beliefs also aim at truth, and they ought to change according to evidence. A belief is also an attitude that one cannot voluntarily adopt. One can form a belief only if one has the evidence to form such a belief: one cannot simply will oneself to believe it. Finally, beliefs are subject to an 'ideal of integration'. One of an agent's beliefs should form a coherent and consistent whole with the agent's other beliefs.

The evidence responsiveness of beliefs is taken to be a particularly distinguishing feature of them as an attitude.¹⁵ Beliefs aim at truth and as such are independent from 'practical considerations'.¹⁶ Reliance does not seem to respond to evidence in this way. You may want to follow up on a citation in this chapter but be unsure if the library is open this evening.¹⁷ As you approach the library, you may see that its lights are off. This ought to affect your beliefs about the library. The question is then whether we can "insulate" beliefs by incorporating them into our plans. This is a difficult question to fully resolve, but it would seem highly desirable that we cannot. A set of beliefs ought to be subject to coherence and consistency. It seems like a bad route to take, if our theory says that you rationally believe the library to be shut (because you see its lights are off), but simultaneously that it is open (because this is the belief you incorporated into your plan).

¹⁴ Bratman (1992, pp.3-4). The features are meant to be common across views of beliefs as credence-based or all or nothing.

¹⁵ See also Smith (2010, p.139). The discussion in this section owes much to Smith's rejection, (pp.138-140) of the Simple Belief View, with which I agree.

¹⁶ Smith (2010, p.139) and Bratman (1992, p.9) also point to the independence of belief from practical pressures to defend an account of distinct attitudes related to planning.

¹⁷ The example is from Smith (2010, p.139).

Beliefs are also not attractive because they are involuntary. Relying does not seem to be involuntary. It often seems very much up to us whether we rely on something or not. So, let us leave this option.

4.2.3 Relying as expectations

An alternative is that to rely is to form an expectation or 'supposition'. This appears to be a belief with a certain credence or a belief about the probability of some event. This has certainly at points been the dominant view – implicitly – in the trust literature – it appears to be the view of Holton, (1994, pp.63-76) and Hieronymi (2008).¹⁸ An expectation can be extremely low, but you cannot rely rationally if you consider the chances of the event to be zero (Holton, 1994, p.72). That is because one can criticise as irrational someone relying on something that they think will definitely not happen. If you have a pet goldfish and your sub-plan to get fed while reading this chapter includes relying on the fish to cook your dinner, then, given that the fish cannot cook dinners and that you also believe this, it seems that you can be criticised for adopting such a plan. You are being incoherent.

Expectation is, however, still an unlikely candidate for the relying attitude. It is still a belief, and as such it should still be responsive to evidence. Again, your expectation that the library is open will decrease as you see its lights are off. It is not clear that you should rely any less on its being open as a result. We seem again to be able to put some distance between expectations and relying. The only route out again seems to be to insulate some expectations, but this still seems like a poor choice.

4.2.3 Relying as practical acceptance

A further alternative candidate for the type of attitude involved in reliance is Michael Bratman's concept of 'practical acceptance'.¹⁹ On this picture, our beliefs

¹⁸ Katherine Hawley has also accepted Holton's view, though she notes that reliance has 'justification conditions that are not purely epistemic, and, unlike believing, relying is sometimes a matter of direct choice' (Hawley, 2014, p.4). I have been unable to fully work out Hawley's view.

¹⁹ It is not clear that Bratman (1992) intends for practical acceptance to be considered as a form of reliance. He views attitudes of practical acceptance as forming the background framing for practical deliberation. It is proposed as a candidate by Smith (2010, pp.140-1) who ultimately rejects it, and by Frost-Arnold (2014), who bases her account on it. This section's account of reliance as practical acceptance draws heavily on Frost-Arnold (2014), account of trust. Frost-Arnold notes that reliance and trust are taken to be distinct by most of the philosophical literature, but her account of trust ignores this distinction (pp.1958-9). This

(satisfying the conditions given above) form a '*default cognitive background*' for our theoretical and practical reasoning (Bratman, 1992, pp.9-10). Specific practical pressures can cause us to alter this default background when deliberating through two mechanisms. The first of these is that we can '*posit*' a proposition for our practical reasoning, when we do not believe that proposition in our default cognitive background. The second mechanism is that we can '*bracket*' or exclude a proposition from our practical deliberations, when we do believe that proposition in our default background. Both mechanisms lead us to accept a proposition that differs from our actual beliefs (e.g., excluding *p*, which is in the default background, means accepting not-*p*). These practically accepted propositions form a '*context-relative adjusted cognitive background*' against which we then deliberate and form intentions.

Bratman's argues for practical acceptance through examples like the following:

'I have a chair and a two-storey ladder. In each case I think it equally and highly likely that it is in good condition. Indeed, if you offered me a monetary bet about whether the chair/ladder was in good condition I would accept exactly the same odds for each object. But when I think about using the chair/ladder things change. When I consider using the chair I simply take it for granted that it is in working order; but when I am about to use the ladder I do not take this for granted.' (Bratman, 1992, p.7)

seems to be a flaw in her overall approach as there are significant differences between trust and reliance, which are better captured in a "reliance plus" account like those discussed in Chapter 5. For example, her account suggests that one may choose to trust to reduce the cognitive load of the world and that this is acceptable if the potential cost of betrayal is low enough. It is unclear, however, whether such betrayal would be appropriate in such a situation. If an agent has trusted another just because doing so was simpler for them, it is unclear that the trusted party would have to live up to such trust. If my partner is worried about how messy the house has become and so chooses to trust that I will sort it, it seems strange that I can now betray my partner by failing to tidy it. (The problem is sharpened if we accept that broken trust also makes resentment appropriate and introduces a need for an apology - as I think we should. Here, even if betrayal would be odd, it seems inappropriate that I should have to apologise to someone who has chosen to trust me purely for their ease.) The same issue does not arise for reliance, however. I can choose to rely, simply to make the world less anxiety inducing. Keeping the distinction is therefore important, as my approach does.

The example demonstrates one of the context-relative practical pressures that can push us to accept a proposition into our practical deliberations: the asymmetries of cost. In taking a bet on whether the chair and the ladder are in working order, my potential loss is the same for either and so I face no practical pressure to modify my default beliefs about the objects. In using a two-storey ladder, I face significantly higher costs should it fail than if the chair fails. These higher costs put greater practical pressure on me to exclude or bracket my beliefs about the ladder before planning to use it. Pressure that is absent when I consider whether or not to use the chair. My beliefs about the chair therefore make it into my adjusted cognitive background, while those about my ladder do not.

We face other context-relative practical pressures to accept a proposition in our practical deliberations from a need to simplify our reasoning, a need to socially cooperate (for example, in accepting some basic points of agreement in discussion), our special standing to and relationships with each other (for example, friendship may apply practical pressure to accept what our friends tell us), and in some cases as a requirement to practically deliberate at all (for example, for a soldier to continue with his day despite believing he is likely to be killed that day). (The examples are all Bratman, 1992; the examples of the pressures are given on p.5, p.7, pp.7-8 and p.8 respectively.)

Acceptances do not share the features of our beliefs in the default cognitive background. They do vary from context to context, as we can see with acceptances about the ladder when we place a bet on it or must use it. Given this, they are not subject to the ideal of integration. They can vary across our different plans. They are mainly influenced by practical, rather than truth-related, considerations. They are also subject to our voluntary control: I can decide to accept some proposition for the purposes of my practical deliberation.

Practical acceptance has been considered as a candidate for our relying attitudes. On this account A relies (on B to) X if and only if the proposition that (B will) X is part of A's adjusted cognitive background – this is adapted from (Frost-Arnold, 2014, pp.1963-4). If (rather than your goldfish) you have relied on your partner to cook

your dinner this evening, so that you can finish reading this chapter, what would this account require you to have done?

The idea is that you have posited the proposition that your partner will cook dinner, and you have accepted that into your adjusted cognitive background. You can then deliberate based on that belief, which allows you to plan to continue reading, rather than to go downstairs and start chopping an onion. This seems like relying, but I do not think that practical acceptance is the right kind of attitude to explain all of the features of relying attitudes.

One difficulty for the practical acceptance account is that it does not require one to want the acceptances to be the case. We can practically accept something, and plan based on this practical acceptance, without thereby wanting the practically accepted belief to come about.²⁰ We can practically accept a high budget for a project and plan accordingly. We might for example delay part of the project, until we have seen if other parts have been delivered on time and to their estimates. We can be actively working for such a high budget to not be hit. We cannot rely on something that we simultaneously try to prevent.

To try to stop something that we are relying on is to self-sabotage our own plans. We can, of course, be unhappy or conflicted that we have to rely on something, but we must nonetheless have some sort of pro-attitude to its coming about. This inherited pro-attitude is a distinctive feature of reliance that it is not present in practical acceptance. It suggests that the relying attitude is more tightly bound up in our planning attitudes than practical acceptance.

We can also accept a proposition that we do not believe.²¹ This is what Bratman takes to be happening in the budgeting example. I do not believe the budgeted cost,

²⁰ Smith, (2010, pp.143-4) makes this criticism. He (pp.142-3) also views attitudinal reliance as involving a pro-attitude inherited from the plan. I want to emphasise how this pro-attitude can be accompanied by large amounts of anxiety or discomfort about the relied upon thing. I may rely on my friend failing to get a promotion in order for me to get a promotion. I can wish that I did not have to rely on her failure for my success, but so long as I want to get promoted and so long as I plan to do so, there is a derivative desire directed at my friend's failure.

²¹ Smith (2010, p.140, suggests that this is not the case in acceptance. He takes Bratman's (1992) view as being that one ought not to accept p when one believes that not- p , as this

though I accept it for the purposes of my deliberations for other practical reasons (here because of the potential prohibitive costs of going over budget) (Bratman 1992, p.6).

Relying does not seem to have this total disregard for our beliefs. There seems to be something unjustifiable in relying on something that one does not believe at all. If I think that there is zero chance of something happening, I am criticisable if I rely on it. We may pretend that we rely on something or run a hypothetical construction of a plan, but to act on a relying attitude in a justifiable way one needs to not think that the relied-upon event or action will not happen.

This means that reliance has some belief-like qualities, but also it seems to respond strangely to evidence. A relying attitude can be criticised as irrational if it fails to pick out something that an agent is actually reliant on to successfully complete her plan.²² If I continue to rely on my alarm clock, even though I have discovered that it does not wake me up, then something seems to have gone wrong with my planning, which I ought to correct. If relying was practical acceptance, it would not explain these features.

Practical acceptance views of reliance also fail to distinguish the place of acceptance and reliance in our planning. Relying is primarily a product of our decisions in a plan, it is not an input to our deliberations through which we reach decisions (as acceptance is, see (Bratman, 1992, §5). Practical acceptance creates the background against which we plan, not the plans themselves. In deliberation we reach decisions on what to do, in part by accepting different propositions based on different

would violate the overall rational consistency of an agent's plans and beliefs. Bratman's (p.6) view of practical acceptance, however, allows us to accept something that we do not believe. Frost-Arnold (2014, pp.1967-8) argues that where there is overwhelming evidence that not- p , one would be able to accept that p in any context in an attempt to save this feature for practical acceptance. It does not seem necessary however for evidence to be overwhelming to make relying on p unjustifiable. I do not need clear evidence that my plans are not reliant on something to make it unreasonable for me to rely on it. I could simply lack evidence that my plans are reliant on it. This would seem to not hit the bar for overwhelming evidence that I am not reliant, yet it ought to stop me from relying.

²² Smith (2010, pp.145-6) also makes this point about relying attitudes as does Alonso (2009, Ft.40).

pressures we find ourselves facing. It is only after we have decided that we then form an intention about what we have to do.

It only seems right to talk about relying on something when I have formed a plan that includes that reliance. For example, you may accept the proposition that the library will be open to go and get the book you need. This will then form the background against which you decide what you are going to do. Only then can you form an intention about what you will do: an intention that you go to the library to get the book. It is only in the context of you having formed this intention that you become reliant on the library being open. It is only then that you seem to be able to appropriately talk of yourself as relying on it being open. You can accept the proposition that the library is open and form different intentions, which do not lead you to rely on it. Practical acceptance is part of planning, but before the plan is enacted. Relying attitudes are part of an enacted plan.

A further distinctive feature of reliance is that it necessarily involves an agent reflectively recognising that the relied upon thing is outside of her control (Smith, 2010), in a way that practical acceptance does not. This also seems correct. We cannot appropriately speak of relying on things that fall directly within our direct control.²³ I cannot rely on me going shopping to go shopping. Practical acceptance does not require this (though this leaves open whether reliance is a sub-species of practical acceptance).

Relying attitudes are not simply the incorporation of a belief or expectation, nor are they practical acceptance. I think that we are better to think of relying attitudes as a distinct intention-like planning attitude which form parts of our plans.

4.2.4 Reliance as a planning attitude

We have seen that relying attitudes seem to interact oddly with evidence yet seem to have some minimal belief-like requirements, to include a contingent pro-attitude tied to our plan and to involve an agent reflexively recognising that their contents are beyond that agent's control.

²³ This would not rule out self-reliance but would suggest that we are self-reliant only on aspects of ourselves over which we do not have direct control. This seems right. I can talk of relying on my memory, but it seems odd to talk of relying on myself to raise my hand.

I think that the simplest explanation – and the one we should prefer – for this is that relying attitudes are distinctive planning attitudes, much like intentions. Unlike intentions, they allow us to plan about that which is outside of our control.

We can see a similarity between intentions and relying attitudes in that they require a minimal coherence with our evidence about the world. An intention is unjustified if whatever it would take for the intention to actually be realised is beyond the control of the agent who has the intention. If I intend to wake up naturally at six a.m. every morning, without my alarm clock, this intention can be criticised if I lack the capacity to wake up naturally at that time. An intention's rationality is - in part - determined by whether the agent holding the intention can actually realise the intention's content.²⁴ An intention's justification is in part conditional on its success. Relying attitudes have a similar success-conditional relation to the world. Having a relying attitude is justified so long as it picks out something that can (potentially) happen and will be useful to our plans.²⁵ This helps to explain why relying attitudes are unjustified when we think that there is no chance of something happening. We contradict these potential success conditions.

Intentions can also have contingent and derivative pro-attitudes, gained from the ends that they pursue. When the plans in which we form our intentions collapse, so can the intentions that go with them. If you find that your kitchen is full of snacks, and so you no longer need to rely on your partner to cook you dinner, your previous intentions to offer to cook your partner dinner tomorrow disappears, along with any desire you may have had to carry out that intention. If relying is a planning attitude, then this explains the desire that accompanies relying on something. When you rely, you gain a contingent pro-attitude for (B to) X derived from your overall pro-attitude towards your plan.²⁶ If you were to decide that you have had enough of

²⁴ This is to put aside many issues that would take us far beyond the scope of this chapter, about the exact nature of intentions or intending to try.

²⁵ On Bratman (1992, §5) view intentions are also more constrained by belief than decisions. If relying attitudes are similar psychological states then they may be similarly more constrained by beliefs, allowing for the fact that one cannot rely on *p* when one believes that not-*p*.

²⁶ This is distinct from hoping that it may be the case, which need not be contingent on a plan. I may hope that something happens (say, that both my friend and I could be promoted, rather than just one of us) although I rely on something else to happen. Hope has also been tied to

reading this chapter, your reliance and your associated pro attitude towards your partner's cooking would disappear. In fact, you may actively not want your partner to cook.

This view also helps to position reliance more accurately as a product of decisions, rather than as part of our deliberations. They come about as part of our planning. This in turn helps to explain how reliance is voluntary: it is a product of a decision, much as an intention is.

This also helps us to understand why we must recognise that on which we rely as outside of our control. It is part of the role they play in our planning. They represent precisely those things we need to come about, but which we will not bring about. We are therefore irrational if we simultaneously think that we should bring them about or if we act to directly bring them about.

4.2.5 Relying Attitudes

To summarise, a relying attitude has a number of distinctive features. It is belief-like, in that it requires some tie to the actual world to be justified. It also has some relation to our beliefs: it is inappropriate to rely on something, when one does not believe that it is the case. Yet it is voluntary, in a way that belief is not. It also has some desire-like qualities: it inherits a contingent pro-attitude towards its content from the plan in which it features. The best candidate for such an attitude is that relying attitudes are a distinct planning attitude. I have attempted to point towards its being a distinct, intention-like attitude, which is part of our plans and a product of our practical deliberations.

4.2.6 Evidence and Relying

Understanding reliance as a distinct intention-like psychological state which is part of our plans allows us to explain its evidence-resistance.

When I mistakenly rely on my alarm clock to wake me up, despite actually being reliant on my neighbour's door slamming, I am unlikely to give up on my alarm

attitudes of trust, for example in Victoria McGeer's (2008) account where a trusting party's hope acts as a motivational impetus to the trusted party to show herself as a trustworthy agent. This provides a further potential source for such a pro-attitude in trust, with hope providing a background for reliance. This further demonstrates the potential multiple realisability in which we are interested.

clock straight away when it fails to wake me up. I will continue to rely on it, despite failing to wake up, for at least a day or so or I will test it by re-setting it and seeing if it goes off. It seems intuitive that such relying would continue, even if one has the belief immediately that that alarm clock had failed to work that morning. If my reliance on an alarm clock, which features in just one of my daily plans, can be this resistant to evidence, then reliance on agents or objects that feature across many interlocking plans is likely to be higher.

This is because of the indirect interactions between beliefs and plans, which come about because of the practical pressures which are also justifying of our plans and the fact that our plans form a defeasible background for action. We may have already set too many other wheels in motion or be too committed to the overall plan to give it up immediately. This does not mean that incoherence between beliefs and relying attitudes would not be criticisable, but it does mean that they are likely to lag and interact in odd ways. Contradictory evidence must push against the dead weight of our plan. It must be sufficient to make us willing to re-plan, and therefore to stop relying.

We ought, therefore, not to be pushed into accepting an affective account on the basis of trust's strange interaction with evidence. As this section has shown, there are alternative explanations available, and they are ones that are at least as plausible. In fact, there are advantages to this account. It explains the close relationship between relying and trusting, for example, better than an emotional account can (the emotion can only presumably motivate us to rely, but it cannot require us to rely). This is why I favour a relying attitude as one of the components of reliance-responsive trust. Trusting in this way will also help us to ensure that we better track the likelihood that the other will be trustworthy.

I have also shown that we can demonstrate a great variety underlying the multiple realisability of some core features of trust, which I hope may remove resistance to my pluralist view.

5.0 Voluntariness

If trust is a judgement and a relying attitude, but if that judgement is bound by second-order reasons which determine how much evidence is required to rationally

make it, then we can easily see how trust may have both voluntary- and involuntary-seeming features.²⁷

The judgement will sometimes require considerable evidence, and so an agent will not feel that they can simply will it. In other cases, such as a trust game, the stakes may be low enough that we can (Holton, 1994, p.63).

On the other hand, relying is voluntary, in most circumstances. You can believe that your friend is trustworthy, but nonetheless you can decide not to rely on her to take you to the airport. There are a number of reasons that you might do this. You might have reasons that you want to carry out a plan that does not involve your reliance on your friend. Perhaps you want to support a local taxi company, or you have just learnt to drive and want to practice. You may not want to put your friend out, and although you consider him trustworthy you may prefer to arrange your own transport. You might not accept the lift from your friend, because you do not want to put the trust between you both out in the open.

This also helps to explain why although trustworthiness merits trust, it does not guarantee it. You do not wrong your friend by refusing to rely on her to do something you consider them trustworthy to do. That is not to say that there are not better and worse ways of not trusting someone. Given that trustworthiness is generally taken to be admirable, we should be clear where we are choosing not to rely that we are doing so for some reason of our own and not because we do not take the other to be trustworthy (Pettit, 2004). We normally make sure of this: we make use of phrases like “I wouldn’t want to put you out” or “I know you’d get me there, but I’ve already made plans.”

5.0 Why trust is contagious

We can rely without making judgements about trustworthiness and equally form judgements about trustworthiness without relying. When we do the latter, we do however face a rational pressure to rely should we need to. After all, we do think that the other agent will act in a way that meets the needs of our reliance. If you think

²⁷ For a good discussion see the opening of Hieronymi (2008). See also Hinchman (2020). Those who view trust as unwillable include Baier (1986, p.244 & Ft.10), Jones (1996, 2012, p.75), and Pettit (1995, p.202). The most prominent defence that it can be willed is Holton (1994).

that your friend is trustworthy to walk your dog, but you do not own one at the moment; should you buy a dog then when considering who to rely on to walk it your judgement of trustworthiness puts pressure on you to rely on your friend. That judgement after all endorses that they will.

I do not, however, think that we should consider a judgement of trustworthiness alone as trust. There seems something wrong, for example, in talking of you being let down or betrayed if your friend turns out to not be able to walk a dog you never relied on him to walk. I think that this is because we should consider a simple judgement of trustworthiness as “trusting” rather than trust.²⁸

We can also easily see why we are responsive to others’ trust if it includes a judgement of trustworthiness. When starting a new job, you may see that everyone directs their finance queries to one particular colleague, even though finance is not part of her job. You might come to trust her advice on some financial problem, despite not having interacted with her before. What mechanism explains the fact that your colleagues’ trust seems to spread to you? On this account the mechanism would be that you infer the judgement of trustworthiness that your colleagues have made (as well as, if we see the trust being fulfilled, gaining evidence that the trusted person is trustworthy). This gives you reason to form a similar judgement. Their trusting serves essentially as testimony to her financial know-how.

Similarly, relying attitudes seem likely to be influenced by other reliance. We commonly take use to be evidence of reliability. When you are deciding to buy a new bike to commute to work, you can look at the kinds of bikes that your colleagues use. Part of the assumptions we often make when we see others using certain objects is that the objects are reliable for the purposes, for which we consider those agents to be using those objects. This suggests that relying is also likely to be either reinforced or encouraged, where we see others relying on some object or agent: we understand relying as implying - through its justification by actual success - a judgement that the relied upon agent or object is reliable. A colleague, who is relied on by other workmates, is therefore likely to become known as reliable, even where this is never

²⁸ Research into trust generally does not make enough use of “trusting”, despite it being a common feature of everyday speech. Trusting usefully captures a proclivity to trust.

explicitly stated. Reliance is therefore likely to breed further reliance. With reliance-responsive trust both components of the attitude contribute to its contagion.

Someone fulfilling their reliance-induced commitments also reinforces the idea that they are trustworthy. In this way trust can be understood as self-reinforcing.²⁹

Successful trust reinforces our judgement that an agent is trustworthy, which in turn gives rational pressure to rely further, leading to more trust, more evidence, and so on.

6.0 Trust and cooperation

If trust includes a judgement that the other will act in the way that you rely on them to act, then it is easy to explain how trust facilitates cooperation. We are more likely to become reliant on those we think will behave as we need them to, and so we face pressure to rely on those we consider trustworthy, and this will enable cooperation.

7.0 Being let down and betrayed

It is standard to note that trust can be responded to with “betrayal” and a weaker form of response: “being let down”. Betrayal has received significant attention; less focus has been given to “being let down”.³⁰ It is enough to note that on this account there is an obligation involved, and so something to which we can appropriately react when another fails to respond. This provides a source of betrayal and being let down, if there is not the distinction that I think there is.

²⁹ Others who have suggested this include Jones (1996, p.960) who describes trust as “trust-philic” and McGeer and Pettit (2017).

³⁰ The distinction is again from Baier (1986, p.235) who in distinguishing reliance and trust says ‘trusting can be betrayed, or at least let down, and not just disappointed’. Despite it being widely accepted as the key marker of trust, ‘betrayal’ has received little philosophical attention and ‘being let down’ even less. It is universally accepted as being a kind of weak betrayal. What discussion there is of betrayal is highly nuanced and interesting. The greatest attention to its relation to trust is given by the discussion of ‘betrayability’ (Kirtan, 2020). As Jackson (2000) argues it can have characteristics of abandonment or desertion (Becker, 1996; ‘The Ambiguities of Betrayal’ in Shklar, 1984, pp.138-191) but also of violation, use and damage. Marder (2020, pp.96-97) suggests that we can ‘faithfully betray’: that trust can be betrayed by doing what it seems one is trusted to do. This is a point, with which I would agree. Though I would say you can faithfully let someone down. That is why I believe the focus of the demands of trustworthiness is separate from what one is trusted to do. How someone is reliant on you settles what reliability one ought to offer; regardless of how they rely on or trust you. Other accounts have noted the potential range of responses to broken trust: see Hinchman (2017) and Bennett (2021).

I think when we do pay attention to being let down, it is not clear that it is simply a weaker form of betrayal. Betrayal includes a sense of abandonment. Being let down does not seem to me to include the same core. If anything, I think it suggests something like a stronger version of disappointment or a sense of having been disregarded, than as if someone has abandoned us.

I think there is good reason to think that being let down is a specific kind of response; and a specific kind of response to someone breaking a reliance-related obligation. It is the way we respond to the fact that they have failed to follow through, as we understood them to be obligated to do in a distinctive way. That there was variation in our response to broken trust would lend further credence to the pluralist view for which I have argued. Nothing rests on this distinction, so I will not push this here.

8.0 Distrust and untrustworthiness

Katherine Hawley has pointed to the importance of an account of trust being able to explain distrust (Hawley, 2014, esp. pp.1-3). I do not wish to go into much detail, as it is unnecessary for the rest of this thesis, but we can consider how this account would also apply to untrustworthiness and distrust.

Reliance-Responsive Untrustworthiness: B is untrustworthy to $x, y, \dots n$, for $A, B \dots N$, if and only if:

FUT1. B is incompetent to $x, y, \dots n$, or

FUT2: B is not independently disposed to fulfil her reliance-induced obligations where she lacks further reasons for doing so, or

FUT3. B does not ensure that she does not invite reliance where she would not be able to fulfil the resulting obligations (in other words, she ignores Due Care)

We can lay out conditions for distrust as follows:

Reliance-Responsive Distrust: A distrusts B to X, if

D1. A relies on B to X

D2. A believes B is “untrustworthy” with regards to X-ing

These conditions for distrust lay out what might be thought of as active distrust. They describe a situation where an agent must rely even though they do not think that the other will behave in a reliable way, even though they are obligated to do so. It may be questioned whether distrust is generally like this.

Such situations do take place. We might think of the difficulty faced by a partner in an abusive relationship, in removing herself from the embedded mesh of reliances involved in many plans, despite the presence of distrust. Of course, in certain situations such extrication may not be possible. Sometimes we have no choice but to rely or are prevented from re-planning. Such situations are deeply harrowing, partly because of the tension between what is practically possible and what the agent believes she ought to do.

Generally, however, we avoid situations of active distrust. More often, it is that we are distrusting of others rather than that we actively distrust them. We simply have the judgement that the other agent will be untrustworthy. A judgement of another's untrustworthiness operates with the opposite effect to a judgement of trustworthiness. It amounts to a judgement that the agent will not follow through on a relied-upon action, and it therefore applies a rational pressure on us not to rely on that agent. It is essentially to think that X will not happen, yet continue to rely on X. We can rely in situations like this where we are committed strongly to the plan in which the reliance is integrated. Generally, however, a rational agent should seek to re-plan in such situations. Where we are yet to come to rely, we would avoid placing reliance on such individuals (absent compelling practical reasons to do so).

That means that standardly we get no further than the judgement of untrustworthiness. I think that this is why we naturally only talk of distrust in someone, i.e. as two place – (D'Cruz, 2020; Faulkner, 2017, p.121): though I also note that the phrase “I do not trust her to X” implies distrusting her in a specific way).

9.0 Conclusion

In this chapter, I have completed the reliance-responsive account of trustworthiness. I have shown how it stands up well to the challenges commonly posed in the philosophy of trust. It is then as good a candidate as any for what

trustworthiness requires. I have also argued for a specific account of trust, which can explain a variety of features commonly found across the literature. This may mean that you accept this account, and that is enough for us to move on and see if it applies to organisations and especially to businesses.

This chapter has also completed the argument which I have pursued across Section II. I argued that we needed a new, more pluralist, and trustworthiness-first approach to understanding our trust-concepts; and that this was because these concepts were about our ability to navigate the various social vulnerabilities that we face in our interactions with each other. I identified the main vulnerability which we face to businesses as being our vulnerability because of how reliant we are on them. I developed an account of trustworthiness which allowed us to reduce this vulnerability robustly and reasonably demandingly. I have shown in this chapter how taking this approach can still lead us to a strong account of our trust-concepts. This account is stronger than others that claim to explain a reliance-based form of trustworthiness and trust, like the “counting-on” and commitment approaches. I do not however think that it can explain all occasions of trust, and I do not claim that it can. Sections 4 to 7 of this chapter are meant to show simply that features of trust can be explained while allowing that trust can have many different forms.

Ideally, the preceding four chapters have demonstrated the merits of the social vulnerability approach. In Section III, we turn to establishing what a business is, and whether it is the sort of thing that can be trustworthy.

Section III: Trustworthy Organisations

7 | Précis of Sections I & II

This chapter summarises the arguments of Sections I and II. In §1, I derive three capacities that an entity needs to be robustly trustworthy: it must be aware of the external effects of its actions, be able to appropriately recognise and reason about its reliance-related obligations and be able to control its behaviour.

*

We are trying to solve a particular problem. An increasing amount of our interactions are with businesses. An increasing number of interactions that take place on our behalf with businesses are focused on trust: that has been the move taken globally by regulators. Yet, it is very unclear that businesses are the kinds of things which we ought to trust. It is certainly unclear if businesses are the kinds of things that can be trustworthy. So, we are trying to get to the bottom of whether they can be: to find the foundations of their trustworthiness. To do this we need to answer two questions. The first of these was:

(1) What do we want when we say we want to be able to trust businesses?

This chapter recaps the argument that I have made to answer this question. It then lays the groundwork for Section III, where we will see if these foundations for trustworthiness are possible and, if they are, describe what a trustworthy business would be like.

Our trust-concepts allow us to navigate a world of interactions, in which we are constantly exposed to social vulnerabilities. These social vulnerabilities vary enormously. Some are ways in which our practical or epistemic agency can be undermined. Others are basic psychological needs: the need to be part of a group, to feel loved, and to be cared for. “Trustworthiness” picks out features of agents which mitigate our vulnerability in interacting with them and which come about through a process of self-regulation: in which an agent can recognise something as a reason for action, reason about it, and have this reason bring about some feature within her. “Trust” picks out a range of simple and complex, cognitive and non-cognitive attitudes which attempt to track these trustworthy features. Trustworthy features vary remarkably in their nature. We should work out what our dominant vulnerability is in an interaction to identify the features that best reduce it. We are

often interested in finding the feature that reduces the vulnerability most robustly across a range of agents, time and domains of action.

Following this approach, I identified our main vulnerability in our interactions with businesses. We are heavily reliant on businesses, and this means that we are vulnerable to them acting in ways which mean our own plans and actions fail. Our actions can fail because reliance picks out a relation in the world, which is a form of dependency. If something we are reliant on fails; then the action which was reliant on it will also fail. Businesses are among the dominant powers in our contemporary world. They are involved across a range of courses of action we can take. Even where we do not directly engage with them; their success undergirds the reliability of the environment in which we live. So, we need a form of trustworthiness which can reduce this vulnerability robustly.

The form of trustworthiness which does this best, I argued, is:

Robust Reliance-Responsive Trustworthiness: *B* is robustly trustworthy to *x*, *y*, ..., *n* for *A*, *B*, ..., *N*, if and only if:

TWY1. *B* is competent to fulfil those reliance-related obligations she produces,

TWY2. *B* is disposed to fulfil her reliance-related obligations, where she lacks further reasons for acting in line with them,

TWY3. *B* ensures that she does not offer reliability where she would not be able to fulfil the resulting obligations (in other words, she adheres to due care).

This requires us to be aware of and appropriately respond to a special class of obligations: our reliance-related obligations. These are non-moral, practical obligations which have a force for us because of standard norms of interpersonal deliberation. When I offer you a reason, I take on some responsibility for its being a good reason. Normally, this responsibility is limited because whether something is a good reason is beyond our control. Often our own reliability is not. We can often make sure that we will behave in reliable ways. We frequently offer our reliability to others as a reason for action. You have probably done so multiple times today: when

you offered to meet your friend for coffee, when you told your partner you would buy ingredients on the way home or when you smiled at the lost student touring your campus. Social norms limit what can reasonably be an offer of reliability. We consider a significant range of activities to be private. Outside of these excluded spheres, we must operate with due care: making it clear when we do not wish to offer our reliability.

If we reason appropriately about these reliance-related obligations, then we can robustly reduce others' vulnerability from their being reliant on us in a way that is not overly demanding. Understanding trustworthiness like this solves a range of puzzling cases of trust. It is the form of trustworthiness we ought to want from businesses, and which if they have then we could appropriately trust them. This allows us to clarify the question to be answered in this section:

- (2) are businesses the sorts of entities which can be robustly reliance-responsive trustworthy?

1.0 The capacity for trustworthiness

To understand if a business can be robustly reliance-responsive trustworthy (for ease, unless otherwise specified this is what I use "trustworthy" to mean), we need to understand what capacities are required to be able to meet the conditions for being trustworthy. These are:

1. Awareness of the external effects of one's actions,
2. The capacity to appropriately recognise and appropriately deliberate about reliance-related obligations consistently, and
3. The capacity to control one's behaviour in line with these reasons.

1.1 Awareness of external effects

The first of these is necessary as part of a perspective on our actions as socially situated. To understand if others are likely to become reliant on us and if we are offering our reliability to them, we need to be able to understand how our own action will affect theirs and where our actions sit in the web of inter-reliance. We need, especially, to be able to tell if an action is going to be perceived by others as an offer of our reliability. As individual people, most of us do this easily. The extent to

which others are reliant on us is likely to be limited, and we can easily perceive how others may or may not be able to become reliant on us. At least, we can if we try.

1.2 Responding to reliance-related obligations

Reliance-related obligations give us reasons, and these are not moral but practical reasons.¹ To be trustworthy we need to be aware of these reasons and to appropriately respond to them: that is, we should understand them as obligating. We ought to not consider other reasons when we have a reliance-related reason, except for any moral reasons or other reasons which override them. If you cannot recognise these reasons, you cannot be trustworthy. Consider an individual who moves to a different country, where the private sphere is smaller than in her home country. She will then potentially offer her reliability when she is unaware of this. She will fail to recognise the relevant reason, and she will also fail to be trustworthy.

An individual who fails to respond appropriately also clearly fails to be trustworthy. You are not trustworthy if your friend tells you a secret and you recognise that you offered to reliably keep schtum about it, then tell a mutual friend who expresses curiosity about it. You are not trustworthy because you have failed to exclude the mutual friend's curiosity from your reasoning as you should have. If we are appropriately responsive to these obligations, then we will be independently disposed to fulfil them. Understanding how we ought to reason about these obligations is also necessary to appropriately exercise due care. It is necessary that this capacity can be exercised consistently if we are to have robust trustworthiness.

1.3 Control

To be trustworthy we also need to be able to control our behaviour in line with those reasons we have. This plays a double role. Firstly, it allows us to fulfil our reliance-related obligations. You are not trustworthy if your friend tells you a secret and you recognise that you offered to reliably keep schtum about it and appropriately ignored your mutual friend's curiosity, but then you told the mutual friend anyway because you carelessly say some of your thoughts aloud. Being unable to control our actions also makes us unable to be robustly competent. Second, control allows us to

¹ If you think that they have a moral basis – as discussed in Chapter 4 – this raises further questions about whether businesses can have moral duties. This is not a question that I address in this thesis. See Chapter 10.

ensure that we are appropriately offering our reliability. It may be difficult to be trustworthy if you have a friendly face which appears to offer reliability (or equally fail to be as trustworthy as you could be if your resting face is angry). This problem will worsen if the lack of control is more serious; we may end up unfeasibly obligated in a way that undermines our trustworthiness.

2.0 Conclusion

The foundations of a trustworthy entity are the capacities to be aware of the external effects of its actions, to appropriately recognise and reason about its reliance-related obligations, and to control its behaviour. We will return to these capacities for trustworthiness in Chapter 7. First, we need to begin to clarify what a business is.

8 | A Joint Obligation Account of Organisations

The aim of this chapter is to lay out a new account of what an organisation is. I argue that organisations are formed by a special kind of reliance-related obligations, where individuals become mutually inter-reliant because they offer each other their reliability in responding to a decision-making structure. §1 discusses a commonsense understanding of what an organisation is and notes that on reflection why we should do what organisations require is unclear. §2 defines the kind of special joint reliance-related obligation in which we are interested. §3 shows how these explain why we should do what organisations require. §4 clarifies the nature of a joint goal, which is part of this special class of reliance-related obligations. §5 lays out this account of organisations, §§6-8 address some specific problems that this account may appear to face when applied to large-scale organisations: entry and exit conditions, alienation, and underdetermination and limits on organisational roles.

*

As we have just seen, to be trustworthy an entity needs to be able to be aware of the external effects of its actions, to appropriately recognise and reason about its reliance-related obligations, and to control its behaviour. We want to know if a business is the kind of entity that can be trustworthy. Before we can answer that, we need to know what kind of an entity a business is. I will not answer this fully until Chapter 12. Part of that answer will be that businesses are a class of organisations. We will start to understand whether a business can be trustworthy by finding out if an organisation can be trustworthy. To do that, we need to understand what sort of a thing an organisation is. In this chapter, I will argue that an organisation is constituted by individuals who are obligated to act in line with a “decision-making structure” (a set including decision-making procedures – e.g. voting systems, hierarchies, constitutions – role-descriptions – e.g. job descriptions – and reasonable derivations from these) because they have mutually offered each other their reliability to do so. The way these decision-making structures lead to decisions and actions by giving reasons to their members means that we ought to consider organisations as collective agents. In the next two chapters, we will see how an

organisation can be a collective agent with the capacities necessary for trustworthiness.

1.0 What is an organisation?

A crowd of shoppers bustling into the Boxing Day sales appears not to be an organisation. The audience at a gig is not one either. Paradigmatic organisations are a committee established to uncover the facts about a national scandal, a company, a legislature, a board of charity trustees or a hobbyists' club.

A first pass at what differentiates an organisation from other kinds of groups is that the individual people who are members of an organisation necessarily have roles within it, and that they operate within a structure which determines (at least in part) what they should do in these roles. Committees have chairs and secretaries, and rules and procedures, as do the other paradigm examples. These are absent from crowds and other “aggregations” of individuals.¹ This suggests a basic definition of an organisation:

A group, *G*, is an organisation if and only if its members have roles and operate within a structure which determines what they will do within these roles.

I think that the basic definition is largely correct. The trouble is that on reflection why we ought to do what an organisation requires us to do is also mysterious.² Why on earth should we do what our role within an organisation demands? What is it that means I ought to follow the outputs of the various structures within that organisation? With little pressing it becomes unclear, for example, why we ought to do what our jobs require.

¹ I take “aggregative” from French (1998b).

² As pointed out by Smith (2023). This chapter and Smith's are concerned with explaining the same phenomenon. Smith points to how we cannot determine everything that seems to be required of a role like this by appeal to having taken on the role (p.243). As will become clear in the main text, I think that we can get around this challenge if we correctly understand that the role to which one is obligated is that of being a “member” and that this gives us derivative reasons to fulfil our specific occupational roles determined by role descriptions and decision-making procedures (which Smith considers occupations). It is this missing ‘master role’ which requires Smith to develop further theoretical machinery to explain why we should do what our jobs require.

In the rest of this chapter, I will develop an account that defends a more complex view of organisations, what our roles in them require, and why we are required to do those things. My account builds on the reliance-related obligations which I developed and defended in Section II. Organisations are formed when individuals “organise”, and to organise is to offer your reliability in a special way. When we organise, we offer our reliability in responding to a range of social phenomena like role-descriptions (e.g. job descriptions) and decision-making procedures (e.g. voting systems, hierarchies, constitutions), which together constitute a decision-making structure. Offering our reliability in this way makes us a member of an organisation, and it is because we are obligated – through a reliance-related obligation – to respond to what is required of us as a member that we have reason to do the specific tasks required when we occupy a specified role (like a chair or secretary of a committee). This picture of organisations helps us explain how roles give us sufficient but defeasible reasons to act and how we gain special standing to criticise fellow organisation members.

These obligations become *joint* obligations because they include offers of reliability in order to achieve some joint goal, and because the kind of inter-reliance which this produces means that we ought to consider the members as collectively obligated to the achievement of that goal.

2.0 What is this special type of reliance-related obligation?

Organising is a way to answer a certain practical question: “What should we do?”

Organisations like a committee seem to settle what the committee’s members ought to do. I believe that we can best understand organisations as being formed by a special type of reliance-related obligation. It involves offering our reliability in a particular way, which answers the question of what we should do. When we are organised, we have a reliance-related obligation to respond to a decision-making structure. This section clarifies what kinds of reliance-related obligation we are interested in, and how they distinctively answer the question, “What should we do?”

2.1 Standard reliance-related obligations

You offer to mow your neighbour’s lawn. In doing so, you offer your reliability to mow her lawn as a reason for her. As I argued in Chapter 4, this gives rise to a reliance-related obligation and so you ought – bar competing moral or other

stronger claims on you – to mow her lawn. This is the standard form of reliance-related obligations, which has particular features.

There is a way in which we might think that offering your reliability to mow a lawn settles some practical matter both for you and for your neighbour. This is certainly the case for you: you should – bar the provisos I have detailed – mow her lawn. That issue is settled. It is unclear, however, that it definitely settles a practical matter for your neighbour in the same way. Generally, I think that offering reliability does not settle a practical matter for the reliant/relying individual in the way that we are interested in it doing so. It seems like your neighbour can, if she wishes, reject this reason that you are offering her. There is also no reason derived from your offer of reliability which settles exactly what she must do. She can use her reliance on you however she likes (she can go and shop, she can take singing lessons, etc.). Equally, she can reject such reliance and, if she wishes, be independent.

Admittedly, she faces some rational pressures not to do this. If she relies on you then she accepts that this practical matter is settled as outside of her control. That is the nature of relying. If she trusts you, as I laid out in Chapter 6, then she has even less reason to do otherwise, as her relying is accompanied by an endorsing judgement. Equally, she will face social pressures: she may not wish to be rude or to signal to you that she does not trust – or even distrusts – you. These are not, however, reasons *directly* related to your offer of reliability and the fact that such pressures remain shows that what she should do remains unsettled. You are just another reliable entity in the world, and so just another part of the background to her deliberations about what to do. Your offer of reliability settles her deliberative background, but it does not settle what she should do. That decision is still hers to make.³

The reliance in standard situations like this goes in a single direction: your neighbour is reliant on you. Plans which she makes which integrate this reliance will

³ I think this is true for other ways we may ensure performance or coordinate without organising. Promising seems similar. Making and accepting a promise settles what the promisor ought to do; but, again, only seems to settle some background deliberative fact for the person to whom the promise is made. This raises doubts about whether promises are – at least standardly – offers in joint deliberation, as suggested by de Kenessey (2020).

fail if you fail to act; but your own plans and actions will not fail depending on what she does.⁴

Standard reliance-related obligations come about within the standard framework of interpersonal practical deliberation. In interpersonal practical deliberation, we can be concerned with what others will do only as far as they settle for us what others will do as part of the background of events against which we must deliberate. We are still thinking about what we are going to do individually. While it may make sense to talk of having in some way settled what we will do, it does so only if we read this “we” in a distributive way: i.e. as having settled what you will do and what she will do. These kinds of situations allow us to coordinate in a loose sense of that term. We coordinate in the same way that a group of robot vacuums might be said to coordinate, when they detect each other and adapt their direction of travel so that each avoids bumping into the others. That is how standard reliance-related obligations work; they are not the kind of reliance-related obligations which I think form the basis for organisations.

2.2 Intersecting Reliance-related obligations with a shared goal

The situation seems different if you offer to maintain your neighbour’s garden, and she offers to grow the plants you will need to do so. This means that you both gain a reliance-related obligation to the other.

This does seem to settle in some way what both of you will do. There is also now reliance in both directions, and you are reliant on each other in a way that plausibly means each of you will fail if the other does not act accordingly. Indeed, it seems to even be appropriate to talk of you as having a shared goal: such as “improving your neighbour’s garden”.

It may make sense in this scenario for you both to talk of what “we should do”, but we need to be careful. There is a way that this scenario can come about which would still mean that it was only appropriate to do so in a distributive sense. To see how, we can consider an analogous case:

⁴ I put aside eccentricities like your reliance on her not stabbing you as you start up the lawnmower.

I make nuts and you make bolts. We each share a common goal: to increase our sales to a large DIY chain. That is, I have the goal of increasing the sale of my nuts to that DIY chain and you have the goal of increasing the sale of your bolts to that DIY chain. We may independently find advantage in achieving this goal in encouraging the DIY chain to deal with the other; and this in turn might make us highly inter-reliant. This can produce reliance-related obligations between us.

We are pursuing independent, though overlapping, plans. We have offered each other our reliability in pursuit of the personal plan which we each have. When you and your neighbour improve her garden, you can do so in this way.

This situation is best understood as two related and intersecting, but standard individual, reliance-related obligations. They each settle only that the other will do something, which forms a background against which each party can deliberate though, admittedly, a particularly prominent part of that background. There are still only your own individual obligations and your individual reasoning. When we think about how you will each act, we can only look to your individual reasoning to explain what happens.⁵

2.3 Joint Reliance-Related Obligations

There is, however, a different kind of situation though it can appear identical. It differs because it involves offering a specific kind of reliability. To see how, we can consider an explicit example of offering this kind of reliability.

⁵ This, I think, shows that “mutual reliance” – which as discussed by Alonso (2009) amounts to intersecting reliance as I have described it here – is insufficient for collective agency. We need something further, as I argue in the main text. I think that it is not the acceptance of interconnected reliance in Alonso’s dance example which makes them form a group. It is that in agreeing to dance we openly offer our reliability in the way I discuss in the main text, and thereby bind ourselves to be responsive to a decision-making structure (in the case of a traditional dance this consists of an implicit role-description derived from a dance pairing – a leader and a follower – and a decision-making procedure formed from paying attention and responding to the physical cues both give to determine specific movements). My account also differs from Alonso’s because I do not think that the underlying obligations are moral (see §7); and because we disagree on the nature of reliance – he thinks it is essentially practical acceptance (Alonso, 2014) and that a relying attitude is necessary for the underlying interpersonal obligation.

Say that you are talking to your neighbour, and she says, “I wish I could improve my garden” and you offer to help her do so.

“What should we do?” she says.

“Well, we better decide” you reply, to which she responds, “Yes. How about I grow the plants, and you do the grunt work?”

I think that you “organised” when you suggested that the two of you needed to reach a decision and your neighbour agreed. This is because that exchange involved offering your reliability in a particular way. You offered to reliably respond to whatever features of your continued interaction determined what you should do (I will call those features a “decision-making structure”) in order to achieve some goal.

In this scenario, what determined this answer may have ended with you accepting the suggestion of your neighbour, though you may have not done so and continued until you found a mutually acceptable outcome through your conversation. What I hope to show is that whatever happened, the reason you should have performed that role derives from your original offer to be reliably responsive to a decision-making structure.

I think that when we offer our reliability in this way, we take on a special form of *joint* reliance-responsive obligation. The obligation is joint because it creates a special kind of mutual reliance. The offer of reliability is constrained by some goal, and this goal becomes a *joint* goal because it is – in part – constitutive of it that it is achieved *together* (see §4). This contrasts with a shared goal, where *who* it is achieved with is incidental to its achievement (as it is if you and I sell bolts and nuts). The reliable responsiveness to some decision-making structure enables the pursuit of this joint goal (and – as we will see in the next chapter – gives the organisation its own reasons). By organising, we get an answer to what we should do, where that “we” is meant in a strong, *joint* sense.⁶

⁶ I use “joint” to refer to the strong sense of “together” which is explained by Gilbert (1992, Ch.4).

This is the class of reliance-related obligations which I believe underlie organising and organisations. They have the general form:

Joint reliance-related obligation: I offer my reliability to respond appropriately to a decision-making structure, so that we can achieve our joint goal.^{7,8}

⁷ The agent need not conceive of being obligated in this way. I think that we are familiar with taking on such obligations, but we never think of them in these theoretical terms. This is in line with Gilbert's thoughts throughout her work on how joint commitments are not thought of as such.

⁸ Bratman (2014, Ch.5, pp.107-120), argues that his 'basic thesis' provides an explanation of 'modest sociality' which renders Margaret Gilbert's joint commitment essentially redundant: as he explains it, the pressure to jointly cohere plans is a rational pressure on each individual agent formed by the appropriate intermeshing of their individual plans and standard elements of planning rationality, such as a need for coherence and consistency. Some of this criticism affects my account, and so I address it here (chiefly that in §3 of Bratman). First, Bratman points to how 'mutual obligation' appears insufficient to produce 'shared intention'. He gives the example of two insincere promises to plough a field, which seem to appropriately obligate but which do not actually result in action. It is possible that people insincerely offer their responsiveness to a decision-making structure as well. Second, Bratman makes a series of criticisms that Gilbert's model fails to appropriately distinguish mutual obligation and shared intention; essentially conflating the two, and that while he recognises the potential existence of mutual obligation, he considers it separate from shared intention, which is better explained as an appropriate intermeshing of plans.

I have four responses to this. (A) As I have said, the obligations on my account wither when no actual reliance is produced: without any actual inter-reliance, the individuals' reliability ceases to be a good reason for the other. Unlike Gilbert's joint commitments, my obligations do not arise in situations where there is no actual mutual inter-reliance. They are not subject to a counterexample like the jointly insincere promisors. (B) Bratman's own theory relies heavily on an idea of 'mutual responsiveness' which leads individuals to appropriately inter-mesh their plans. Reliance-related obligations are one way to clarify how mutual responsiveness can come about. Without something like these sorts of obligations in our theory the reason that we ought to accept something like a decision-making structure remains rather mysterious. (C) Bratman makes much of the relative parsimoniousness of his own theory over Gilbert's (see §4 of Bratman). My own approach has two advantages over Gilbert's and *contra* Bratman: (i) I do not appeal, as Gilbert does, to a *sui generis* form of normativity. I do not claim that joint reliance-related obligations are some sort of primitive. They are a special form of an obligation which is rooted in something that Bratman would accept: interpersonal norms of practical rationality. I do not think that I am postulating anything that requires basic theoretical commitments beyond those Bratman himself holds. (ii) Completeness is as much a theoretical virtue as parsimony. In at least some cases of shared intention, we do seem to acquire a standing to criticise those who fail to jointly intend in a way that suggests obligation, in a way that seems absent from Bratman's basic picture. In the case that I introduce shortly in the main text, Alan and Barbara seem to have the "standing" to criticise Cheryl for her performance in her role in a way that an unrelated observer does not. This sort of standing normally follows some kind of obligation directed towards those with the standing. So, given (i) and (ii), joint reliance-related obligations are a further theoretical construct in our theory, but one which does not introduce anything extra into our basic theoretical commitments and which can more fully explain the range of phenomena we

The next sections will further justify why we should accept that there are joint reliance-related obligations, provide further detail on what they are, explain how they are joint, and show how they allow us to organise.

3.0 The nature of organisational roles & decision-making structures

When we offer to reliably respond to some decision-making structure, we gain a standard reliance-related obligation but with a specific content: responsiveness to a decision-making structure. To explain this, I first need to clarify some further aspects of roles and what I mean by a decision-making structure.

A useful idea here is that of a “master role”.⁹ A master role is one which forms the normative basis for subordinate and derived roles. That is, it is the ultimate source of those roles’ normative force and the reasons we have to act as those roles require can be explained by the master role.

We can think of offering our reliability to respond to a decision-making structure as taking on a master role, and I will call this master role “member of the organisation” (for ease, just member or membership).

Being a member means that we are obligated to respond to a decision-making structure. A decision-making structure are those aspects of our continued interactions which allow us to answer the question of “How should we do this?”

encounter in our social world. (D) I intend this only as an account of a prominent form of sociality, not as the only potential account of sociality. I note that there are significant crossovers between this view and that of Bratman’s as presented more widely in his work especially if we view joint reliance-related obligations as a way of producing ‘mutual responsiveness’.

⁹ The term comes from Blackman (2023, p.161). The map of the conceptual terrain that Blackman provides for considering role-based reasons is very useful. The account offered here is normative, direct and hierarchical. It is normative because it takes the basis of the normativity of role-based reasons to itself be some normative fact. It is direct because it denies ‘that the normativity of roles is explained by non-role-based normative facts’ (p.166). It is hierarchical because it argues that ‘some roles are normatively prior to other roles in the sense that the normativity of some roles explains the normativity of others’ (p.161). This account is attractive in part because it is a hierarchical view that does not need to appeal to a universalising or moralised ‘master role’. As, for example, the Aristotelian account constructed by Blackman does which appeals to a “human” or “rational being” as the master role. The limited scope of the account (focusing on organisational rather than social roles) keeps it theoretically simply in a way Blackman suggests is difficult for such views (p.167).

There are two main components of a decision-making structure: decision-making procedures and role-descriptions. A decision-making procedure can take various forms.¹⁰ They include all of the following: pulling out a straw and doing some action if the shortest straw is pulled; offering and being responsive in a discussion aimed at reaching a consensus; raising one's hand at the required time and following the outcome which attracts a simple majority's support; following the instructions of one individual; and standing at a specific spot to observe a woodpecker and avoiding an action should it fly in a specific way which we take to express a god's will.¹¹ We should stay open-minded as to what kinds of decision-making procedures may be adopted by an organisation, as there seems to be little reason to limit their scope. All that is required is that they are ways that a group of individuals can gain an answer to the question "How should we do this?" and that they can offer to reliably respond to them.¹²

We can see how we can organise in line with this: how offering our reliable responsiveness to a decision-making structure seems to place us in an intuitive organisation. Alan, Barbara and Cheryl may be three friends considering how to sell artworks at a local craft fair. They may explicitly agree that they will reach decisions about what each of them should do by discussion until consensus is reached. In making this agreement, they offer each other their own reliability in responding to the consensus position. In doing so, they take on a reliance-related obligation.

We can also make use of role-descriptions to decide how each of us should act. A role-description will standardly be an explicit description of the tasks associated with the role. These may take the form of lists of tasks or role responsibilities, as they do, for example, in a job description. They can also be formed when each party

¹⁰ I mean decision-making procedures, broadly in line with how it is standardly used in the collective intentionality literature. See French (1979) and List and Pettit (2013).

¹¹ That the latter is a "supernatural" form of decision-making does not seem to be reason to reject it as part of a category of decision-making procedures (though it may give us reason to reject it as a *good* decision-making procedure).

¹² I am not claiming that decision-making procedures may not be reason-giving independently of this force. For example, the commands of God may be reason-giving and could be adopted by the Roman Catholic Church. They may have independent normative force. My argument is that they will also have normative force as requirements of a role, and that that is what is explained by joint obligation.

explains what she will do, or the tasks required may be implicit in a broader (societal) conception of what the role requires – we know for example, broadly what a teacher does (though see my discussion of societal roles below, which I am not attempting to explain). Alan, Barbara and Cheryl may determine specific role descriptions. Cheryl may say she will be the artist; Alan that he will be the salesman; and Barbara that she will do the accounts. In offering to perform these roles, they each offer their reliability to be responsive to the requirements of the role: that is, they offer their reliability in completing the tasks which are required of them.

Decision-making procedures and role-descriptions affect and interact with each other. For example, a decision-making procedure can determine what to do when a role-description is underdetermined (see §8). While a role-description can determine how one is allowed to input into a decision-making procedure, and to which of its outputs you will have to respond. Cheryl may be unsure what sort of artwork to produce, and the three of them may therefore seek a consensus that specifies this. Alternatively, they may accept that Alan and Barbara do not get to decide what kind of art Cheryl makes; and that within their roles they have to accept her decision on this.

We can call the total range of requirements that affect an individual and are derived from a decision-making structure “job roles”. Secretary, lecturer, abattoir worker and boatswain are all job roles. Some job roles are less paradigmatic. Your maintenance of your neighbour’s garden can amount to a job role. Together with membership they constitute our organisational roles.

The reasons we have to do what our job roles require are derivative. When we become obligated to respond to a decision-making structure, this involves responding to those specific requirements of our job roles, which are determined by that decision-making structure. This means that the requirements of our job roles originate in our membership.

These reasons give us *pro tanto* reasons to do what our roles require, as reliance-related obligations do generally. They are also exclusionary, which matches the experience of an organisational role-requirement. Finding your work tedious is (unfortunately) not usually a reason you can consider when deciding whether to do

it. These parallels give further weight to an account resting on reliance-related obligations.

To be clear, I am claiming only that this provides an explanation for the nature of organisational roles and why they come to give us reasons for action. I think that organisational roles likely have a distinct nature from other social roles. Though I cannot establish that here, and so simply bracket the wider question. We can recognise an intuitive difference. Barbara may be a financial officer and the wife of Cheryl. Both “financial officer” and “wife” are roles. The former is largely a job role. The latter is chiefly a societal role. There is often crossover between organisational and societal roles. A teacher occupies a societal role but will also often have an organisational role. The societal role of “teacher” is subject to a wide range of norms to which an occupier will be subject, such as the prioritisation of the education of her pupils over her professional advancement. The school which employs her in the job role of a “German teacher” will also make specific demands of her. The relationship and priority of the requirements of societal and organisational roles is also not something that I will resolve here. I am only defending an account of *organisational roles*.¹³

¹³ This is one difference between my account and Margaret Gilbert’s account of joint commitment, which she has developed in a series of prominent and influential works since the 1980s (see especially Gilbert, 1992, Ch.4, 1990, and 2008, Chs.7 & 8). I do not want to endorse any of Gilbert’s view beyond that which I explicitly endorse here. The picture I present may also not be one Gilbert would want to endorse. My account is far more minimal in its goals than Gilbert’s. Gilbert claims that joint commitments are the “social atom”: the fundamental building block of our entire social world, potentially including morality. The “social atom” claim is found throughout Gilbert’s work and defended most extensively in Gilbert (1992). She makes the claim regarding morality in (Gilbert, 2008, p.158.). I do not agree with either view. My ambition is smaller. The limits of my aim in this chapter are to provide a plausible account of what organisations are. I am not claiming that it is the only plausible account of organisations (or of their agency). I make no claims about the ontology of anything other than organisational roles; I certainly make no general claims about the nature of social roles or social life. My account may be understood as an attempt to defend a reductive and minimal version of Gilbert’s account. In particular, I try to do away with a particular objectionable part – in my view – of Gilbert’s picture: the idea that joint commitment has a *sui generis* form of normativity. Gilbert has described the joint commitment as a way to commit one’s ‘will’ in such a way that another party gains ‘ownership’ over one’s actions – (Gilbert 2008, pp.127-134, on “ownership” p.153). It is the notion of ownership that is most unclear. It seems to be intended to explain how one gains standing, but it is unclear how one gains ownership when Gilbert considers the other to only be committing her will in the same way that we decide or form an intention. It is also unclear that ownership is the right way to characterise owing another some outcome. Say that you are harmed by falling masonry on an ill-maintained

There are two ways such requirements may be limited. You may offer your reliability in responding to only a limited point. For example, it may be implicit in your offer to decide with your neighbour how to garden that you will not respond to immoral suggestions (like raiding the nearest garden centre). Alternatively, when joining a pre-existing organisation, you may be required to offer your reliability in a specified way. For example, as an engineer you may be required to offer your reliability to respond to a decision-making structure in line with reliably acting as required by your professional body's regulations. Failures to behave as you have offered to reliably behave mean you fail to meet the demands of your obligation. As we will see in §8, these conditions on your offer are also important in determining what your role can require of you.

4.0 Putting the joint in joint obligations

I have called these obligations *joint* reliance-related obligations, and they are joint because they can only be fulfilled by achieving a joint goal. A joint goal is one that can only be achieved by the individuals acting together, because their acting together is part of that goal.

This need not be the case for goals held by more than one person. As we saw, we may both have the shared goal of selling more to a national DIY chain. This is a shared goal because while it has the same content, it is not constitutive of the goal for either of us that we do it with the other.

Joint goals can be goals to achieve a form of action which is only achievable collectively. Examples include playing a duet, dancing the conga, playing cricket and co-authoring a paper. We may also adopt a joint goal because we think some end is better achieved cooperatively, even though it may be possible (if not feasible) to achieve the goal alone. Examples include selling artworks, building a cathedral, producing and selling motorcars globally and providing retail banking services.

building and so are owed compensation of £10,000. It would be wrong to think that you now own that £10,000: you have a right to it, but part of recognising that you are *owed* it involves recognising that it is someone else's. I am unclear exactly how joint commitments as Gilbert understands them are meant to gain their normative force on us. I try to get around this using the idea of reliance-related obligations which I developed in Chapter 4, as I argue in the main text. This provides a better grounding and simpler explanation than that of Gilbert.

We can also adopt self-referential joint goals. These are simply goals to do *something* together, but which leave it open as to what that specific aim is. You and a friend may simply get along and fancy doing something together. We can then offer each other's reliability in the ways described above and so become members of that organisation; but we can leave it up to the decision-making structure of that organisation what exactly it is that we will do. When we decide what to do, we then collectively adopt this as our goal. All these goals are joint because they cannot be achieved by an individual; they have to be done together.

An organisation may have a complex concatenation of goals. An organisation may start off selling art, but it may come to adopt the goal of increasing its profits yearly as well. This may push it to adopt further goals such as providing art workshops or producing artistic materials. These goals again need not be explicitly conceived in these ways by members. Goals can be interpreted from the overall orientation towards an end that an organisation's decision-making structure pushes it towards. Alan, Barbara and Cheryl may claim to not care about profit, working only out of a love of art. If, however, they make decisions on what to do based on what is the most profitable course of action, we may appropriately consider this one of their joint goals.

Shared goals will often give rise to joint goals. The best way for you and me to sell your bolts and my nuts to the national DIY chain may be for us to organise; to adopt a joint goal of selling our nuts and bolts to them, and to offer our reliability in adapting appropriately so that we can achieve this. Sometimes they will be explicitly adopted; they may, for example, be the object of a specific agreement – as when Alan, Barbara and Cheryl agree to sell artwork – or feature in a constitutional document, which explicitly lays out what the organisation is for. In this way, they are integrated into the decision-making structure and usually set the bounds of that structure.

Joint goals are part of the obligation because they form the boundaries of the reliability that we offer each other. Suggesting a joint goal is often a precursor to offering one's reliability to respond to a decision-making structure. They set what we should do, and our reliability in responding to a decision-making structure then

allows each of us to determine what we each must do to achieve this goal. It is constitutive of our doing it, however, that we do it together.¹⁴ To put it in the terms introduced in Chapter 4, each member has endorsed and owns their own reliability in responding to a decision-making procedure as a reason for all members to pursue some joint goal together.

The nature of the reliance in these situations differs from the one directional and intersecting reliance we have already seen, in a way that I think can appropriately be thought of as mutual reliance. The reliability we offer is limited: we offer a reliable responsiveness to a decision-making structure towards some particular end. Within that scope, the reliability is highly open. We offer to respond appropriately to whatever things it is we are required to do by the decision-making structure. For this to work, we need each member to also be responsive. If they are not, then we cannot achieve our overall joint goal and we cannot therefore fulfil the reliance-related obligation which we have (as fulfilling a reliance-related obligation requires performing the action it specifies). So, we can only fulfil these obligations jointly.

5.0 What an organisation is

I have argued that “to organise” is to accept a special kind of joint reliance-related obligation. We can offer our reliability to respond appropriately to a decision-making structure, so that we can achieve our joint goal, and when we do then we can properly be thought of as having organised. In doing so, we become a member of the organisation. Our job roles are specified by this decision-making structure, and our obligation to respond to it is why we ought to do what our job roles require. This explains those otherwise mysterious features which we found in our basic definition of an organisation.

These sorts of obligations can only be fulfilled if we act together, and if we do so in a specific way – by responding to a decision-making structure. They are not fulfilled by each member being reliably responsive, they are only fulfilled if all the members are reliably responsive in such a way that together they have appropriately tried to achieve their joint goal.

¹⁴ This should be read to include the fact that we may only each know who we are doing this with under some generic description like “the other members of the organisation”.

This would mean that properly understood an organisation is constituted by all those individuals who are its members; that is, who have offered each other their reliability to respond to a decision-making structure in pursuit of a joint goal.

This seems expansive. This account does allow for small, informal groups where mutual inter-reliance is present in the right way to be organisations. Say that I regularly ring my mother, but that my mobile signal frequently drops out. We may have adopted a decision-making structure, whereby I am the one to call back once I have signal. I have offered to reliably do so; my mother has offered to reliably avoid calling herself and jamming the line. This is an organisation on my account. Is that worrying? I do not think that it is. Though it is odd in everyday speech to say that there is an organisation here, it is not to consider my mother and me as having organised ourselves. I am happy for there to be many such small, limited and *ad hoc* organisations.¹⁵ Given the focus of this thesis, I will not spend much time focusing on these sorts of organisation.

There is another potential objection. What should we make of an individual like Daniel, who offers his reliability to Alan, Barbara and Cheryl, responds to their decision-making structure in pursuit of a joint goal, but is not recognised as a member by Alan, Barbara and Cheryl? I do not think that Daniel is a member of the organisation, and the reason for this lies in the nature of the others' reliance on him. The others are not reliant on Daniel being responsive to the decision-making structure, and they can ensure this by acting so that they can all act independently of him to ensure this. If this is the case, then Daniel does not have a good reason to continue to offer his reliability. He has no reason to continue to endorse his reliability, and his reliance-related obligation will wither, as they standardly do when an individual is not actually reliant on us. Daniel therefore does not have a joint reliance-related obligation and so is not a member.

The account can also handle large organisations. In the next three sections, I will attempt to address some specific problems that this account might face when

¹⁵ To emphasise how limited this is, I am not saying that my mother and I form a general organisation. The nature of our joint reliance-related obligation is limited by our joint goal: calling each other back. It is only in this very limited domain that we form an organisation.

applied to larger organisations and to provide further detail on the nature of joint obligations.

6.0 Entering and exiting a joint reliance-responsive organisation

6.1 Ways to enter

Joint reliance-related obligations are still reliance-related obligations, and so the observations about how we can become obligated in this way which I made in Chapter 4 still apply. We need to be aware of what others may reasonably rely on us to do. Many similar operative social and moral norms may apply. Some of these seem likely to be more stringent. It may be appropriate for a standard obligation to arise directed towards someone who you do not know well, but privacy norms may be more stringent for joint obligations to such strangers. This helps explain what feels strange when a stranger attempts to walk somewhere together with us; she has transgressed that norm, which more tightly limits our joint reliance obligations. We can avoid acquiring them by acting in line with due care.

We can become jointly obligated through ‘basic’ entry (Gilbert, 2008, pp.138-141 – my account better deals with questions about how such obligations come about over time and without explicit commitments). Alan, Barbara and Cheryl may be talking over dinner one night when Alan says, “Shall we work together to sell Cheryl’s art?” raising a potential joint goal. Barbara and Cheryl agree. In doing so, they have directly and explicitly accepted the importance of their reliability to each other in bringing it about that they sell art, and they have adopted a joint goal. They have left this reliability open, in a way that requires their responsiveness to specifying requirements if they are to be successful in achieving that joint goal. That is, they have jointly obligated themselves as described above.

We need not explicitly invite others’ reliance to enter a joint obligation. If Alan, Barbara and Cheryl simply start “helping each other out”, but do not exercise due care in doing so, then they can offer their reliability and end up forming obligations as we can whenever we fail to exercise due care.

These kinds of explicit offers seem unlikely in a large-scale organisation. An employee of a large firm can surely hardly be understood to have offered her reliability to the other members. That is right, but it does not pose a problem for this account.

The decision-making structure may specify who else can be added to the organisation. Specific individuals' job roles may require them to decide whose offers of reliability to integrate into the organisation as members.¹⁶ That is, they may have the authority (because of their role or through a decision of the organisation) to accept others' offer of reliability of responsiveness and to allow them to become mutually inter-reliant. This will mean that their job role requires them to get others to jointly obligate themselves. Where this is appropriately derived from the overarching "membership" obligation, then the other parties will be obliged to accept this as doing so will be part of being appropriately responsive to the decision-making structure.

In large-scale organisations there are also usually specific ways of offering our reliability. In many large-scale organisations this is done through the acceptance of a contract, like an employment contract or membership terms. We ought though to understand the contract as a sign that we are offering our reliability to be responsive in the right way, rather than as the actual source of our membership and its related reasons – though, legal reasons may be further reasons to act as our role requires. This means that legal and organisational roles and legal organisations and actual organisations can mismatch. Understanding organisations like this explains situations like those where we have started to take on other responsibilities in ways that mean our current actual role does not match our legal job description. We in part have employment practices like annual reviews to allow us to correct for such drift. If we allow only for the legal entities, then this drift becomes mysterious. It would seem that we are mistaken in thinking we have any reason to do what is not in our legal job description. If this is right, then this means we have no reason at all to do these different tasks. This seems, on the one hand, correct. We sometimes may feel that we do not have reason to do anything beyond our legal job description. My

¹⁶ This is, I think, a simpler model than that proposed by Gilbert – e.g., (Gilbert, 2008, pp.140-3 for basic cases, pp.179-180 for non-basic cases). These are not special forms of joint obligation (as Gilbert suggests is required for joint commitments). They are simply particular ways of being responsive to a general organisational joint obligation. This account does not require common knowledge for a joint obligation to arise, in the way that Gilbert's does (Gilbert, 2008, pp.138-141). Though I note that usually common knowledge is achieved as we are aware of others to whom we owe our reliability at least under a generic description such as "member of this organisation".

joint obligation account can explain this. It is because we offered our reliability only to do what was within that initial scope of our role-description. On the other hand, there does seem to be some reason to do these other tasks, and this account can explain why. We did in fact signal a broader reliable responsiveness to the decision-making structure, and this has determined that we ought to do these further tasks. As the legalistic account can only explain the former of these, we ought to prefer my account as the more complete.

6.2 *Ways to exit*

A joint obligation will end when it is fulfilled. That is, when the joint goal has been achieved. Say that you and a colleague decide to co-author a paper on populism. You finish it, submit it, and it is accepted. Your joint obligation is fulfilled. As it is fulfilled, you are no longer obligated and so you are no longer organised.

You might also mutually agree to end a joint obligation. This may happen explicitly. Alternatively, it can happen because your mutual reliance dissipates. Alan, Barbara and Cheryl may find that selling the art at the art fair is taking up too much time and giving them too little reward. They may each begin to focus more on other tasks, which make them increasingly less reliant on each other. This may eventually happen to a great enough degree that they are no longer obligated to each other, because a lack of reliance stops us from having reliance-related obligations.

You cannot, however, unilaterally free yourself from a joint obligation. This is because your “membership” obligation requires you to be reliable in responding to a decision-making structure to achieve an overall goal. This means that unless the overall goal is achieved or the decision-making structure determines that a member can leave, she cannot do so without breaking the obligation that she has. Leaving unilaterally is to ignore the way that membership obligates us to others. It is to fail to reliably respond to the decision-making structure by simply refusing to be subject to it.¹⁷

¹⁷ This offers a much clearer way in which joint obligations cannot be unilaterally rescinded than is available on Gilbert’s joint commitment account. Again, specific decision-making structures may alter who has to respond to relieve you of your reliance-related obligations.

You may however be able to leave an organisation where this is accepted by others and steps are taken to end the mutual inter-reliance in which you stand. This frees you from being party to the joint organisation. Alan, for example, may find that selling the artwork is reducing his time with his family. He may negotiate with Barbara and Cheryl that he will recruit a new member to take on his responsibilities; and to end his personal part in their joint obligation. Similarly, were Alan to violate his obligation significantly enough (for example by selling stolen artworks) then he may “exit” the organisation by Barbara and Cheryl shifting their mutual reliance so that he is no longer included.

These forms of entry and exit are prevalent in actual organisations. Consider taking on a new job. In signing an employment contract, you arguably express a readiness to be jointly obligated in the way that I have outlined. A key element of this will be that you take responsibility for your reliability in responding to how the organisation reaches decisions, to accepting its management structures and the way that tasks will be assigned to you: that is, you will accept an organisational joint obligation. You also accept that your employer’s decision-making structure can determine when you ought to leave the organisation and agree to be responsive to that.

7.0 Alienation

It may be objected that this is not how many people join organisations, particularly business organisations and cases of employment. The objector accuses me of idealism. The above works, but it has no relevance to how we actually join organisations, especially not businesses. Look at how we approach our work. Many (most or all) of us just turn up, act because of other reasons (like making sure we get paid) and go home. We are alienated from our jobs. Few (next to none or none) of us seem to be jointly obligated in the way my account would have it.

There is something to this objection, but it is important to not overstate it and let its intuitive appeal run ahead of its strength. The objection rests on the fact that individuals who join organisations are alienated from the organisation’s decision-making structure. Its appeal comes from an ambiguity in our notion of alienation. One of these senses is limiting on my account; the other is not.

The less worrying form of alienation might be called task-alienation or compartmentalisation.¹⁸ This involves an individual understanding her work as part of some broader, overarching joint enterprise; but not identifying herself with that joint enterprise. She may view her work as somehow not an authentic part of who she is or something similar. This does not preclude joint obligation, in fact, the kind of understanding that joint obligation requires seems implicit within her view of her role. The compartmentalised individual understands the role she has *qua* member as part of an organisation's joint enterprise; that is why she recognises the reasons that it provides. She simply wishes that she were not party to that obligation or refuses to identify herself with it. She is reluctantly obligated, but that does not stop her from being obligated.

The more worrying form of alienation is “full alienation”. Here, an individual is simply going through the motions of occupying some role. She views her tasks as essentially meaningless and pursues them simply for unrelated personal ends (e.g., wages). The individual will have to be careful not to offer her reliability, as this could still obligate her; but she may be able to do so. It may be clear that she is doing this only for the cash. In these cases, I do agree that she is not a “member of the organisation”. Her organisational role is not a source of reasons for her. This is the right result. It seems right to say that such an alienated person does not have reasons to act as her job demands. It would seem odd for her to say, “I do it because it's my job” in the sense that the requirements of her job are an independent source of reasons for her. We would need to interpret her as meaning, “I do it because it's what I get paid to do” which refers to her personal reasons. She is then not properly thought of as a member, and so she is not really part of the organisation.

The fully alienated individual will often still act in line with what her job requires. When she does, her actions can still be appropriately counted as the organisation's actions as they appropriately derive from the organisational decision-making structure, and so ultimately are why that task is done rather than some other. So,

¹⁸ Which may or may not be the right thing to do (cf. Brunning, 2023).

the organisation still gives the reasons why that specific task is done, even if it does not feature appropriately in the alienated individual's reasoning.

I think that cases of genuine full alienation may not be as prevalent as we may think, though this is of course an empirical question. The requirements to offer mutual reliability in the way required for a joint obligation are quite thin and seem likely to happen quite frequently even in large-scale organisations.

When it does happen, I do not think that it need worry us that an organisation sometimes acts through something that is not a part of it. Humans use tools to act; we are responsible for the use to which we put the tool. It does not matter if I hit a stranger with my fist or a cricket bat; I still hit her. An alienated individual can be a tool for an organisation. They are means to produce its actions. The action is still the organisation's, as I will argue in the next chapter.¹⁹

8.0 Constitutions, underdetermination and the limits of organisational roles

8.1 Constitutional principles

Joint obligations are offers of reliability in responding to a decision-making structure in order to achieve some joint goal. It is important to note the limits that such obligations impose on what an organisation can require (which are separate from those reasons that cancel joint obligations or mean they should never have come about) and which help when elements of a role are underdetermined, as I discuss in the rest of this section.

I call these “constitutional” principles because they rule out certain demands. They can take the form of stated or unstated but commonly acknowledged overarching principles or policies, which are binding on role-requirements and decision-making

¹⁹ I am not going to attempt to defend the moral responsibility of organisations here. If you are sceptical, then I refer you to the arguments in List and Pettit (2013) (especially on programme control which fits the picture offered here) and in Gilbert (2000, Ch.7 ‘Collective Remorse’), on functional analogues of the reactive attitudes. I think a convincing argument for their moral responsibility can be found in some combination of these, which addresses concerns such as those that organisations lack appropriate self-reactive attitudes.

procedures.²⁰ Sometimes these will be explicitly laid out in key documents stating the organisation's purpose, etc., but often they will not be.

Every organisation has at least one constitutional principle: its joint goal.

Understanding why this is a constitutional principle helps us to understand how others can come about and become binding on decision-making structures.

The joint goal is the end for the achievement of which someone offers her reliable responsiveness to an organisation's decision-making structure. As with standard reliance-responsive obligations this offer sets the limit of what can be demanded from the obliged person. Say that Cheryl and Barbara vote that Alan should clean their cars, in a way that counts as an exercise of the decision-making procedure that they have adopted. We can assume that Alan's cleaning of their cars does not contribute to their joint goal of selling art. Selling art is, though, what Alan offered his reliable responsiveness to the decision-making structure to do. Alan is, therefore, not obligated to be responsive in this way. So, he does not have to respond to this output of the decision-making procedure. The limits of what he offered set the limits of what he can be obligated to do. This – in effect – renders the decision that Alan must clean the car (absent other reasons he ought to do so) void. It is in this sense constitutional.

Organisations can also stipulate what kind of reliability someone must offer when she joins. It may require that she offer her reliability in a specific way (say, that she must conform with legal norms or uphold certain values) within her role. If someone offers their reliability and she is aware of these requirements, then she ought to recognise that they constitute part of how she has offered to reliably respond to the decision-making structure. For example, Barbara may have been required to offer to take on her role while adhering to the professional standards of accountants. As with the joint goal, when these are contradicted then she will not be obligated unless there is some further overriding reason.

²⁰ The "commonly acknowledged" requirement allows for the fact that in offering mutual reliability individuals can expect some reasonable limits on what may be required of them such as compliance with operative social norms in that community or moral norms.

8.2 Underdetermination

I accept that often role-descriptions will leave it underdetermined exactly what is required of a role-occupier. Circumstances change and it may often be unfeasible to fully detail everything that a specific role would be required to do. Where role-descriptions leave this underdetermined, they will often refer a role-occupier to a specific decision-making procedure. They will require specific requirements to be determined by more senior roles or to require a consensus to be found across all members. Even with a decision-making procedure, however, we may still be left with underdetermination of what is required. How, for example, does the more senior role-occupier correctly determine what ought to be done? Decision-making structures may also leave it up to the role-occupier to decide what she ought to do. That is, they may leave it at her discretion. How should she determine the exact requirements of her role?

A role-occupier will be free to reason about what she has to do, as far as this is reasonable within the confines of the decision-making structure which determines her organisational role. She ought not to reason in ways that contradict this, because she is obligated to be responsive to it. This will standardly mean that we have to pay attention in her reasoning to anything that is incorporated into the organisation's decision-making structure. For example, organisations may adopt specific principles. Alan, Barbara and Cheryl may have said as they agreed to sell artwork together, "But let's make sure it's always fun." This would limit the scope of what they can require of each other. Barbara and Cheryl may not be able to instruct Alan to drive eight-hour round trips every day to get to craft fairs. Barbara ought not to think that she is required to work late into her weekend evenings to complete the accounts. Cheryl need not explore the darkest depths of her psyche for the art's subject matter. To do so is unreasonable, because the organisation's decision-making structure has determined that these are not acceptable reasons for its members. Other such limits include the organisation's joint goal. Barbara ought not to buy 30,000 novelty pinatas for Alan to sell, as this does not serve their joint goal of selling Cheryl's art. In larger scale organisations, these may often be put into "constitutional" documents, like documents which outline the broad principles on which the organisation operates. When an organisation fails to set such principles then it leaves it open to a member to individually determine what she ought to do.

8.3 Limits of occupational role requirements

There are limits to what our occupational roles can require of us. The force of joint obligations still lies in the fact that they are reliance-related obligations, and like all such obligations they are non-moral, practical obligations. Reliance-related obligations give exclusionary reasons, but they are only sufficient and not necessarily conclusive reasons (Raz, 2002, Ch.1, 1.2). That is, they may not be what we ought to do all things considered. They can be overridden by other reasons. If the art-selling organisation is corrupt, Barbara may have a role-based reason to falsify the accounts. If falsifying the accounts is morally impermissible then she ought not to act on the role-based reason.

It follows that a member ought not to determine an underdetermined aspect of their role in ways that violate these stronger reasons. If they ought not to act on these obligations if they require that they act in immoral or socially unacceptable ways, then they ought not to infer a reason from that obligation which demands that they act in these ways. For example, Alan may notice that Cheryl's art is not selling as well as the art of the person on the stall next to theirs. He may decide to steal the other stall's art and sell it as an interpretation of the joint obligation to sell art. Stealing is immoral, and so Alan ought not to do it and ought not to reason that he should, if what he ought to do is left unclear.

If the offer of reliability did not explicitly offer a responsiveness to immoral reasons – for example – then this is beyond the scope of what the member is obligated to do and so the organisation may not require such behaviour of them.

8.4 Immoral organisations

It may not always be unreasonable to infer such an immoral requirement (though it will always be unreasonable to act upon it, as it misunderstands the respective strength of the demands). As with standard reliance-related obligations, joint obligations can give rise to immoral requirements. A decision-making structure may adopt immoral principles. The Mafia considers intimidation an acceptable mode of action, and this is incorporated into its decision-making structures. Certain members of the mafia will therefore sometimes have it as a requirement of their role that they intimidate someone.

Reid Blackman has suggested that such a result shows an account of role-based reasons to be wrong. Blackman's immorality challenge runs something like this:

- (1) If my account of role-based reasons is true, then all roles (including roles like concentration camp guard and Ponzi-scheme runner) have normative reasons to act for their role-based reasons.
- (2) Occupants of such roles do not have normative reasons to act for their role-based reasons.
- (3) So, the given account is wrong.

I disagree with (2). I think it is important to note that there are reasons here. If a mafiosa recognises a role-based reason to intimidate her local shopkeeper, she ought to have recognised that her moral reason to not intimidate others is stronger. It does though seem to be a case where she has had a genuine reason to do so. That at least is the description of the situation I think most accurate. I think that capturing this is important if we are to explain how immoral requirements come about, why we take them to have force for us, and how we ought to avoid them.²¹

9.0 Conclusion

In this chapter, I have put forwards an account of organisations as consisting of individuals bound together by a joint reliance-related obligation, which has the general form of an offer of reliability to respond appropriately to a decision-making structure, so that we can achieve our joint goal, which leads to appropriate mutual inter-reliance. This gives us a general reason to be responsive to an organisation's decision-making structure: consisting of role-descriptions and decision-making procedures. This is what it means to be a member of an organisation. This allows us to largely determine the specific requirements of our organisational roles: the things that we usually think of as jobs.

²¹ For the immorality challenge see Blackman (2023, pp.170-1). I sympathise with Gilbert (2008, pp.9-10) and Gilbert, (2008, p.129), that immoral role-based reasons are nonetheless genuine reasons. Ultimately, the disagreement may rest on settling whether there can be normative reasons that contradict moral reasons. That goes beyond the ambitions of this thesis. Not much really rides on it here. The practical upshots are identical; only the nature of the mistake in practical reasoning which has occurred differs.

With this model in place, we can now turn to the required capacities for trustworthiness and see if organisations are the kinds of entities which can have these capacities. I will first deal with the issue of control, as I argue in the next chapter that we ought to understand organisations as a form of collective agent. Then in Chapter 10, I will turn to the other capacities required: an awareness of the external effects of an entity's actions, and an ability to appropriately recognise and reason about its reliance-related obligations.

9 | Organisations as Agents

The aim of this chapter is to show that organisations can be agents, because they can act intentionally. §1 sets out how this model of organisations can be understood to give rise to intentions when decision-making structures issue outputs directing members to complete specific acts. §2 shows how the nature of organisations puts pressures on them to rationally shape their agency.

*

To be trustworthy, I said in Chapter 7, an entity must be able to control its behaviour in line with those reasons it has. This allows it to fulfil its reliance-related obligations, and to control the way that it offers its reliability to others. So, the question is whether an organisation – understood as individuals bound by a joint obligation in the ways just described – can have this kind of control. If they do not, then organisations do not have a hope of being trustworthy, and so neither do businesses.

I take control to be able to deliberate about what to do, form an intention on the basis of the deliberation, and act on the basis of that intention. Establishing whether an entity has control then requires us to work out if it is an agent – at least if we understand agency in line with a standard conception on which it is to be able to act, where an action is an event with the correct sort of causal chain: it is one that comes about from an intention.¹ I take intention to be a state that settles a practical matter for the intender of what she will do, and which brings about that she does whatever it is that she will do. On this view, an agent is any entity that can intentionally act.

¹ The most standard view (Anscombe, 2000; Davidson, 2006, Ch.1 'Actions, Reasons and Causes (1963)', pp.23-36). For an overview see Schlosser (2019). To be clear, I am not making a general claim that this is the nature of all collective agency or intentionality, but only that this is how organisations can be collective agents. I am not trying to defend this view against potential alternatives, to which I am largely open as potential explanations for these phenomena. The collective agency/intentionality literature has grown hugely (some influential works include French, 1979, 1982; Gilbert, 1990, 1992, 2000, 2008; Searle, 1990, 1995; Miller, 1995; Velleman, 1997; Tuomela, 2005, 2006; List and Pettit, 2013; Rovane, 2014; Collins, 2017).

This chapter will argue that organisation can be collective agents. It will further argue that they face pressures to behave in rational ways in order to achieve their goals.

1.0 Organisational agents

On Margaret Gilbert's picture of group intentions, a group intends so long as its members have a joint commitment to intend. That is, a group of individuals may agree to go for a walk together, and in doing so they form a joint commitment to intend to go for a walk together.

My picture is similar to that of Gilbert's. The forming of a joint obligation involves the adoption of a joint goal or joint goals. The members of an organisation are then obligated to take the pursuit of that joint goal as settled for them, and in this sense may be understood as being obligated to intend that joint goal. When Alan, Barbara and Cheryl agree to sell artwork, their obligation to each other means that they ought to intend to sell their artwork.

The connection on my picture between a joint obligation and an organisational intention is not as direct as that of Gilbert. For Gilbert, to intend as a group that intention must be part of the joint commitment. This has led some, like David Velleman, to suggest that Gilbert's account of group intentions is redundant (Velleman, 1997). Velleman's critique is that the joint commitment is what does the work here. If people jointly commit to going for a walk together, it is that joint commitment that results in their doing so. It is unclear, according to Velleman, what further role the idea of group intention plays. We can explain everything that happens without reference to a group intention.²

This is not the case with my account, as the following discussion will show. The joint obligation is a general obligation to reliably respond to the decision-making

² It may be that this is not a concern for Gilbert, but that talk of group intentions is a way of standardly presenting how action comes about, without the technical terms that are characteristic of her account. She does view joint commitments as 'commitments of the will' and has (at least at certain points in the presentation of her views) described them as being essentially intention-like. On this interpretation, there would then be no difference between the two presentations; except that the latter is couched in the more standard language of action theory. A joint commitment is a group intention. So, Velleman's critique is correct, but perhaps that is what Gilbert intended.

structure to achieve some joint goal. Specific practical matters are therefore not settled by the joint obligation. All that the joint obligation settles is the method by which further practical matters will be settled. Alan, Barbara and Cheryl have a practical obligation to respond to the way their roles and discussions settle what specific tasks they must do, but what these specific tasks are is not derivable from their general obligation, and it would mischaracterise the general obligation to describe it as a joint obligation to intend to do the specific task. That is not the content of the joint obligation: that is simply to reliably respond to the decision-making structure to achieve some joint goal.

On my account, the notion of an organisational intention has a significant role in explaining what happens; and it helps us to understand how an organisation is a collective agent. To see why, we need to think about how intentions come about. An intention is normally not taken to be identical to acting for a reason: we can form intentions that have no reasoning behind them. Intentions are, nevertheless, standardly the outcome of deliberation and the product of decisions. They are the end point of practical reasoning.

In this section, I will show how organisations can undertake their own deliberations and reach their own decisions; and that these can result in specific events happening. This means that organisations can have their own intentions.

1.1 Organisational deliberation

We can understand deliberation to rest on the ability to have some motivational states, which push us to do certain things, and some representational states, which interact with those motivational states to determine what to do and which should respond in some appropriate way to evidence. In individual human agents, these are standardly taken to be desires and beliefs.

Decision-making structures mean that deliberation and decisions about what to do happen at a group-level independently of the personal deliberation and decisions that each member may have. They mean that the motivations and beliefs of the organisation may come apart from those of its members *qua* individuals. This section argues for this claim.

1.2 Organisational beliefs

We can understand a belief in a minimally functional way as those representational states which interact with those motivational states to determine what to do, and which should respond in some appropriate way to evidence.³

Something will be properly considered an organisation belief if it is part of its decision-making structure or can be reasonably derived by members from it, in line with how other elements of the decision-making structure may be reasonably derived from it (discussed in Chapter 8, §8) and fulfils that function.

Organisational beliefs will come about when information – used as a general term to mean anything that represents how the world is meant to be – is incorporated into the decision-making structure (and what can be reasonably derived from this incorporated information). What is it though to incorporate information into the decision-making structure? Incorporation happens when information is either taken up as part of a role or included in deliberations required by the organisation's decision-making procedure. Let me illustrate what this requires with some examples.

Alan, Barbara and Cheryl might each wonder what artwork sells worst. Alan might note which artworks are selling out of his personal curiosity. Alan may then,

³ I aim to show that organisations can have *functional equivalents* of some key features including beliefs and awareness in this and the next chapter. That is, that organisations can have features that do the same thing as features of individual human agents, even though these features may be very different in their nature. The rationale for this approach is that organisations are fundamentally different in their ontology and structure to individual human agents, and so we ought not to expect them to realise the features which individual human agents have in identical ways. See also the approaches of List and Pettit (2013) and especially Gilbert (2000, Ch.7).

For example, the approach to belief that this chapter will take will understand belief through a *minimal functional definition*. In the case of belief: a representational state about some proposition. We will then be interested in establishing whether some feature of an organisation can satisfy this minimal functional definition. There is nothing essential to belief on this understanding of it that requires beliefs to be mental states. That is fortunate, such an approach would likely rule out any possibility of belief for organisations.

Functionalism has a long and influential history in our philosophical theorising about the mind and action. I draw on this approach. It may dissatisfy some. The essentialisation of features of agency to their realisation in human agents should, however, bear the burden of argument. Unless we are to beg the question that agency is somehow restricted to human and human-like organisms, then the assumption appears difficult to justify.

appropriately, form the personal belief that Cheryl's daring photorealist paintings of roadkill sell worst. That will not however count as an organisational belief; it does not fulfil the function we require of beliefs.

Eventually, Alan, Barbara and Cheryl may agree that the question requires answering. Alan already has an answer to it, but if he keeps this to himself it still does not count as an organisational belief. This is because it has not served the function we require of beliefs for the organisation. It has played no roles for Alan as a member.

If the outcome of that meeting is that Alan must determine what the lowest-selling artwork is and aggressively sell it, then Alan's belief can become an organisational belief. This is because it has become incorporated into the decision-making structure. It is now some information which Alan's role requires him to have and to respond to it. It has now been taken up as part of Alan's role.

Organisational beliefs need not, however, be personal beliefs involved in roles.⁴ The outcome of Alan, Barbara and Cheryl's discussion may be that they agree that each piece will be logged by its content and Barbara's role will be to calculate the worst selling artwork each month. Barbara may then report the outcome of those calculations at the next meeting. That it has been raised means that it has been incorporated into the decision-making structure, but information raised within an organisation will only count as an organisational belief if it serves the function outlined above. This will happen if, for example, the role-requirements within the art-selling organisation are changed to require them to be responsive to the information provided by these calculations. Alan may have to sell the lowest-selling artworks aggressively and Cheryl to produce less of these artworks.⁵

These ought to be taken as organisational beliefs because they do not require that any of Alan, Barbara or Cheryl have a *personal* belief that aligns with the organisational belief. Say that Barbara calculates that rural sunsets sell worst. She may doubt the accuracy of the coding of the pictures and so the accuracy of the

⁴ They could be in some forms of decision-making structures like dictatorships.

⁵ Note that incorporated information reinforcing a particular decision-making structure or way that things are done will still allow it to fulfil the function of a belief for the organisation.

calculation enough that she rationally does not form her own belief; Alan may retain his personal belief that the roadkill paintings are the worst sellers; Cheryl may simply not have any beliefs about how her artworks sell. The incorporation of that information and its function – and the fact that each of the members are required to treat it as a belief *qua* role-occupier because of their joint obligation – make it an organisational belief.

Information that is incorporated into decision-making structures but does not serve the function of beliefs will not count as an organisational belief. It may still be relevant to us. We may, for example, be able to criticise an organisation for failing to incorporate some information that was raised within it as a belief or for having a decision-making structure which impairs the incorporation of some information as an organisational belief. For example, if Alan, Barbara and Cheryl regularly discussed how well various types of artworks sold at their meetings but never responded to this within their roles. Doing so may be irrational and, in some cases, immoral. They are not though organisational beliefs because they fail to function as such. They seem to have a more similar function for the organisation to something like perceptions – or memories of perceptions – that individual people have and may fail to form beliefs about appropriately.

The above is an example of a belief that is incorporated into the organisation's decision-making structure. We should also consider as organisational beliefs information which plays a belief-like function and can be reasonably derived from other organisational beliefs (within the confines of reasonably deriving role-requirements, which I outlined in the previous chapter). For example, we can go back to the start of the above example and imagine that Alan's role requires him to aggressively sell the low-selling artworks. The decision-making structure may leave it unanswered as to what these are or how that is to be determined. Alan may reasonably form his own belief about which pieces are low selling. As with other reasonable derivations of what our role requires, these are part of the decision-making structure and so can correctly be attributed to the organisation.

1.3 Organisational values

Being reliably responsive to a decision-making structure can also require acting in line with values which individual members do not personally hold, when one is acting within one's role. Consider the following:

Teacher: Diane is a teacher. Her role requires that she teaches her pupils that knowledge is valuable for its own sake because that is the policy of the school in which she teaches. Diane thinks knowledge is only valuable if it is directly applicable and valued in the current job market.

To uphold her part in the joint obligation Diane should teach her students that knowledge is valuable for its own sake. To do so, let's stipulate that her personal (i.e., non-role-related) motivating reason is a belief that her job requires her to teach this and, say, a desire for the financial reward which accompanies doing her job. That is why she fulfils her obligation.⁶

This, however, seems to leave something unexplained when considering why it is that Diane is teaching that knowledge is valuable for its own sake. We cannot necessarily explain this by appealing to the personal motivating reasons of the rest of the school's members. It may be that across the board they personally do not think that knowledge matters for its own sake.

That action seems to require some further justificatory reasons, and these must appeal to something about the organisation. In this case, the school's policy that knowledge is valuable for its own sake.

That policy has a larger role. Within the decision-making structure it plays a similar role to a motivating, desire-like force in individual reasoning. It is, in a sense, what the organisation wants. It will push the organisation to make choices that further the policy and help to achieve it, so long as members are responsive to it in their roles and organisational decision-making. It will do so even if – as the example has just illustrated – it does not actually motivate any of its members in their individual reasoning.

⁶ To be clear, acting because Diane recognises her joint obligation and values upholding them would also usually be to act on a personal reason.

The joint goal for which the organisation exists and its policies, missions, values and related organisational features, which are appropriately incorporated into a decision-making structure and serve this motivating function, are appropriately attributed to the organisation. They functionally motivate the organisation. For ease, I will call them ‘organisational values’. These can again be implicit or explicit within the decision-making structure or appropriately derived from it.

1.4 Organisational intentions

There are features of organisations which function as beliefs and desires act for individual agents. Organisational beliefs are information which is incorporated into the decision-making procedure in a way that ought to be responsive to evidence and which plays a role in determining what the organisation does in conjunction with its motivations. Organisational values are statements of goals, policies, etc., which push the organisation to do certain things through its members.

Together organisational values and organisational beliefs mean, I think, that organisations can act intentionally and ought to be classed as agents. Organisational deliberation on its values and beliefs can settle what an organisation should do and bring it about that it is done, so the outcome of these deliberations function as organisational intentions.⁷

We saw this in *Teacher*. Diane’s personal motivating reason is a belief that her job requires her to teach that knowledge is valuable for its own sake and a desire for its pay. She has therefore formed the intention to do so. There is also a justifying reason, which requires appeal to organisational beliefs and values (e.g., a policy that knowledge is valuable for its own sake and a belief that the best way to act in line with this policy is to directly teach that to pupils).

⁷ There are a couple of ways of understanding what is happening here. One is to understand what is happening as a case of ‘programme control’ (List and Pettit, 2013, pp.161-2). List & Pettit’s example is to consider a flask containing boiling water, which smashes. We have two sets of events: a high-level one (the water boiling) and a lower-level one (the specific high-energy water molecule) that can explain the same event (the smashing of the vase). We can see how a higher-level event (water boiling) “programs” for the collapse of the flask, and the lower-level event “implements” that program by actually producing the break’. This is possible so long as (i) the higher-level event is multiply realisable; (ii) it is realisable generally only through events of the lower-level type; and (iii) it is actually realised by a lower-level event. These conditions hold in the case of an organisation, and so the decision-making structure can be understood as programming for specific actions which individual role-holders implement.

We can also appropriately talk of the organisation having acted. *It* has taught that knowledge is valuable for its own sake because it made a decision to do so – enacted through the role-requirements to teach this – on the basis of considerations which are appropriately taken to be its own, rather than those of any of its members. The creation of those role-requirements was an outcome of deliberation, which settled what was to be done and brought it about that ‘teaching that knowledge is valuable for its own sake’ has been done. They have therefore served as an organisational intention.

As with its beliefs and values, organisational intentions will sometimes be attributable to explicit outcomes of the decision-making structure (e.g., specific role-requirements). Alternatively, they may have to be appropriately derived from it. For example, when they are left underdetermined. So long as they are reasonably derived role-requirements that settle and bring it about that something is done, they can appropriately be considered organisational intentions.

Examples where personal reasoning and organisational reasoning misalign provide the clearest illustration of why we ought to consider organisations to be agents. This need not be the case. One way of being reliably responsive to a decision-making structure is to “switch” into deliberating in line with it. That is, to adopt a form of something like “team reasoning” (e.g., Gold and Sugden, 2007). This would be to respond to organisational reasons as motivating reasons for you *qua* member. Diane may, for example, accept that *as a teacher at this school* she does believe that knowledge is valuable for its own sake and wish to see that value promoted to her pupils, though she would be dismissive of this outside of her work. Such switches are, I think, possible, though difficult to maintain particularly where they misalign as they do in Diane’s case.⁸ My account does not require team reasoning but is open to it having a role in enabling organisational agency.

⁸ This is to require individual members to “value switch”. That is, to adopt some deliberative perspective separate from their own personal deliberative perspective. This would require something like the team reasoning model to be correct. It is plausible that we can do this. For example, they may be a form of ‘layered agency’ (Nguyen, 2020). Gilbert has also argued for joint commitments as offering a theoretical framework for team reasoning (Gilbert, 2005) and others have defended a combination of joint action and the ‘framing’ of team reasoning

Organisations can properly be thought to deliberate and to act on the basis of this deliberation. They do so in a way that makes it appropriate to consider them organisational agents. Organisations' actions will be those that we can appropriately reconstruct as having happened as a result of organisational agency.⁹

2.0 Pressures to be rational

It may be that the above explanation of organisational agency misses something. The objection here is that to be an agent is more than simply to act intentionally; it is for there to be some coherence to this intentional action. To be an agent is to be rationally coherent, at least to some minimal degree.¹⁰

I do not want to get mired in discussions of agency. I think that I can show that an organisation based on a joint obligation would face some pressures towards being more practically rational. In practice, I think that organisations are likely to require good decision-making structure which ensures some norms of practical rationality are followed. They certainly can, and ultimately that is all that matters for their ability to be trustworthy. An organisational agent without diachronic, rational consistency to its actions will struggle to be robustly trustworthy.

Achieving a joint goal is the aim of a joint obligation. Recall that joint obligations have the general form that a party offers her reliability in responding to a decision-making structure in order to achieve a joint goal.

The joint goal gives the organisation an end, something for which it aims. The broadest boundaries of what individuals can be obligated to do are set by this joint

(Pacherie, 2011). My account is neutral on the role that team reasoning plays, though it seems likely to enable a high-level of responsiveness to the demands of joint obligations.

⁹ Again, to be clear. What it is to appropriately reconstruct will require considering what is reasonable internally to derive within the organisation's decision-making structure, as outlined in the last chapter. It may be that we think that an organisation ought to have been structured differently so that it could have had certain beliefs or that – had it been reasonable – it would have had these beliefs, but that is not part of the view that I am advancing here. We are, I think, right to scrutinise, re-organise and – sometimes – criticise organisations for failing in these ways, but they are unhelpful in working out what an organisation *did* believe, value, etc.

¹⁰ Cf. Rovane (2004). I do not intend to argue against this view; as I say, I do not intend to settle the nature of agency here. I show that it is possible on this model, and indeed likely to come about. The level of reasoning required to be robustly trustworthy will require this.

goal, because – as we saw – it sets the limit of what the parties to it are offering to be reliable to a decision-making structure to achieve.

The presence of a joint goal which must be accomplished produces pressures on organisational agent to be practically rational in similar ways to an individual agent.

If will, for example, face pressures towards coherence. If the organisation adopts values which contradict or undermine the joint goal, then it may not be able to act on the basis of these contradictory values, because – if a reliable responsiveness to them does not contribute to the joint goal – the members will not be obligated to be responsive in this way. Likewise, there will be some direct pressure to avoid poor instrumental reasoning. If an organisation directs its members to pursue a means that clearly does not achieve the joint goal – say by undermining it – then its members will not be obligated to carry out those actions, because they never offered that kind of reliability (they only offered reliability in pursuit of the joint goal) and so cannot be obligated to reliably respond to the decision-making structure in this way. These will produce pressure to improve the organisation's practical reasoning.

For example, Alan, Barbara and Cheryl may realise that they are selling no artworks, because they have agreed that Cheryl should make her highly provocative, politicised paintings and that these should be sold at local craft fairs. The joint goal of selling art will then force the organisation to adapt its activities to reconcile the goal and its means. The three of them will have to either modify the artwork made, where they sell it, or their joint goal (if they can).

These rule out at least some major ways in which an organisational agent may be irrational and show how pressure internal to an organisational agent may push it towards adopting a better decision-making structure. It may, however, still be possible that within those boundaries an organisation can do things which are poor – but not directly contradictory – means towards a certain end or decide to constantly change what it is trying to do (if, for example, it has the joint goal simply that some group of people do *something* together) in a way that makes its actions incoherent.

Ultimately, an organisation will face pressure to achieve the joint goal for which it exists. Achieving such a goal will best be achieved by adopting a good decision-

making structure, where a good organisational decision-making structure will share the features of good individual practical reasoning which helps us to achieve our ends. This will be particularly important when an organisation wants to exist over a longer period of time. An organisation can adopt a good decision-making structure. It can do so, for example, by adopting further constitutional principles alongside its joint goal which limit what can be required of role-occupiers and how decision-making procedures ought to be approached so that there is greater consistency across the organisation.

It may also institute mechanisms which help to ensure greater rational coherence, such as a hierarchical structure where lower-level decisions are reviewed to ensure that they cohere with the organisation's overall approach (such as managerial reviews of teams operating under them, which are themselves subject to review, etc.).

The ability to produce this more rationally coherent agency is necessary for robust trustworthiness. Without it we cannot guarantee that organisational actions will conform, consistently to those reliance-responsive obligations that the organisation acquires. This is because the organisation could – for example – simply stop appropriately valuing these obligations or behave erratically. Regardless of whether they face pressure to act more rationally anyway, a trustworthy organisation will need to be an agent of this more demanding kind. They can be.

3.0 Conclusion

The model of organisations developed in the last chapter allows for organisations to be agents, capable of intentional action because their deliberations are independent of those of their members, settle practical matters, and bring it about that some specific action happens. They can also form independent beliefs about the world and about their internal composition, which again can differ from those held by their members personally. Finally, although there is no guarantee that an organisational agent must be rationally coherent, they are constituted such that they are well positioned to be rationally coherent, and they will face rational pressure to become so in order to achieve their goals.

Organisations represent and reason about the world in ways that can separate from their members. Their reasoning leads to actions happening, which their intentions bring about. They are an important set of agents in our practical world.

Overall, this gives us a clear picture of what organisations are and how they can come to act. We can now return to the question of whether they can be trustworthy. In the next chapter, we will see how an organisational agent can meet the required capacities for trustworthiness.

10 | Are Organisations Capable of Being Trustworthy?

The aim of this chapter is to argue that an organisation can complete the three capacities required for trustworthiness which were explained in Chapter 7. §1 links the previous chapter's account of agency to an account of control. §2 argues that organisations are capable of adopting or being given their own values about different issues, and so they are capable of appropriately considering their reliance-related obligations. It also argues that this can be ensured consistently, through mechanisms including the adoption of "constitutional principles". §3 argues that organisations can have a functional equivalent to awareness and attention, which allows them to be aware of the external effects of their actions.

*

In Chapter 7, I argued that to have the foundations for trustworthiness an organisation needed to have the following capacities:

1. Awareness of the external effects of its actions,
2. The capacity to appropriately recognise and appropriately deliberate about reliance-related obligations consistently, and
3. The capacity to control its behaviour in line with these reasons.

We have now seen how an organisation can be understood as an entity which exists because a group of individuals offer their reliability in a specific way: their reliability to pursue a joint goal by responding to a decision-making structure. In the last chapter, we saw why we ought to consider organisations as collective agents. In this chapter, I will argue that they are collective agents who are capable of these capacities which provide the foundations for trustworthiness. In the next chapter, I will describe what a robustly trustworthy organisation would look like and consider a case study of an untrustworthy organisations which my account illuminates.

I will approach these three capacities one at a time. We will first consider control; then awareness and finally the ability to appropriately recognise and respond to reliance-related obligations.

1.0 Organisational control

By control, I mean the ability to control one's behaviour in line with reasons. Control matters for trustworthiness for two reasons. First, it allows us to fulfil our reliance-related obligations. You are not trustworthy if your friend tells you a secret and you recognise that you offered to reliably keep schtum about it and appropriately ignored your mutual friend's curiosity, but then you told the mutual friend anyway because you carelessly say some of your thoughts aloud. Being unable to control our actions also makes us unable to be robustly competent. Second, control allows us to ensure that we appropriately offer our reliability.

In individual agents, control is standardly exercised by deliberating about reasons, deciding as we (subjectively) think we ought to decide on the basis of those reasons, and then forming an intention which brings about an action in line with that decision. (Note that this allows for, e.g., impulsive action as uncontrolled action because such action misses the deliberative step in which we reach a decision about what we ought to do.) The discussion of organisations in the previous two chapters has demonstrated that organisations can have control. They can recognise reasons, deliberate about those reasons, and act as the outcome of their deliberations. My account has also provided a mechanism by which organisations can – often – ensure such actions are done: by reminding members of their joint obligation to act as their role requires. Generally, this will give rise to a sufficient reason for an individual to act in line with the requirements of her role and so for the organisation to bring about the action which it wants to bring about.

Organisations can implement other means of control. They can implement systems which help to ensure responsiveness in an agent by appealing to those members' personal reasons. This allows them to enact their intentions through alienated members' actions. This form of control standardly uses incentivisation: e.g., pay and the threat of firing (though this is not the only such means, the Mafia for example may use extortion or threats of violence to ensure compliance of alienated members).¹ So, organisations are capable of control.

¹ The Mafia may also use such methods to ensure compliance for members who have reasons that cancel the joint obligation; for example, because they have (rightly) noted that their moral

2.0 Organisational reasoning about reliance-related obligations

To appropriately reason about our reliance-related obligations is (i) to recognise them as *pro tanto* exclusionary reasons, and (ii) to ensure that we can appropriately reason about how to reasonably exercise due care (that is, understand how we are socially situated, to understand the importance of our offer of reliability, and the social and moral norms which may constrain what reliability we should or should not offer).

I discussed in the last two chapters how organisations can have their own reasons and deliberate independently. In this chapter, I want to say a little more about how organisations can come to have different values to their members. This can affect their ability to appropriately reason about their reliance-related obligations.

I will argue that organisations can find how they conceive of different aspects of the world to have different importance in their decision-making either because (i) they adopt some value directly, or (ii) the nature of the organisation's activity means that there are some related values which it will commonly find important in pursuing that activity. The different importance is the amount to which some reason – or belief or perception – appears to matter to us in our deliberation. This can allow organisations to ensure that they appropriately reason about reliance-related obligations.

We also need this appropriate reasoning to be maintained over time; I will argue that this can be done by organisations establishing constitutional principles which limit the scope of the reliability others can offer to them and so ensure that they maintain these principles across time.

2.1 Organisational values

To be trustworthy it is not enough that an individual recognises her reliance-related obligations, they need to carry the right kind of weight for her when she decides what to do. This raises several issues, not all of which I can address. We will focus on the issue of how values can affect the importance which certain ways we think about the world – that is, our reasons, beliefs, perceptions – have when we deliberate. This

reasons cancel the practical obligations they have undertaken. This turns such members into alienated members.

is because of its particular relevance to businesses. Businesses face particular economic pressures, which may distort the importance of certain reasons for them, because of their values. I will discuss this further in Chapter 13.

I want to show that there are two major ways that the importance of some beliefs can be set by an organisation. One is to do with the organisation itself directly adopting a value. We saw this in the last chapter. This can happen simply through their incorporation into the decision-making structure, either by placing consideration of them as requirements of roles or by instituting constitutional policies. Such value setting is an important aspect of modern organisational design. Missions, visions and statements of organisational values abound. In our simple art selling organisation, we can imagine that the members implicitly accept some values such as legality and having fun when they initially jointly obligate themselves. In large-scale organisations, it will be harder to ensure this and so the organisation will be better served by making such values explicit and thereby obligating a member to accept them when they offer their reliability in accepting a role. We saw how values can be adopted in the last chapter.

Organisational values will affect reasoning because they will affect how important and/or relevant certain considerations seem for the organisation. The art-selling organisation that values fun will have reasons to consider how enjoyable some process is, in a way that another organisation will not. In this way, the organisation's values alter how important considerations about enjoyability (or whatever else) are for the organisation.

While organisational values can be adopted, I also think that organisations will often come to have a particular value because it is salient to the activity for which they exist. As I have argued, an organisation exists to *do* something. It exists because multiple individuals have undertaken to behave in particular ways in order to achieve some joint goal (which, as noted, can simply include the acting together.) The pursuit of this activity will determine the salience of particular values for the organisation. As organisations will often be focused on a particular range of activities, they are likely to have a far smaller range of salient values than individuals.

So, for example, an organisation that exists to treat the sick may find that the value of hygiene is particularly salient for it because of its joint goal. It may find itself often making decisions which reflect this value, though it may not have ever explicitly adopted it. This will mean that particular aspects of its deliberation will be important for it: such as forming accurate beliefs on how to practice good hygiene or ensuring that it can notice threats to good hygiene – such as individuals being sick on its premises.

We can distinguish values that an organisation adopts from those that arise from its sphere of activity. To distinguish them I will use the terms *adopted* and *given values*. The latter are grounded in the organisation's sphere of activity. In business, economic reasons will be highly salient and so it may well seem that their associated value is too. (I will object to the need to make this move in Chapter 13.)

There is no reason that I can think of why organisations would not be able to appropriately value reliance-related obligations. There are no concerns here, as there are for moral obligation, that organisations might not be subject to reliance-related obligations.² Reliance-related obligations are practical commitments; they rely on offering one's reliability and being taken as a source of reasons for others. Organisations are capable of both of these, just as they can otherwise practically commit themselves. They can adopt values, and so – if they wish to be trustworthy – they ought to assign the appropriate importance to considerations related to the reliance-responsive trustworthiness.

2.2 Long-term reasoning

To be robustly trustworthy an entity needs to be able to appropriately reason about its reliance-related obligations consistently over whatever period of time it exists. This requires some stability to its reasoning over time. Again, this is possible for an organisation. It should adopt principles that ensure this.³

² Arguments to this end are available (e.g., Collins, 2019).

³ To be clear, I am not saying that the only constitutional principle that needs to be adopted is that the organisation will act in line with its appropriately made offers of reliability. There are complementary principles that may support trustworthiness. If it is to be robustly trustworthy though – given the observations previously made about the difficulty of finding other reasons to always act as we have offered to reliably act – then such a principle will likely be necessary.

There are two potential ways to do this. Organisations can integrate key principles into their decision-making structures. This will be less binding on the organisation, as there is greater potential for the organisation to alter these principles. This is because members may still be obligated to respond to changes in the organisation's decision-making structures.

The principle will be more secure diachronically if it is a constitutional principle (as we met at the end of Chapter 8). Constitutional principles can limit changes to the decision-making structure and so what the organisation can do. They do so because they are ways in which the joint obligation – and so what the members should do as members – is limited. This is because they form the boundaries for which the members each offered their reliability in responding to the decision-making structure. We saw this when Alan, Barbara and Cheryl agree to form an organisation but only so long as their work is fun. This is the limit of what they are obligated to do, and they cannot be required to do something that is not fun even if the decision-making structure tells them to do so.

These can also be used to bind new members and so ensure a greater consistency within the organisation's values despite changes in membership. Offers of reliability are made to join a joint obligation, and these can be accepted by the organisation integrating the individual who makes the offer into their mutual inter-reliance. When Alan, Barbara and Cheryl hire Darren, he offers his reliability to them, and they integrate this. They may make it explicit to Darren that he must offer his reliability to respond in line with their organisational value of having fun. When this is known to Darren, it is part of what he has offered to reliably do and so he ought not to do things like suggesting at their next meeting that they each work fifteen-hour days. In doing so, he violates the joint obligation. As this request violates the joint obligation the others have no reasons related to their membership in the organisation to consider it or act on it (if, say, Darren had been appointed their boss). Constitutional principles are a more robust route to integrate values. I can see no reason to think that a similar principle ensuring that the organisation is trustworthy could not be adopted.

3.0 Organisational attention and awareness

We also need to be aware of the external effects of our own actions to be robustly trustworthy, because this is part of what is required to appropriately reason about the demands of trustworthiness. We need to understand ourselves as socially situated to recognise what others may take as offers of reliability and whether our reliability is a good reason for them. Adopting this socially situated view requires being aware of the external effects of our own actions.

What it is to be aware and to pay attention is itself a controversial philosophical topic. Here, I therefore adopt a minimal view of both. Exactly what is required for an individual to be aware is debateable partly because it must accommodate complex findings from neuroscience and psychology. Organisations are structurally dissimilar from us, and so we do not want an account of awareness that bakes in such requirements. I therefore propose to focus on a minimal function that awareness and attention play for us.

Awareness must be more expansive than our attention, but somehow appropriately linked to it. Imagine that you are giving a talk, and two attendees are sat at the back of the room lost in – what you must assume they consider – discrete conversation. We would standardly say that you are aware of this conversation, if it is playing some role in your experience of the situation but is not the focus of your attention.⁴

The minimal definition of awareness that I propose is that someone is aware of something if and only if there is some mechanism internal to that individual by which that individual can be brought to attend to that thing. This fits the basic elements of the situation just described. If you are aware of the chattering, it can be brought about that you pay attention to it, perhaps because they begin to deride your presentation. If this were not the case, it would seem inappropriate to talk of your having been aware of their conversation. The mechanism (whatever it is) must be internal to you. If you could only be brought to pay attention to the conversation once the chair had interrupted you and asked for silence, we would rightly take you

⁴ To be clear, if you are paying attention to something it follows that you are aware of it.

to have been unaware of it. In the rest of this section, I will argue that organisations can pay attention and be aware.

3.2 *Organisational attention*

It will be helpful to first establish an account of organisational attention. Again, this is a task which I could well require a further thesis to fulfil. The account here should be taken as a starting point of an account, which may be more fully developed in future work.

My starting point is the observation that organisations are structurally dissimilar to individual human agents. They are characterised by a decision-making structure. Any account of organisational attention ought to be aware of this difference, and an account of organisational attention need not meet all of the requirements of an account of human attention.

We can think of attention for human agents as basically the ‘selective directedness of our mental lives’ (Mole, 2024). Organisations, however, do not have distinct mental lives like ours. We can instead identify two aspects of our attention and see if organisations can have something functionally similar.

These two aspects we can call *perceptual* and *deliberative attention*. We perceptually attend to something when it is the focus of our senses. I may attend to the mug on my desk visually or by touch. Perceptual attention affects the inputs to our reasoning: it alters the quantity and quality of the information we gain about the object to which we are attending. I can see that someone has not washed the mug that well, which I could not before I paid attention to it. I can feel that it is cold, which was also information I lacked until I attended to it. The second aspect of attention we might call *deliberative attention*. For us as individuals, we deliberately attend to something when it is the focus of our reasoning. Your attention may currently be held by considering this minimal account, or it may have wandered to what you will cook for dinner.

We can minimally consider attention to be the ability of an entity to direct its resources (a) for gathering information about a particular object or objects, or (b) for deliberating about some set of considerations. This captures, in a broad way, our own attention. Both (a) and (b) also need to be successful for this direction to count

as attention. I may be able to exclude focusing on background noise as you talk, but I cannot be said to have paid attention to you if I do not process at some level any of what you say. Equally, to have paid attention to a belief I must have given it some consideration (though attention does not require that I have appropriately reasoned about the belief). Let us see if an organisation can do these.

Organisations are structurally dissimilar to human beings and are characterised by decision-making structures. These include the determination of roles, which in turn determine what “inputs” a member may put into the organisation’s procedures, and so in part how the organisation gathers information or deliberates. These procedures will themselves be characterizable as specific sorts of systems: such as the passing of memoranda to a particular role-occupier or the input of information into an algorithm which spits out data to be considered by a committee.

This points us in the direction to explain organisational attention. Attention is a way of directing resources and for an organisation this requires the use of its decision-making structure to direct individual members to do specific tasks. So, the question of organisational attention will be its ability to gather information and dedicate appropriate deliberative time to a specific object or problem.

Recall our small organisation of three friends who sell art at a craft fair: Alan who runs the stall, Barbara who does the accounts and Catherine who makes the art. We can think within this simple model how the organisation they compose might come to pay attention.

We can see this in their consideration of what their lowest-selling artworks are. Alan, Barbara and Catherine agree that Alan’s role includes fully detailing the content of each artwork that is sold and that Barbara’s role includes totting up which subjects sell better at the end of each month. In doing this, they have restructured the organisation’s resources, and they have done so in a way that constitutes them paying perceptual attention (in the minimal sense above), because by altering the role-requirements they have enabled the organisation to gather further information on the artworks sold. The perceptual process involving the direction of resources towards a specific object or objects which culminates in a potential input to the organisation’s reasoning.

It is a separate issue whether the organisation would then deliberately attend to this information. They may forget to ever bring it up at a future meeting. It may not therefore be considered as part of their deliberation. They would need to put aside some time to consider it. Deliberative attention requires the allocation of resources – say time in a meeting – towards considering specific things as part of the decision-making.

Note that both forms of organisational attention need not be tied to the attention or other mental states of any of the individual members. Alan and Barbara do not by themselves gather sufficient information to have paid attention to which artwork is bestselling. The organisation having a belief may also not amount to it having paid attention to it. Barbara may form a belief that rural sunset scenes sell worse and if it is part of her role to have this information this will class as an organisational belief. If, however, it is never raised in line with the organisation's decision-making structure then the organisation cannot be said to have paid attention to it.

Organisations can pay attention both perceptually and deliberately.

They may face specific problems in doing so. One type of problem arises across both processes: the organisation may be unable to dedicate sufficient resources to allow it to either gather information on an object or to consider inputs into its decision-making procedure. This may happen because the organisation cannot allocate any resources to these ends. Alternatively, an organisation may overstretch its resources. We can imagine Alan's role already placing a high-level of demands on him, which mean that he does not have the time to successfully gather the required details about each artwork sold. This would lead to the organisation failing to pay attention. This is analogous to how we can fail to pay attention to a friend either by attending to the more interesting argument between people on the neighbouring table or by spreading our resources for attention too thinly, by trying to attend to both.

There are also problems related to inputting the information attained into the organisation's reasoning. The first of these arises at the end of the process of perceptual attention. The organisation may fail to collate this information in an appropriate way to allow it to form an input in its reasoning. Our art-making organisation could hold the information on what the content of each sold artwork

was, and which artworks have been sold, but fail to collate this in a way that allows it to form a belief about which sort of artwork has sold best. The process of deliberative attention would then not have anything to which it could attend. It can also fail, as we have seen, because it is simply not resourced (as when an item never again appears on a meeting's agenda).

This picture of organisational attention suggests that it follows a broadly capacity-limitation model (Mole, 2024, §2). The capacity is constituted both by the capacities of its role-occupiers and in its decision-making structure but are likely to appear throughout the organisation rather than at any particular point, though certain organisational structures make a single bottleneck more likely. For example, we can imagine a dictatorially run company where all information must pass through a single chief-executive, who is the sole decision maker for the entire organisation. The company will face significant problems with organisational attention, as it is likely that the chief-executive will fail to appropriately pick up information passed to her or to allocate sufficient resources to considering all the inputs she receives.

None of these problems are though insurmountable. Organisations can have the ability to pay attention because they can direct their resources to gather further information or to consider information more fully about specific objects.

3.3 Organisational awareness

I proposed that we could minimally define awareness as:

Minimal awareness: an entity, E, is aware of X if and only if there is some mechanism internal to E by which E can be brought to attend to X when X is present in E's environment.

We now have a fuller understanding of how an organisation can attend to something. To understand awareness, we need an explanation of the kind of internal mechanism that can bring about the organisation's attention.

What this requires is a way of directing organisational resources so that they are directed towards particular objects. This will be possible where organisations have ways of inputting to their decision-making structure that allow for it to adapt. This requires having a mechanism by which potential objects for attention can be raised. This can be simply satisfied. An organisation can simply have a system in place by

which a subject can be raised to which the organisation ought to direct its resources. For Alan, Barbara and Catherine this may amount to little more than Alan being able to raise the fact that it appears to him that certain artworks are selling more than others, allowing for the organisation to direct its attention in the way described above. More complex organisations may require more complex mechanisms, especially ones that allow it to overcome potential capacity-limitations. For example, a process may be required where individual members can highlight pressing matters allowing the organisation's attention to be redirected promptly and without blockages: internal whistleblowing would be an example of such a mechanism.

A common block on organisational awareness will be restrictions on concerns that can be inputted. For example, if a role is constituted such that it is only allowed to input financial information to the organisation, say through data entry, then the role will be barred from raising other concerns. This again is not an insurmountable problem; such a role ought not to be set up.

A further issue will be the priming effects of organisational values. As discussed in the previous section, organisations can have their own values, which may limit or change organisational awareness. That role that is only meant to input financial information to the organisation may not consider other concerns to be of interest to the organisation. I will discuss this issue in more detail in Chapter 13.

Overall, however, organisations can direct their resources to gather information about or to deliberate on a specific object. They can, then, pay attention. They can also ensure that they have mechanisms which allow them to direct this attention, and so they can be aware. These abilities can be applied to external effects of their actions, and the organisation can – as we saw in the last chapter – form beliefs about this on which it can reason.

4.0 Conclusion

An organisation can meet the three capacities required to be robustly trustworthy. They can control their behaviour in line with their reasoning. They can reason appropriately about reasons, by adopting appropriate values, and they can integrate these into their decision-making structures or joint obligations so that they continue to do so consistently. They can also be aware of and pay attention to the external

effects of their actions, and as we have seen form beliefs about these about which they can then reason. So, it is possible for organisation to be trustworthy. The foundations for robust trustworthiness can be there. In the next chapter, we will build on these foundations and see what a robustly trustworthy organisation would be like.

11 | A Trustworthy Organisation

This chapter lays out what a trustworthy organisation would be like, completing my argument that an organisation can be trustworthy. §1 lays out what is required for reliance-responsive trustworthiness. §2 shows how this can illuminate how an organisation fails to be trustworthy with a case study of London's Metropolitan Police. §3 gives a picture of a trustworthy organisation. §4 argues that we have a form of coordination duty to produce trustworthy organisations. §5 explains why we cannot guarantee trust because of trustworthiness, but that this should not stop us from aiming at trustworthiness and that we will at least succeed in creating the conditions for it.

*

We have seen that a trustworthy organisation is possible (Chs.9 & 10). This chapter lays out what a robustly trustworthy organisation would be like, fleshing out this skeleton sketch.

1.0 Reliance-Responsive Trustworthiness

The kind of trustworthiness in which we are interested is a robust reliance-responsive trustworthiness which requires:

TWY1. *B* is competent to fulfil those reliance-related obligations she produces,

TWY2. *B* is disposed to fulfil her reliance-related obligations, where she lacks further reasons for acting in line with them,

TWY3. *B* ensures that she does not offer reliability where she would not be able to fulfil the resulting obligations (in other words, she adheres to due care).

To meet these conditions an organisation needs to have the capacities to control what it can do (so that it can both be competent and control its offers of reliability), that it can appropriately reason about its reliance-related obligations (so that it can be appropriately motivated by them and figure out which reliance-related obligations to value), and be able to understand the external effects of its actions by viewing itself as socially situated.

In the next section, I illustrate how this account illuminates what goes wrong when an organisation fails to be trustworthy. I then describe a robustly trustworthy organisation.

2.0 A not-trustworthy organisation

A good case study for an organisation that is not being trustworthy is London's Metropolitan Police or the Met. The Met was recently subjected to extensive scrutiny in *The Casey Review* (Casey, 2023) providing detailed information about their internal failings. It is an organisation that ought to be trustworthy, trust being taken as the foundation of 'policing by consent' (Home Office, 2012). As Baroness Casey (Casey, 2023, p.7) says, 'We give the police exceptional powers and we trust them to use them responsibly'. Although, Londoners' trust in the Met fell from 89% of respondents in 2016 to 66% in 2022 (Casey, 2023). It is also a good candidate because the public's chief vulnerability to the police is likely a reliance-related one, and so – following the social vulnerabilities approach – we would require it to be reliance-responsive trustworthy.

I think that the model developed here clarifies some of the problems at the Met. It is though, obviously, a limited application of a theoretical account which cannot and does not aim to capture all of the force's issues. Baroness Casey's findings are deeply concerning, and I do not intend the fact that I do not engage directly with some of them to underplay their importance.

2.1 Offering reliability and understanding competence

Baroness Casey details several ways that the Met has failed to appropriately reason about how it ought to offer its reliability to the public. First, she notes that the Met – as with many British public services – has very limited resources after budget-cuts over the past decade. This means that as an organisation it lacks adequate resources to deliver certain services and it struggles to appropriately reallocate resources when it needs to (Casey, 2023, pp.9-10, Ch.3). These financial constraints have changed what the Met can reliably do.

The Met has not, however, constrained its offers of reliability to align with its reduced abilities. It has continued to publicise and promote strategies to the public saying that it will do things that it cannot because they are not funded appropriately. For example, it has publicised its strategy to reduce violence against women and

girls (VAWG). It has not funded this strategy adequately to deliver it. Baroness Casey explains that, '[t]hose investigating domestic abuse are ... under considerable pressure, with unmanageable caseloads and poor support for victims.' (Casey, 2023, p.15, Ch. 5.5-5.8).

A misalignment of offers of reliability and our understanding of our competencies always leads to a failure to be trustworthy, when it means that offers of reliability are made which cannot be matched. The Met appears to have favoured appearing responsive or worried about appearing inadequate over a realism about what reliability it can offer to others. It has bound itself to be reliable, only to fail to meet its obligations – as it had to because it could not meet them. So, it has been untrustworthy.

2.2 Disconnection inside an organisation

The Met has also become increasingly disconnected in its organisational design, in a way which undermines its ability to know how it has offered its reliability. There is an increasing gulf between decision-makers and teams of police working in the communities the Met serves. It has massively decreased funding to neighbourhood policing teams and engagement with community groups. Decision-makers in its headquarters increasingly disregard the insights and reports of its officers and staff who work closest to the Londoners who rely on the Met.

My account suggests that a disconnect like this would produce two problems for the Met. First, it would make it ever harder for the Met to understand how it is socially situated. That is, it will lack an understanding of itself as it stands in the wider web of reliance. This will lead it to struggle to see how its actions may be reasonably interpreted as offers of reliability by others, and to misunderstand the relative vulnerabilities others face to it. This will produce unexpected reliance-related obligation and lead it to mis-prioritise the obligations that it has. This will undermine its trustworthiness.

This problem can, I think, be seen in the way that the capital's police force has increasingly focused on its own priorities, over those of Londoners. It has been especially concerned with securing sufficient resourcing for prestigious national command centres and high-profile specialist units, in ways which Baroness Casey

suggests has undermined its ability to combat the crime which most affects Londoners (Casey, 2023, pp.10-11, pp.14-15, Ch.3.10-3.11, Ch.4).

This kind of disconnect will also make it harder to align offers of reliability with actions taken. Those officers who engage most frequently with the public are likely to be unaware of what decisions will actually be taken, which may lead them to make offers on behalf of the organisation which are not kept. Decision-makers will be unaware of these and unable to respond to them. The disconnect in the Met severs decision-making from the information needed to appropriately reason about what it should do in a trustworthy way.

2.3 Inappropriately motivated by reliance-related obligations

The Met is an example of an organisation that has failed to appropriately value reliance-related obligations, let alone to establish this as a long-term value for the organisation. An organisation that does so fails to appropriately reason about its reliance-related obligations, and so it cannot be trustworthy.

This is clear in the Met's approach to prioritising its actions. *The Casey Review* highlights how the Met has consistently favoured technocratic solution-focused policy implementation over understanding the needs of those who rely on it to fight crime (Casey, 2023, Ch.10.3).¹ It fails to consider the relative vulnerability of those to whom it offers its reliability. It prefers to operate under the assumption that it knows best what needs to be done.

The Met also fails to recognise that offers of reliability commit it to acting on those offers. Instead, Baroness Casey makes clear that it is common for the Met to offer reliability and then view this as having done enough. This can be seen in its “box-ticking” approach to community engagement. It claims it will listen to the needs of those it serves and in doing so offers to reliably act on those concerns. Instead, it simply arranges meetings which – in part because of the lack of funding and disconnected organisational structures already discussed – go no further.

¹ This obviously brackets off considerations about whether such an approach to evidence-based policy would even work or whether it incorrectly considers what works *there* to be something that would work *here* (Cartwright and Hardie, 2012, esp. Part II).

Overall, the Met inappropriately values its trustworthiness against other values it holds, especially those regarding its own internal expertise and reputation, which outweigh considerations which could make it more trustworthy.

The Met also fails to embed trustworthiness. Its values are highly changeable, as it lurches between its response to the latest “crisis” without fully integrating the values it claims to have into its decision-making structure in a pattern Baroness Casey terms ‘initiative-itis’ (Casey, 2023, Ch.3.11).

2.4 A lack of control

The Met struggles to bring it about that its members and those subject to its organisational decision-making structure are actually under its direction or control. In part, this is because the Met lacks appropriate internal accountability structures to ensure that its officers and civilian employees held to the requirements of their roles and dismissed if they do not meet them (Casey, 2023, Ch.7).

The Met also claims to have constitutional values but fails to ensure that these are its actual values. It lacks proper vetting procedures to ensure that prospective members are offering it reliability in the ways it claims the organisation requires. This allows entry into the organisation (even if as, in my terms, alienated members) those who will abuse their positions or, as Baroness Casey puts it, ‘who want power over their fellow citizens’ (Casey, 2023, p.7). This leads to a lack of organisational control which is deeply damaging to its trustworthiness.

The Met also fails to embed values in its members. It ‘does not make ethical standards as clear as it could, and it has no systems in place to ensure staff and officers adhere to them, nor clear consequences if they do not’ (Casey, 2023, pp.12-13, Chs.3.3-3.4). This leaves significant unclarity for those occupying job roles, making it difficult to appropriately determine how one ought to act within them.

An organisation which cannot control what its members and other role-occupiers do cannot be trustworthy. It cannot ensure that its reliance-related obligations are fulfilled.

2.5 Organisational blindness

Finally, the Met faces a further obstacle to developing the socially situated decision-making and to ensure that it has control over its members sufficient to support

robust trustworthiness. It suffers from organisational blindness. By this, I mean it has an inability to appropriately attend to how those outside of it are vulnerable to it, because its organisational structure prevents it from appropriately considering this.

This blindness can be explained within my model of an organisation. It occurs as a by-product of internal trust. I have argued that organisations involve reliance-related obligations, and so they are likely to be “hot spots” for trust between members who require each other’s reliability in being a member. Organisations will likely function better with such internal trust. As Bernd Lahno (2017) has argued, this can be problematic because of the effects that trust has on the acceptance and interpretation of evidence (something I accept though for different reasons to Lahno, see Ch.6, §4). This can lead to group members becoming “blinded” to evidence of wrongdoing directed at another group member because they trust her.

On my account this undermines the organisation’s trustworthiness when it deadens an organisation’s awareness of others’ vulnerabilities to it. This will mean that it fails to appropriately reason about its obligations, and so it fails to be trustworthy. This blindness can also stop an organisation from realising it is failing to be trustworthy, and so from taking the appropriate steps to rectify this. That will be the case, for example, when a victim of wrongdoing is ignored in favour of a member.

Blindness of some kind has certainly been present at the Met. There are well-documented cases where accusations against officers have been overlooked and their testimony preferred over their victims. The causes of this are complex, but a plausible mechanism is blindness through internal trust. The Met has fostered an internal culture which I think has likely contributed to an inappropriate form and level of trust taking root between its members. This has been encouraged in part by a culture of “specialness”, notably through the internal view that the Met “knows best”. It has been reinforced by ‘a profound culture across the Met that incentives people to look, act and sound the same, and a resistance to difference’ (Casey, 2023, pp.16-17). This has been enforced through widespread bullying and accepted homophobic, misogynistic and racist discrimination directed at the public and other members of staff. This has led to a Met where many people do look, act and sound

the same: 82% of its employees are White and 71% of them male (Casey, 2023, Ch.9). Such a homogenous membership with high levels of “in-group” thinking will likely encourage group identification among its members. I think this will often produce inappropriate forms of in-group trust, producing organisational blindness.

2.6 A non-trustworthy organisation

The Metropolitan Police are, I think, certainly not a trustworthy organisation. They seem to be an untrustworthy one. They demonstrate much of what can go wrong in an organisation in ways which undermine trustworthiness. We can now turn to how an organisation would have to be in order to be trustworthy.

3.0 A trustworthy organisation

Untrustworthy organisations like the Metropolitan Police harm people. Generally, our lives go worse when we have to interact with untrustworthy and non-trustworthy organisations. Things go better when organisations can be trustworthy, even if this does not mean that we have to trust them (see below).

3.1 A trustworthy organisation

In overview, a robustly trustworthy organisation will be realistic about what it can do and make sure that its offers of reliability align with what it can do. It will be appropriately committed to acting on its reliance-related obligations and embed this in its decision-making structure. This will be part of its efforts to ensure that it has the control required to fulfil the demands of its trustworthiness. It will ensure that it can do so not just briefly or intermittently but for a significant span of the time for which it exists. Its ability to appropriately reason about its reliance-responsive obligations will be strengthened if it understands itself and its actions as socially situated. This will allow it to better understand how its actions may reasonably be interpreted as offers of reliability, and how others are vulnerable to it so that it can appropriately consider how important a reliance-responsive obligation will be. It will think about its effects on others reliance on it before it acts. It will be careful about when it offers its reliability to others and understand that it cannot always refuse to do so.

There are many trustworthy organisations which we encounter in our day-to-day lives. Sometimes, we will be interested in only a limited trustworthiness from them. My local coffee shop is trustworthy; but then it has offered only limited reliability to

me and those who live around me: decent coffee at a reasonable price and hangover-numbing Turkish or Full English breakfast options. They consider the limited disposable income of their clientele; they appear to do so not just out of economic concerns, but as an independent concern, which would motivate them significantly even if their own overheads were to face an increase. They have, however, always been clear about the struggles that come with running a small business, and the pressure that this puts them under. It may also be robustly trustworthy, though to know if it is I would need to know about its interactions beyond its customers: how it offers and responds to reliability offered to its suppliers, creditors and others.

Robust trustworthiness becomes harder to satisfy the more expansive the demands of trustworthiness become and the more diverse the range of actions the organisation offers, and so the greater variety of demands for reliability it acquires. A medical research charity that forms a partnership with a commercial enterprise to develop a medical device, for example, will have to be careful that it can maintain its offered reliability to its beneficiaries while sticking to the terms of its new commercial interests. Developing robust trustworthiness will not always be easy and exactly what is required will vary just as organisations are varied. The account I have given means they may range from a couple of people to a sprawling inter-governmental agency. Some will last for mere minutes, some for centuries. The nature of the organisation will affect exactly what is required of it. This picture aims to capture what is common across this wide array of entities.

3.2 Offering reliability and understanding competence

First, an organisation must know how it will be able to allocate its resources and what its limitations are (both those internal to it and from external risks). This allows it to understand what it can do.

A clear understanding of what we can do is the bedrock of trustworthiness. It allows an organisation to limit its offers of reliability to those it can fulfil. A robustly trustworthy organisation will be honest about what it cannot do – no matter how much it wishes things may be otherwise. An organisation that overplays its hand or tries to hide its limitations in ambiguities will offer reliability it cannot achieve. This will obligate it in ways it cannot fulfil. This is more likely to happen when the organisation does not know what it can do.

3.3 *Appropriately responding to its obligations*

A robustly trustworthy organisation will understand what it can do and will appropriately offer its reliability where it sees that non-members are reliant on it. As I have made clear, this requires excluding other reasons, but defeasibly: especially if there are moral considerations. The simplest way for an organisation to realise this is for it to adopt values that assign this sort of importance to its offers of reliability (I will consider some potential problems in doing this, especially for businesses, in the final chapter).

A robustly trustworthy organisation is also going to have to ensure that it can embed the importance of acting in line with its offers of reliability consistently.² It would be untrustworthy to signal that a course of action is “forever” if, for example, it is dependent on the views of the current C.E.O. Here, “constitutional” features that set the scope of the underlying organisational obligation will be helpful. These can bar changes in some features of the organisation, ensuring such behaviours ought to continue over a greater period. (I will consider these using a case study of *Reuters* in the next chapter.) This will only work if the organisation ensures that those who occupy its roles are aware of the importance of these considerations or that the decision-making structure embeds this in each decision (something which the Met fails to do).

To be trustworthy this responsiveness should be internal. Complaints procedures and similar mechanisms will be inappropriate ways to trigger responsiveness. This is because it triggers a responsiveness only after the vulnerable person has been harmed. To be trustworthy is to not harm in the first place. Complaints procedures and other mechanisms by which non-members can highlight problems with organisational behaviour are best understood as a fail-safe: they provide a route to raise untrustworthy behaviour (or other wrongdoing). If such concerns are appropriately considered, they can though serve as ways of building more robust trustworthiness.

² This is of course relative. A Christmas market stall may be able to be robustly trustworthy by signalling that it is not offering its reliability beyond the season in which it is trading, but meeting all of the conditions of robust trustworthiness within that timeframe.

3.4 Knowing its social situation

The robustly trustworthy organisation can face a number of competing demands. As I argued in Chapter 5, it will help the robustly trustworthy agent to fulfil the demands of trustworthiness that it faces, if it considers the priority that it would have to give to the various obligations that it could acquire. This requires a robustly trustworthy organisational agent to think about how vulnerable the individuals who are reliant on it would be to its failure to be reliable. This requires consideration about whether (a) there are other options for reliability, (b) how important its reliability is for others' success, and (c) how important that overall success is for those who are reliant upon it.

The robustly trustworthy organisation would need to understand these to appropriately understand the importance of the reliance-related obligations it has and could acquire. This is best done by developing a socially situated view of itself. As I put this earlier, it must understand its own reliability as part of a web of potential reliance to understand how others are interpreting its actions.

As we have seen, to enable this kind of socially situated view an organisation will need to ensure that (i) appropriate mechanisms are in place to enable such awareness and to gain appropriate information; (ii) appropriate resources are dedicated to this. Appropriate mechanisms will take various forms. "Front-line" members (that is, those who engage with people external to the organisation the most) may be best placed to notice potential and actual reliances, and a trustworthy organisation may institute mechanisms that allow such members to easily feed into decision-making. Alternatively, organisations may commission research or scoping exercises before taking decisions, which focus on how others may become reliant on the organisation. Equally, organisations may engage in external consultations to better understand how their decisions will affect non-members. For these mechanisms to function, the organisation will have to ensure that it has allocated sufficient resources to them to allow them to properly function, and that an organisational culture does not take root which distorts the information it gains (such as one which disregards lower-status or front-line staff in favour of higher-level managers).

The resources required for this will depend on the scale of the organisation and the number of individuals who may be reliant on it. A small-scale local food bank, for example, may be able to gain the required self-understanding simply through its interactions with its users.

Developing a socially situated view also supports robust trustworthiness because it helps you to understand which of your actions others may reasonably take as offers of reliability. This is because what it is reasonable to take as an offer of reliability is heavily socially dependent, because the boundaries of what is reasonable are in part set by social norms. This may cause problems for organisations which operate across different societies and cultures. They are likely to misunderstand how their actions will be taken by others, what reliability they have offered and so what reliance-related obligations they may have. This means that organisations which operate in diverse societies ought to ensure that they understand those societies, and the differing demands that may reasonably arise in them.

3.5 Trustworthy control

A robustly trustworthy organisation will ensure that its members act in line with its reliance-related obligations and control the offers of reliability which they give, so that they align with those obligations that it can fulfil.

For inviting reliance, this will have to take into account the realism about what an organisation can do. A robustly trustworthy organisation will also need to carefully control its signalling or “organisational messaging” to avoid members inviting reliance in areas where the organisation as a whole is not able to be reliable.³

Organisational trustworthiness will be undermined, for example, if a role-occupier signals in ways that can be reasonably interpreted as being part of her role, but where she is speaking personally. An example of this is the meeting at the end of World War II between President Franklin Delano Roosevelt and the King of Saudi Arabia, Ibn Saud, on the Bitter Lake. Among the topics they discussed was the

³ Sometimes members may clearly signal in ways that signal too much, and they may do so in ways that are not authorised by their roles. These are, on my picture, not part of the organisation’s signalling. They therefore do not make the organisation less trustworthy. Nonetheless, they may undermine trust (see §5.1).

ongoing settlement of Jewish people in the lands of Palestine and the future policy of the USA towards the establishment of a Jewish state. Roosevelt undertook to carry out no policy without consultation with both the Arabian and Jewish people. Ibn Saud understood Roosevelt to be acting on behalf of the US government; Roosevelt meant the commitment personally. President Truman abandoned the consultation a matter of months later (Holden, Jones and Johns, 1982, Ch.10). In my view, Ibn Saud reasonably took this as an offer of organisational reliability, and so Roosevelt's actions were damaging to the trustworthiness of the US government. This is because, standardly, a pronouncement on policy by the President of the United States within a meeting with another head of state counts as a signal of the policy of the government of the United States. So, reliability is reasonably taken to have been offered here by the US government. This makes it important that certain role-occupiers take greater steps to clearly distinguish their personal commitments from those of their role; and a trustworthy organisation will need to embed this into the role requirements.

This careful signalling will require significant control over member behaviours, but it is important to note that by this I do not necessarily mean highly restrictive practices like the kind of performance management that occurs when a script is given to a call-centre operative (though such measures may allow organisational trustworthiness). The role of 'President of the United States' could have among its role requirements the need to clearly distinguish personal and governmental pronouncements. That it does not, is a weakness in the role-description, at least from the point of view of trustworthiness (the ambiguity may be a gift to diplomats).

A robustly trustworthy organisation will also ensure that its members realise the behaviours that are required to fulfil the demands that the organisation has to fulfil if it is to be trustworthy. How this is possible has been discussed enough but it should be emphasised that such control is available in different ways. An organisation may be possible that is highly restrictive of its members' behaviour. We can think of this as members having role-requirements which severely limit the options open to a role-occupier to fulfil that role. For example, a call-centre operative stuck to a specific phone script. If the organisation has sufficiently flexible decision-making that it can be responsive to non-members' reliance on it – by for

example, adapting these stringent requirements – then it would be possible for such an organisation to be trustworthy. In fact, in certain ways (such as being able to control signalling) it may be able to be trustworthy more easily than other organisational structures will allow. Such organisational designs may be objectionable from other perspectives, but they can be means to trustworthiness.

So too are organisational structures that allow significant leeway for individual role-occupiers. For example, within a barristers' chambers where individuals may have significant latitude to invite reliance and overall decision-making structures are heavily decentralised. What is needed here is a clear understanding of the chambers' overall values by its members, so that offers are not later revoked and so that members do not act against each other, in ways that stop the organisation from fulfilling its obligations (as would happen were one member of the chambers to offer something which another refuses).

What seems common across these structures, is that the robustly trustworthy organisation has mechanisms by which it can hold an individual accountable for their organisational role. That, and the concomitant ability to realise control, is what appears essential for trustworthiness. This raises a problem for organisations that “outsource” their activities. Where this reduces accountability and control, it is likely to reduce the ability of the organisation to be trustworthy.⁴

3.5.1 Correcting organisational blindness

This necessary kind of accountability can also, as we saw in the Met, be undermined by a kind of organisational blindness which my account suggests is particularly likely to occur. That caused by trust between members.

A robustly trustworthy organisation needs to avoid this. Affected organisations will struggle to root out poor internal behaviour which could undermine organisational trustworthiness.

⁴ The last two chapters are – together – a rebuttal of the arguments in Hawley (2017) regarding trustworthiness in groups and their members (though I note her focus is on testimonial trust). I hope to have shown that the trustworthiness is rightly thought of as the organisation's. I take the latter section to be particularly important. Members need not be taken to be trustworthy as individuals for the organisation to be trustworthy, if the organisation can appropriately control action. For further critical discussion of Hawley and trust in groups see Faulkner (2018), Pouryousefi and Tallant (2023) and Bennett (2024).

I have argued that our trust-concepts pick out various kinds of trust. I mentioned in Chapter 2 that there may be good reasons to better discriminate between them. I think that organisational blindness provides a good motivation for doing just that.

It is common in organisations to try to foster a sense of “identification” or *esprit de corps* which seem likely to promote one form of trust between members: that which is grounded in our group membership. That form of trust is – I think – most likely an affective form of trust, and it is likely to also oddly respond to evidence. This sort of trust may then further weaken the extent to which members accept and respond to reports about behaviour from those external to the organisation. It looks like that is part of the picture in the Met.

That form of trust is not though, on the account I have presented, the important one for the good functioning of the organisation. It is not a form of trust which best tracks reliance-responsive trustworthiness, and it is that form of trustworthiness which we require from fellow members of an organisation who have offered us their reliability to follow the decision-making structure. We can try to limit our trust. In doing so, we may not eliminate these effects – which may be, as Lahno (2017) argues, unavoidable – but we can at least lessen them. Here then is one good reason for being pickier between the forms of trust we seek out in different contexts.

Eliminating such effects may be possible, but they may require eliminating trust. On the picture presented here that may not be fatal, but it is likely to make an organisation less stable. Other members’ responsiveness will be called into question more regularly, in a way that will make the joint obligation difficult to maintain (though it may be possible to maintain confidence in them with sufficient internal mechanisms to ensure reliability).

Organisations can adopt mechanisms which counteract this trust. Examples include strong whistleblowing mechanisms and grievance procedures which are independent of management structures. These are likely to undermine internal trust between members (as all regulation does, on my account), but that is a fair trade-off to ensure external trustworthiness: at least for large, powerful organisations or those who deal with the most vulnerable.

3.6 Transparency and audit

What is notably not included in these requirements is that the organisation is open or transparent, at least not in the senses in which these have often been taken to contribute to trust building (see the discussion of transparency, governance and trust in Hood, 2006). An organisation that is open about its decisions and deliberation may gain some advantages in being trustworthy. This may help it to improve its signalling about when it is offering its reliability. This though seems closer to Onora O'Neill's (2002a, 2002b, 2014 – see also O'Brien, 2014; Nguyen, 2022a) understanding that trustworthiness and trust in organisations requires justification and clear explanation, rather than openness and transparency as they are often practised. Openness and transparency as simply releasing or reporting information about oneself does not clearly contribute to trustworthiness. In fact, in confusing the picture of what the organisation values and how such internal information is used it may damage the signalling done by an organisation, in ways that make it difficult for the organisation to be trustworthy and certainly damage trust. That will be the case if an organisation releases information that is accessible but difficult for individuals to understand (Manson and O'Neill, 2007), where this introduces ambiguities about what the organisation will do.

Furthermore, current systems of openness and transparency will not contribute to greater trustworthiness because they are aspects of external regulation of an organisation. A feature must be self-regulated to be a trustworthy feature. Freedom of Information requirements, for example, undercut trust because they require non-members to actively seek information to help regulate and interpret the potential trustworthiness of the organisation. Robust trustworthiness requires that the trustworthy entity does this itself. Similar observations stand for whether the organisation and its members are externally accountable; again, this is external regulation and so will not contribute towards a trustworthy organisation. Though if those regulations are internalised and embedded as the organisation's values, then this could contribute to trustworthiness.⁵

⁵ I think that avoiding accountability is likely to reduce overall trust, but this is because it suggests a general attempt to avoid being held to account which may be important for the trustor were her trust to fail. Likewise, an individual may be less likely to be trusted if she

4.0 A duty to be trustworthy?

I have argued that there can be robustly trustworthy organisations and described how they would have to be. There are untrustworthy or non-trustworthy organisations. To become trustworthy these organisations need to reform.

This raises a further question, which I have not yet addressed. Why should an organisation be trustworthy?

I have already suggested some answers to this. Things are better when there are more trustworthy agents in the world. An organisation may recognise this and simply wish to be trustworthy. It may equally find that there are advantages that it gains by being so; a trustworthy organisation may well be an organisation individuals act with more often, for example. It may therefore simply decide that it wants to be trustworthy, and if it does, then it should try to emulate what I have just described.

That though is probably not what is meant when we ask why an organisation should be trustworthy. It will, at least, not be what most philosophers mean by that question. We might wonder if there is some stronger reason why an organisation should be trustworthy. Is there even a duty for organisations to be trustworthy, as some have argued (Kirby, Kirton and Crean, 2018)?⁶

Organisations may certainly have a legal duty to become trustworthy. That may happen in instances of trust-focused regulation. Again, however, the question is more likely whether there is a moral duty to become trustworthy.

I note first that on this account trustworthiness is likely to be more self-generating than on many others. It is, as I have argued, common to be in situations where we can offer reliability and often tricky not to do so. This will especially be the case for organisations which want others to engage with them. When an organisation goes

avoids accountability in other areas of her life (though she may, in fact, be trustworthy and be happy to be held accountable for her reliance-related obligations).

⁶ Specifically, corporations. I think that the below captures better how the duties that this paper argues for come about. It requires specific situations in order for trustworthiness to become 'salient' enough for a duty to arise. If organisations have duties, then I think this can be caught by a general moral duty not to harm which is best realised by being trustworthy. I also think that there are duties which fall on the members as individual people, rather than on the organisation as a whole. See also Alfano and Huijts (2020).

out there looking for others to interact with them and likely become reliant on them, it will be hard to avoid offering reliability. Reliance-related obligations do obligate, and so organisations ought to follow them regardless of how committed they are to being trustworthy. If they fail to do so, they will have wronged the person to whom they are obligated. They should then fulfil those obligations. Trustworthiness can at least get going without a broader duty.

I have two issues with arguing for a duty to be trustworthy. First, I am not arguing here for a form of trustworthiness which I think is always morally valuable. It is not clear to me that we can have a moral duty to promote it. So, we do not have a duty to be trustworthy in a direct sense. However, I do think we have a duty to sometimes reform our organisations so that they can be trustworthy. I think that this occurs when not being trustworthy would sufficiently harm some person or people that it would contravene a moral duty to not harm them. This will best be fulfilled by being trustworthy. If an organisation can have a duty, then I think it only indirectly requires trustworthiness.

Second, ascribing a duty to be trustworthy to an organisation is complicated by the question of whether an organisation can have a moral duty. What exactly this requires is not something that I need to settle here. We have discussed how an organisation can fail to be trustworthy as often because an organisation lacks control or has some other defect as because it has aims that do not align with the demands of trustworthiness. This suggests that there will be a significant overlap between organisations with weak internal control (and so weak responsiveness to reasons) and organisations that fail to be trustworthy. Others may be moral delinquents – unable to appropriately morally reason, like psychopaths. This calls into question the organisational duty to be trustworthy approach, because – at least on some prominent accounts (Collins, 2019) – what is in part required for an organisation to have a duty is for it to be able to fulfil a duty, but both a lack of control and an inability to properly morally reason would undermine that ability. That calls into question whether these defective organisations could even be duty bearers. If they are not, then they cannot have a duty to be trustworthy. Given that poorly behaving organisations motivated our concern, establishing a duty to be trustworthy would then not help us with the problem with which we started.

I leave it open, but think that it is unlikely, that there is a duty to be trustworthy for an organisation. Instead, I will borrow an idea from Stephanie Collins: coordination duties. She summarises it as:

if a non-collective group is able to produce a morally valuable outcome in a morally valuable way—or is able (in a morally valuable way) to form a collective that will produce that morally valuable outcome—and if no other group will do so in a better way, then each member of the group has a duty to be responsive to the others with a view to the outcome being produced or to the collective being formed. (Collins, 2019, pp.117-8)

Collins' (2013, 2019, Ch.4) argument for the existence of these duties is complex, and I will not cover the argumentative ground that she does in justifying them. I agree with her that they exist, and I only aim to show that the situation we are concerned with can meet the requirements of the Coordination Principle she lays out (Collins, 2019, pp.116-7).

Coordination duties either require us to be responsive to each other or to collectivise. The former of these requires that we are open to coordinating with each other, without organising. The latter, for our purposes, amounts to organising.

A coordination duty arises when a number of conditions are met, which I will now show are met in the cases we are interested in. We need some state of affairs which is 'morally important'. On my account, while trustworthiness itself will not meet this bar, it does seem some organisations becoming robustly trustworthy organisations would. This state of affairs would be – as I have argued – a world in which many less people would be vulnerable to things going badly for them and so being harmed.

That seems morally valuable. For coordination duties, we then need that there is no group of individuals or organisations which have that duty (because they cannot have such a duty, either because it is beyond their abilities, or they are not the sort of thing that can have a moral duty).

At the same time, however, there is a group constituted by certain individuals which could produce an organisation that would have a duty which would produce that

state of affairs. To be clear, in an organisation that cannot have a duty, this would mean that the organisation could be reorganised so that it can have a duty and fulfil it. In our case, this is the members of a non-trustworthy organisation. It then requires that that trustworthily-organised collective exercising its ability would be expected to contribute to a more morally valuable state of affairs. Which again is satisfied in our case: an organisation which is more trustworthy will lessen the harm caused to people through that which they are reliant on failing. Collins then argues that we need there to be no other group which could better realise this state of affairs. This again is met in our situation. The pre-existing organisation is best placed to alter itself.

Collins argues that when all of this is in place, each individual has a duty to be responsive to others efforts to coordinate so as to bring about the state of affairs (or, if that would only be effective where enough of the others act in similar ways – as in, for example, preventing climate change – that they take action – bearing ‘reasonable costs’ – to make it clear that they would respond to those efforts, should enough also do so).

If I am right, then this gives us a way of promoting organisational trustworthiness even absent organisations that currently have duties or of a more general duty to be trustworthy. I think that such an approach is the right one. What it amounts to, is that each of an organisation’s members has a duty to promote the organisation’s trustworthiness *if* this will contribute to a wider, more morally valuable state of the world. These are individual duties on each member to either work with other members to reform the organisation so that it is trustworthy or to make it clear that they are willing to do so, if enough of the others are also willing. The state of affairs is a lower level of vulnerability, but the best way for that to be achieved is to reorganise in ways that promote trustworthiness.

This would mean that it is on us as members to bring about more trustworthy organisations. What is required of each of us will, in practice, rest on the individual ability we have to reform that organisation. Those with greater power to reorganise the decision-making structure to promote organisational trustworthiness ought to be responsive to each other so that it is reformed or at least show that they would do

so if others joined them. For many organisation members, it may only be open to show willingness to push such a reform with enough fellow members – for example through action like a strike – if we faced significant costs – like unemployment – were we to do more.

There is then, I think, a duty to be promote more trustworthy organisations in many contexts. That duty is not, however, a duty of the organisation. It is its members' duties. Many people will then have a much larger share of the responsibility for promoting more trustworthy organisations than we might standardly think.

5.0 Trust in organisations and organisational trustworthiness

Finally, it is worth making a few observations about the relation between organisational trustworthiness and individual trust in an organisation.

A feature of my account of reliance-responsive trustworthiness is that it can require an agent to behave in the appropriately trustworthy ways even when the could-be-trusting agent does not form a trusting attitude. Reliance-related obligations bind regardless of whether others have a specific attitude towards you. A robustly trustworthy organisation will need to be trustworthy even if no one trusts it, and an absence of trust should not be taken as a sign that an organisation can avoid the demands of trustworthiness.

Trustworthiness does not guarantee trust, and the trustworthy person ought not to expect trust because it is up to each individual whether they choose to rely.⁷ We can only be trustworthy and hope that others will trust us.

There are, however, steps we can take to cultivate trust from others or at least not destroy trust where this trustworthiness would otherwise make it appropriate. Acting as I have described will though make trust fitting and so likely promote it.

Internal actions towards members are not part of organisational trustworthiness (directly – they may, for example, undermine the embedding of value or encourage

⁷ Where we take ourselves to be wronged because we are not trusted this seem to involve other norms which embed trust in specific relationships. For example, it may be wrong to not trust your friend if she is trustworthy; but this seems to have more to do with the fact that she is your friend and you should trust friends, than with the fact that she is just trustworthy. Standard cases of trustworthiness do not seem to be accompanied by a wronging if the trustworthy person is not trusted.

poor behaviour from staff). It may, though, undermine the likelihood that an organisation is trusted. This is not guaranteed, but it seems likely that in assessing an organisation's potential trustworthiness we would consider its behaviour towards those individuals who are reliant on it. As being an organisation is constituted by a specific type of reliance-responsive obligation, we may view organisations that do this poorly as less likely to behave in trustworthy ways towards non-members. Trust may be better encouraged by treating members better.

A further problem in establishing trust is that for an external observer it may not be clear exactly what the organisation is. As I have made clear, organisations are not identical to, say, the legal parameters that bind them. Two legally distinct organisations – like a wholly owned subsidiary and its parent – will form one organisation. The nature and extent of such an organisation may be unclear to those who seek to trust it, but equally I think everyday judgements about what an organisation is will often overlap more with the account given here than with legal structures. I will not enter here into what exactly constitutes an organisation (see e.g. Hess, 2020), but for our purposes, organisations who fail to make clear what they can do by failing to make clear what they are, will be more likely fail to be trustworthy because they will be less able to clearly offer their reliability. This will best be counteracted by signalling that clearly to others. This should also help to promote trust.

We cannot guarantee that building trustworthiness will build trust. It does however lower the level at which it would be rational to form a judgement of trustworthiness, and this makes it easier for individuals to trust. It promotes what Thomas Simpson has called a 'culture of trust' (Simpson, 2023, Ch.6).

6.0 Conclusion

In making organisations, we still each have an individual coordination duty to try to do so in a way that will bring about the best state of affairs we can. We should then try to build more trustworthy organisations. Organisations can offer their reliability in ways that will give rise to reliance-related obligations. For this reason as well, they ought to aim to be trustworthy.

I have shown over the past two chapters that this is philosophically possible, and from the discussion in this chapter that it seems plausible that we can and often do create such organisations. A trustworthy organisation will be realistic about what it can do, appropriately committed to acting on its reliance-related obligations and understand itself and its actions as socially situated. It will think about the effects of others' reliance on it before it acts. It will be careful about when it offers its reliability to others but understand that it cannot always refuse to do so. It will ensure that it can bring about what it claims it can reliably do, and it will ensure that it can do so not just briefly or intermittently but for a significant span of time. If we build trustworthy organisations, then we cannot guarantee trust will accompany them; but we do make the placing of trust rational and easy.

We have seen that organisations can be trustworthy. This was part of our attempt to understand if a business could be. With this part of the puzzle in place, we can turn to what a business is and whether it faces any specific challenges to its trustworthiness.

12 | What is a business?

The aim of this chapter is to say what a business is. I will argue that it is a type of organisation: one which makes a “surplus” – an amount of something, usually money, which exceeds the amount it needs to operate (and which, vitally, is not a “profit”). §2 discusses some potential approaches to understanding business and rules out appealing to social purpose. §3 bypasses two more approaches: that businesses are nothing more than legal entities or that we ought to conflate them with corporations. §4 lays out my own account. §5 explains the circumstances in which businesses face specific economic pressures. §6 considers some implications of this account for thinking about whether businesses are justified, though that is not a line of argument I pursue.

*

So far, I have been focusing on organisations. We have seen that they can be trustworthy. That gets us some way towards answering our question: can a business be trustworthy? Only some of the way, we now need to know what a business is. So far, I have used the term without very much precision, relying on our commonsense understanding of it. I think a more precise description can be found that is useful because it gives us an insight into the pressures businesses face, which might affect their trustworthiness.

These challenges to trustworthiness and how they can be overcome, I address in the next chapter. In this chapter, I aim to give a descriptive account of business. My positive account argues that businesses are organisations which make a surplus (which is distinct from a profit), and which distribute this internally – that is, either to their members or to themselves. This understanding of business helps us to explain various aspects of our lives including what we mean when we say something is getting too like a business. It is, I think, an important and neglected social kind.

1.0 What is a business?

What is a business? Walking from the University of Leeds, I pass entities that seem, pre-theoretically, to be businesses. There is a sandwich shop owned by one man, a supermarket run by a multinational publicly listed company, a hairdressers where sole traders hire space, a pub run as a private limited company, another supermarket owned by an American private equity firm and another pub, which is

owned by a national publicly listed company. What, if anything, makes these entities (and countless others we each interact with every day) a business?

Our conversations with each other and in politics are filled with talk of business. Yet, little philosophical attention has been paid to explaining exactly what sort of thing business is. This may be because our everyday, pre-theoretical grasp of what is or is not a business feels secure. So far in this thesis, I have used “business” quite freely. What I have been getting at has been clear.

One way this question has been approached has been the main approach in business ethics. There, business has been taken to have a specific purpose and to be best defined as something that ought to fulfil this purpose.

Researchers in social ontology have more or less ignored “business” instead considering the “corporation”. This is not a route I propose to follow. “Corporation” may be a good social classification. In many cases the same entity may both be a corporation and a business. This seems unlikely to always be the case. Many entities which are intuitively corporations are not intuitively businesses – e.g., the University of Leeds or the Corporation of London – and *vice versa* – e.g., the sandwich shop or hairdressers I pass by.

My approach will be to first explain why I am planning to not take the social purpose or corporation-based routes to understanding what a business is. I will then turn to my own positive account.

2.0 Social purposes and description

A popular approach to understanding what a business is has been to answer the question by saying that a business is an entity that serves some particular social purpose. This may be to facilitate the exchange of goods and services or to promote greater economic efficiency in a market, driving overall welfare by increasing growth, to more expansive purposes such as serving “stakeholders” or that businesses exist because they are “licenced” by a state to serve these purposes, but gain further benefits – such as recourse to legal remedies, protection and

recognition of property, the training of employees, provision of infrastructure, direct financial support – which obligate them towards that state.¹

These are all social purposes accounts because they argue that what is characteristic of an entity that makes them “businesses” is some distinctive social purpose that they have which other entities do not have. As others have pointed out, despite its popularity clearly defining what a social purpose is can be difficult (Hsieh *et al.*, 2018). What I consider to be common across these accounts is that they take an entity to be characterizable as a business when it is the nexus of some specific set of ethical values or requirements. (I use “nexus” because “subject” misdescribes – for example – Friedman’s view, where ethical and fiduciary duties fall on individuals, but where what a business is appears to be a specific focal point of such duties.) The defender of the exchange of goods and services as a social purpose view would need to make some claim about how this contributes to a social good (perhaps coupling it with the second view), and so be claiming that what is characteristic of an entity’s being a business is that it helps promote this value. The view that businesses are just those entities that are created in certain ways by states and owe them specific duties similarly identifies a business as being an entity characterised as the nexus of different ethical requirements. The set of these values or requirements is the social purpose.

One route now would be to pursue this social purpose debate, and to similarly define a business as an entity that faces a particular distinctive set of ethical requirements or promotes some distinctive ethical values. That is not the route we will take. The social purpose route suggests that the descriptive project of determining what a business is requires the settling of some ethical debate about what businesses should be for. I think that these questions can be answered separately.

It may be that we cannot get a good descriptive account of what a business is, separate from answering these ethical questions. The social purpose approach seems

¹ The former view can be found in Hsieh *et al.* (2018, p.56); the second is my interpretation of the basic justification for businesses in the work of Milton Friedman (1970; Friedman and Friedman, 2002) and his followers; the third is the classic stakeholder view popular in much business ethics (e.g., Agle *et al.*, 2008) – and commonly endorsed by businesses (e.g., Business Roundtable, 2019) – and the latter is my summary of the argument in Hsieh *et al.* (2018, pp.57-58).

to assume, rather than argue, that this is the case. I do not want to, and so we will see if we can deliver a good descriptive account which does not address the ethical issues. I think that if we do, then we can get clearer on what is required to settle the ethical debate (see §5).

3.0 Clarifying the approach

I will advance my own account in the next section. First, I want to reject two prominent alternatives: defining businesses in terms of legal entities and as corporations.

3.1 Legal entities?

An obvious account of business would be to simply pick out those entities which have a certain legal status. A first pass would be that an entity is a business if and only if it has the legal status of a business (e.g., it is a registered company). Such a view is attractive. Legal recognition as a business has numerous effects: for example, it changes tax liabilities, criminal liabilities and property rights. Some of the concerns we started with in this thesis were questions about regulation, and those are – at heart – questions about the law. So, it may also help us to answer them.

I think that pursuing this course requires us to settle further philosophical debates, largely in the philosophy of law, which it is preferable if a descriptive account remains neutral between.

To say that a business is a legal entity is to adopt a concessory view of the legal status of businesses. On this view, businesses exist at the legal pleasure of the state and their existence is exhausted by the laws that define them. Concessory assumptions are prominent in discussions of business. They frequently lie behind social purpose views, for example, where they justify the idea that states can and ought to modify the legal status of businesses to better serve specific social ends (again, cf. Hsieh *et al.*, 2018).

Concessory views are not, however, the settled position of discussions in the philosophy of law.² There they are opposed by real entity views (that take business to exist entirely independently of their legal status) and institutionalist views (that

² Such discussions focus on the legal status of the “corporation”, but what I say here still stands.

take business to have an existence only partially dependent on their legal status) (Kurki, 2019, p.156, p.158).³ Adopting a concessory view requires settling this debate, and settling it in favour of the concessory position. That is a stance I think is hard to defend, but even if it can be we can avoid having to take a stance in this debate. I will do so in my account.

Your objection may be that any alternative account surely assumes the settlement of this debate, but more likely in favour of the real entity view. After all, I am hoping to be able to claim that there are some things and that these things are businesses. I do not need to take a side in the debate. I will pursue a descriptive account of business, but I do not intend that to settle the debate about what the objects to which the law refers are, when the law talks of “business”. For example, it may be that businesses are real entities; they may not however be real entities of the kind that can have rights and duties that the law ought to respect (unlike us). Then the concessory view may still be correct. The law assigns the legal status and duties to a business and their legal normative status is entirely so determined. Other combinations of legal status and ontological view are compatible.⁴ The adopter of the concessory view in their ontology of businesses cannot make this move. Her ontological account of business includes the legal mechanisms (contracts, licensing, etc.) which on that account constitute the business. If they actually refer to another entity, then her account is internally inconsistent. That problem does not affect a descriptive account and so leaves the legal question open.

³ I agree with Kurki fully that ontological questions and questions of legal status need to be treated separately. The issue here is, again, the mixing of normative and descriptive projects. It is present in social purpose views (Hsieh *et al.*, 2018) who assume a concessory view in their argument for social licensing. It is present in views who assume a participant view: that the existence of corporations is exhausted by the inter-personal contracts and adopt it as a view on their existence (Ewin, 1991) and, arguably (Ludwig, 2017, Ch.14). This obscures the kinds of arguments that need making. The debate would move more clearly if we clearly distinguished these descriptive and ethical question.

⁴ We may even be real entity theorists but think that businesses are not the right sort of real entity. For example, perhaps a corporation is composed of contractual relations. If it is, then perhaps that is all that the corporation is, but there is also a business (of the sort I describe) “tagged” by this corporation, but ontologically distinct. I think corporations are just a specific kind of decision-making structure, but I leave it open whether they have some distinct form as well.

There are also further advantages to separating arguments about what the legal status of businesses is. We intuitively identify things as businesses that lack such a status. We can identify illicit businesses, and these do not seem defective businesses (in terms of what they are). If a group of scammers does not register a business because it will engage in illegal activities, it seems odd to claim that this enterprise was never really a business. Informal businesses also seem to be aptly called businesses, even if they lack a particular legal status. Our three friends who decide to sell art seem to be a business, even if they have not registered their enterprise. In fact, it seems that we could not make sense of our notions of legitimate and illegitimate business if we lacked separate understandings of what a business is and its legal status. For these reasons, I will not go down a legalistic route and I will try to produce an account which is neutral on this question.

3.2 Some arguments against corporation-based approaches

The corporation, by which standardly we mean the joint-stock corporation, is certainly a powerful and concerning part of our world. Some have called them a ‘key institutional site in the production of crime and harm’ which ‘is rapacious, thirsty for profit and systematically criminal’ and so ought to be significantly curtailed and eventually abolished (Tombs and Whyte, 2015, p.69 & p.180). Others have called for the abolition of corporations because of the historical roots of the form in racist systems of oppression including colonisation and slavery (Logan, 2019). Exact figures to measure corporate dominance are difficult to attain. By one estimate from McKinsey, around 5,000 ‘large corporations’ with revenue exceeding \$1 billion represented approximately 28% (\$17 trillion of \$61 trillion) of ‘value added’ to the economies of countries in the OECD (Manyika et al., 2021). They undoubtedly hold much financial clout, and we should not underestimate the risk caused by other forms of power they may exercise, especially their political influence and regulatory capture.

Given the prominence of corporations for those of us who live in today's developed economies, I am not surprised that they have formed a major focus for theorising.⁵ That makes sense. They also pose concerning ethical problems, which need solutions. I have two concerns, however, which mean that I will not focus on the corporation myself.

We might think that the focus on corporations is justified because the paradigmatic form of a business is the joint-stock company. That appears to be unjustified. For example, the “Big Four” companies that dominate consultancy and accountancy – and which play a key regulatory role as “gatekeepers” – are not joint-stock companies. If they are not paradigmatic businesses, why not? Perhaps, we instead say that what is paradigmatic of businesses is that they have limited liability, and that this is what makes something “corporate”. Still, this leaves the question of why it should be paradigmatic of *business* to be corporate in this way.

I also think this corporate turn is an unhelpful course to follow when trying to theorise ethically about businesses. It is to essentialise an element of businesses that we have good *prima facie* reasons to consider contingent. It risks essentialising a form of business that has only become dominant in recent centuries (Davoudi, McKenna and Olegario, 2018). It takes a certain (admittedly dominant) form of organisational design and packs it into our understanding of what a business is. In doing so, it is likely to obscure potential capabilities that businesses can have and ethical demands that they can meet. For example, it may obscure potential trustworthy businesses' options for organisational designs by overemphasising features unique to the corporate form.

The assumption by many of a concessory view perhaps justifies this, but I am not willing to assume the correctness of that view. The historical focus of theorising on responsibility also perhaps explains the theoretical draw of limited liability. It poses

⁵ The moral agency has been argued for in debates on moral responsibility including Björnsson and Hess (2017) and Silver (2019). Their moral agency has been argued for by French (1979) and argued against by Dewey (1926) and McMahon (1995).

Specific attention has been given to the nature of the corporation and the part that collective intentionality plays in that (Ludwig, 2017, Ch. 14; Hess, 2020; Bratman, 2022).

a challenge: a potential major discrepancy or block to moral responsibility formed by a specific group structure. The challenge can certainly be an interesting and important one to address, but we ought to be clear about the limited scope of our arguments when we are doing so. They apply to a specific way that businesses can be, not to anything essential about what they are.

So, I will look for a descriptive account of business which does not pack in this corporate design. We want something more general.

4.0 What a Business Is

I aim to provide a different account of what a business is. One that leaves open the ethical and legal questions about them, and which does not prioritise a specific structure over others.

I think that businesses are organisations who operate in an economic environment, which produce a surplus (which is not identical to a profit) and distribute that surplus internally (that is, either into the organisation itself or to its members). I will establish each of these points in turn.

4.1 Businesses as organisations

Businesses are intuitively often also organisations. Recall the general definition of an organisation as a group, G , whose members have roles and operate within a structure which determines what they will do within these roles. That seems to be true of most businesses. We saw how to specify this definition as members of a joint obligation to reliably respond to a decision-making structure in Chapter 8.

That a business is a form of organisation is helpful in picking out certain differences. Take our three friends who sell artworks, they are an organisation and intuitively a business. Now imagine three friends who each invest in the retail stock-market for profit but do so individually. They are not an organisation. It is also not clear that they are businesses.

We do sometimes call individuals businesses. I gave some examples at the start of this chapter of sole traders, and I agree that there is some intuitive pull to these individuals being businesses. I believe that this is because we understand such individuals to be acting as pseudo-organisations. They adopt a role *as if* there was an organisation (or in the hope that there will be such an organisation), and this is

why we extend the scope of business to include them. However, for the broader purposes of this theses, individual traders are relatively uninteresting. If they are to be trustworthy then they simply need to be trustworthy as described for any individual in Chapter 7 within the domain of action in which they carry out their business.

We are interested in how collective businesses can be trustworthy, and collective businesses appear to always be organisations. So, we will take them as being organisations.

4.2 Economic environments

When we talk of businesses as entities, we are talking about entities which operate in some kind of economic environment.⁶ This need not be a market environment. A state-owned supermarket in a control economy still seems to be a business. Yet, considering situations entirely removed from economic environments talk of there being businesses operating seems inappropriate. It seems essential to something being a business that it operates in an environment in which a predominant value for interactions and exchanges is an economic (and nowadays usually monetary) value.⁷

The pressures that such an environment produces are essential to understanding the problems that businesses face and will form the focus of my discussion in the next chapter. I will return to these pressures once we have discussed the rest of my account of what a business is.

4.3 Surplus-producing organisations

Consider two otherwise identical entities, HelpUK and Big Healthcare.⁸ Both entities provide a service to healthcare providers. They employ specially trained

⁶ We may use business more generally in the case of something that we have to do, but in this case are clearly not picking out entities in the world.

⁷ Of course, economic environments thus defined can expand. We can theoretically assign economic value to other exchanges, which can bring with it explanatory gains. We can actually extend the importance of economic values into other spheres (often through financialisation processes). What matters is that such values are the dominating ones in that sphere of activity.

⁸ By otherwise identical I mean in terms of organisational structure and other potential features. Both entities would have, for example, the same legal structure.

nurses for a specific medical condition. They are paid to do so by those healthcare providers. They supply that service at a lower cost than the payment they receive from the healthcare providers.

This means that they financially gain from their transactions. We would standardly call this a profit. After paying the costs of providing the service, their payment for it leaves them with some money. Profit, however, has a technical meaning that I want to avoid. Imagine that HelpUK gives this leftover money to other individuals affected by the medical condition in the form of grants. Big Healthcare gives that leftover money to all its employees in the form of end of year bonuses. In both cases, we would say that they have increased their expenditure and so reduced their profits. I think we can clearly understand what a business is by paying attention to that left-over money and what happens with it, but what that left-over money is does not seem to be “profit”. Instead, I will call this left-over money a “surplus”. I will define it further below.

Both HelpUK and Big Healthcare have a surplus and are otherwise identical. They use that surplus differently, and this seems – intuitively – to make Big Healthcare a business and HelpUK not. This example helps to clarify two things about business. First, it shows the limits of the commonsense view that businesses are entities involved in the exchange of goods and services, where this is intended purely descriptively. This description is clearly insufficient. It fails to pick something out, which should be picked out between HelpUK and Big Healthcare. It is also not clearly necessary. It is not obvious that a fraudulent investment scheme is not a business, yet it is unclear that this involves the exchange of goods and services.

The example also points us in a fruitful direction if we want to work out what a business is. I want to show that an entity is a business if and only if it is a surplus-producing organisation where the beneficiaries of that surplus are either the entity itself or some members of the entity.

To argue this, I will first attempt to show that there is a genus of organisations which operate in an economic environment. These are surplus-producing organisations. Businesses are a species of this genus. A species which is defined by the purpose to which it puts its surplus: helping itself and its members.

4.4 Surplus-Producing Organisations

In this section,⁹ I am going to show why given certain background assumptions we should think that (1) an entity like an organisation would come about, and (2) some of these entities would produce a surplus.

4.4.1 Organising reduces transaction costs

I have said that to organise is to offer one's reliability in responding to a decision-making structure. What I want to show here, is how accepting such a structure can reduce "transaction costs": difficulties caused by interacting in non-organised ways in conditions of imperfect knowledge.¹⁰

Imagine that you are one of three adult siblings. Your siblings and you are trying to arrange a holiday together. How ought you to go about doing so? One option is to individually argue your case for where, when and how to go. Note that you can do this without adopting a decision-making structure. You may, for example, say "I know what I want to do, but you're going to have to convince me." This does not necessarily imply that you offer your reliability in being convinced, or that you will accept the outcome of any discussion which you take part in. As you have not offered your reliability in being responsive, you are not in a decision-making structure, and so not organised.

Each of you knows that taking part in such discussions will have significant costs. Although you usually get on, discussion about holidays tend to become heated which takes an emotional toll on each of you. Whatever their nature, we can think of these costs in something you value as "transaction costs".

These transaction costs may be avoidable if you had better information and so could navigate them better. It could be that you each take a course in reading emotional

⁹ This section is heavily influenced by the economic theory of the firm, especially as it has been classically laid out (Coase, 1937). It also aligns with the view of human organisation as a sort of social simplification (Luhmann, 1997)

¹⁰ I am not arguing that organising necessarily reduces transaction costs. It may not. Certain organisations may, for example, exist specifically to increase transaction costs. They may exist to make it more difficult to do something that would otherwise be simpler. For example, one interpretation of legal institutions is that they increase the transaction costs on punishing an individual, making it more costly to do so in various ways (greater use of time, more evidence required to reach a threshold "beyond reasonable doubt", etc.) and by imposing further costs if someone subject to them imposes "vigilante" punishment, so that we can serve a further end of justice (by, for example, protecting the innocent).

cues and so are better able to avoid taking routes that lead to emotional hurt. Under such conditions, you may be able to avoid your transaction costs. Standardly, however, it is difficult to achieve such conditions. For example, there are further costs to gain the better information.

Imagine instead that you adopt a decision-making structure: say, that one of you will decide and the others accept that decision. Such decision-making procedures reduce transaction costs. You avoid the potential for hurtful comments which each of you could make to the other.

Forming an organisation can reduce our transaction costs. Often, organising does little more than save time. Decision-making structures tend to be more efficient than individuals reaching consensus to bring about some shared outcome.

Organisations can reduce transaction costs and in doing so participating in an organisation can save you certain things you value. Your siblings and you may each save time and emotional capacity which can be used elsewhere. You won't be snapping at your partner this evening. We can think of this as the organisation producing a surplus. That is, any "extra" produced by a reduction in transaction costs.

Talk of surpluses may seem odd where the surplus is something that cannot be easily deposited and redistributed (as in the case of emotional capacity). I am happy for the term to be fairly technical. Many transaction costs that can be reduced will be of this form, and money in particular has these features. For ease of exposition, I focus on one specific type of surplus: financial surplus.

4.4.2 Financial surplus-producing organisations

When organisations operate in economic systems of monetary exchange the reduction of transaction costs can produce a financial surplus (for example, through greater efficiency).¹¹ Participating in an organisation is unlikely to totally remove all costs. In fact, establishing and maintaining an organisation may produce its own costs. Organisations operating in a monetary economy will, nevertheless, be able to produce a monetary surplus. This will be what is left from its inputs minus the bare

¹¹ For simplicity, I stick to money. The account can easily be extended to other material surpluses. For example, if the agents were involved in a system of barter for goods.

cost levels of running itself. Equally, where the inputs are smaller than its running costs, deficits will be produced (see Fig.1).

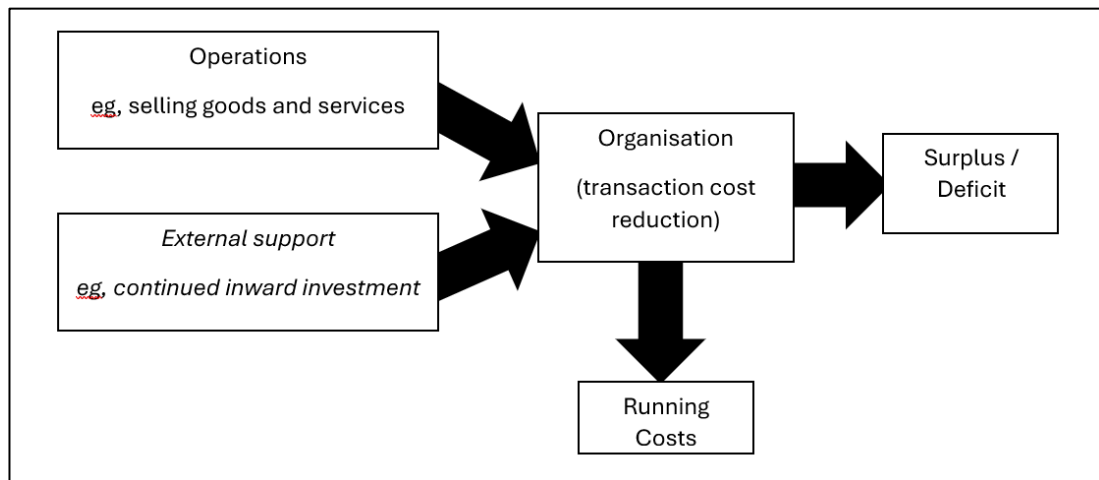


Figure 1: A diagram illustrating inflows and outflows producing a surplus or a deficit.

Exactly what running costs are in monetary terms is too detailed an economic question for me to fully settle. It will amount roughly to the reasonable market rate for which the organisation can do what it needs, like pay salaries, buy goods, etc.¹² If an organisation's running costs are larger than its inputs, then it will produce a deficit. Some organisations can produce a monetary surplus, as illustrated by the example of HelpUK and Big Healthcare.

¹² In the model we are operating in, we should be able to set parameters for "reasonable" running costs. This seems likely to be set by considering the market rate for some transaction were there no organisation, and the costs required by an organisation to maximally reduce its overall internal transactions costs compared to that market rate. In operating businesses, we may expect a more rough and ready notion of reasonable running costs. While the aim is to maximise as in the model, our notion of reasonableness should account for the information that the entity (or actors within it) can reasonably be expected to have (including information unavailable due to distortions in the price mechanism) and the social norms that shape expectations of such costs (e.g., whether greater health insurance benefits are expected). To reiterate, it is for ease that I am focusing on monetary surpluses. This is, I believe, the standard case for business. However, other tangible and intangible things may class as a "surplus". If a business can reduce time spent on work for all its employees through adopting new computer software, a question may arise about how to use the freed-up time. This is a "surplus", though a non-standard sort and the same distributive question arises. It could be distributed internally (e.g., reducing working hours for all or laying off certain workers) or externally (e.g., mandating voluntary service in the freed-up time). However, businesses are likely to usually be chiefly concerned with the distribution of their monetary surplus for the reasons of survival laid out in the main text.

4.5 What to do with the surplus?

Some organisations produce surpluses, some of which are monetary. The pressure to produce a monetary surplus will be heightened if the organisation operates in an economic environment, as businesses do. I want to argue that what differentiates a business from other kinds of organisation is that it uses its surplus for the benefit either of itself or some part of itself (including its members).

A surplus can raise a distributive question, when it is the right sort of thing to be distributed.¹³ Some cannot be, like you and your siblings' emotional capacities in the example we met earlier. Others like time and especially money – because it is easily distributable and exchangeable – will raise distributive questions. Once a financial surplus is produced, an organisation that has produced it would have to decide what to do with it.

Taking the model of organisations that I developed in Chapter 8, there is a distinction that we can make between those who are internal to the organisation and those who are external to it. Those who are internal to the organisation are those who are subject to its decision-making structure. This includes members, and also those alienated individuals who nonetheless act as they do because of its decision-making structure. Those who are external to the organisation are those who are not subject to its decision-making structure.

A business is characterised by the fact that it answers the redistributive question about its surplus, by distributing the surplus among its members. Figures 2 and 3 illustrate the classification of organisations that I have so far presented.

¹³ Equally, some forms of deficit raise distributive questions.

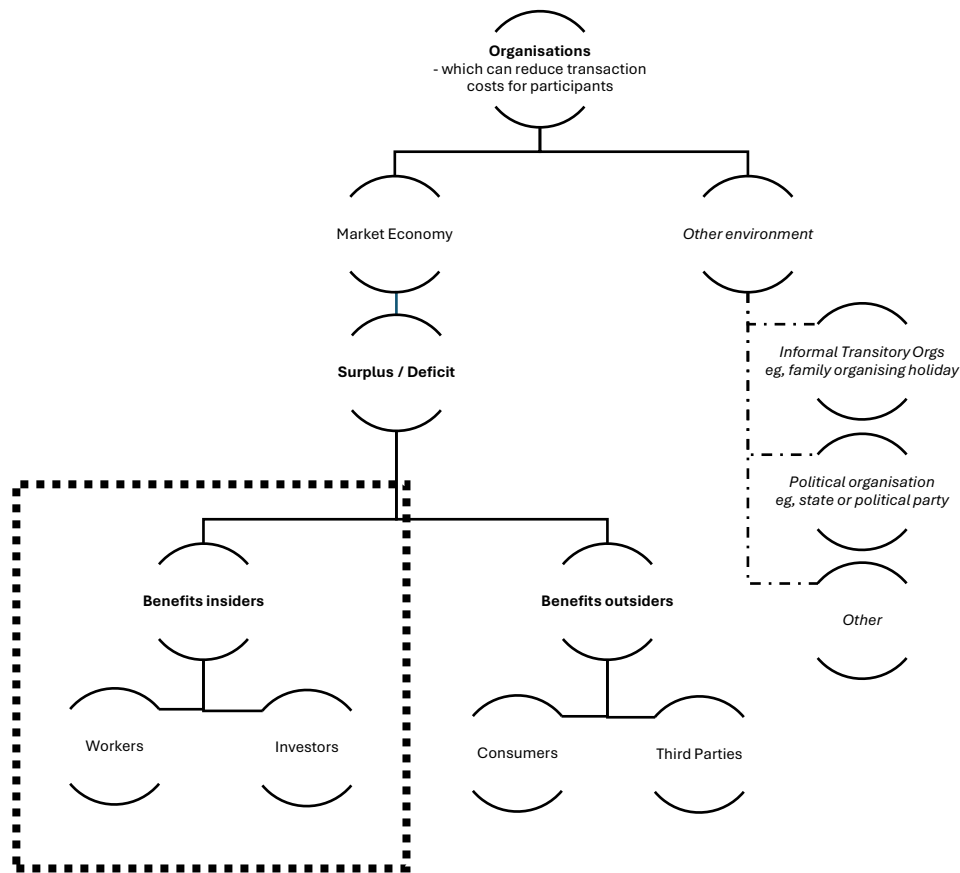


Figure 2 A schematic presentation of how to classify organisations. Businesses are those entities in the bottom left corner within the dashed square. The dot-and-dash line on the right branch represents potential further levels of classification, like the dashed square for businesses on the bottom left corner, which could help classify the kinds of organisations given as examples on the right of the diagram.

An organisation can decide to redistribute the surplus it produces to those who are internal or external to it. Say that Alan, Barbara and Cheryl do make a surplus of £600 and as an organisation decide to give themselves each a third of this. On this account they are then a business. They may also decide to spend the surplus “on the organisation itself”. This would happen if they spent the £600 on improving its art supplies or otherwise investing in the organisation’s activities. This is still an internal beneficiary, and so still counts as a business on this classification.¹⁴

¹⁴ It may be asked whether an investor is really internal to an organisation. On my account they undoubtedly are. An investor accepts (standardly) that they are becoming the owner of an organisation; in accepting that they are its owner they offer their reliability to fulfil this role and so become members. Owners do, standardly, have a role in a decision-making structure. For example, in many joint-stock companies they are taken to set the overall goal of the organisation even if this is simply its making them money.

Before clarifying why, I think this makes them a business, it is worth noting exactly what is included in this classification.

It is important to recall that a surplus is not identical with a profit.¹⁵ Non-profit organisations can still be businesses on this account, so long as they are non-profit in a specific way. Say that an organisation is a cooperative which redistributes all its surplus to its workers who are its sole members.¹⁶ It will be a non-profit and is – on my account – a business. It will not, however, include non-profits of other kinds. For example, a non-profit like the example of HelpUK we considered at the outset of this chapter, which redistributes the majority of its surplus in grants to those who have a specific medical condition. This is still a non-profit, but not a business because its surplus is distributed externally.

My account does not include as businesses organisations which redistribute all their surplus to consumers who are external to them. This might seem worrying. It seems, for example, to rule out as a business an organisation that decides to use its surplus to pass savings on to its customers. That is not what my account requires.

The aim of this chapter is to classify a business as a social kind. When thinking about such classifications what will matter is the general tendency of an organisation rather than each individual action. So, whether an organisation which uses its surplus to pass savings on to its customers is a business depends on whether (i) it still uses the majority of its surplus to help itself or its members, or (ii) whether this use of its surplus is part of a long-term disposition or a (near) one-off. For example, if the organisation does this to increase its surplus overall so that it can

¹⁵ Profit will be a part of a surplus. We use the term “profit” in a couple of ways colloquially. Let us consider profit the money left after all other costs by an organisation have been met. This is a more expansive set of costs than those included when considering the surplus (what I have termed “running costs”), so profits will be that part of a surplus left once these further costs are deducted (e.g., surpluses can be distributed as internal investments, which decreases profit).

¹⁶ The examples in Figure 2 are illustrative, and it is important to note that “consumers” may be members and so internal to the organisation. This is traditionally the way that it has been with cooperatives that are owned by those who shop with them. They are members in my technical sense if they have offered their reliability by becoming members; say to play a role in deciding how the coop is run.

redistribute a bigger surplus to its members then it still classes as a business because of (ii).

4.4 What a business is

I am suggesting that the correct way to classify which entities are businesses is to consider them as organisations which operate in an economic environment, and which produce a financial surplus which they redistribute to beneficiaries who are internal to the organisation.

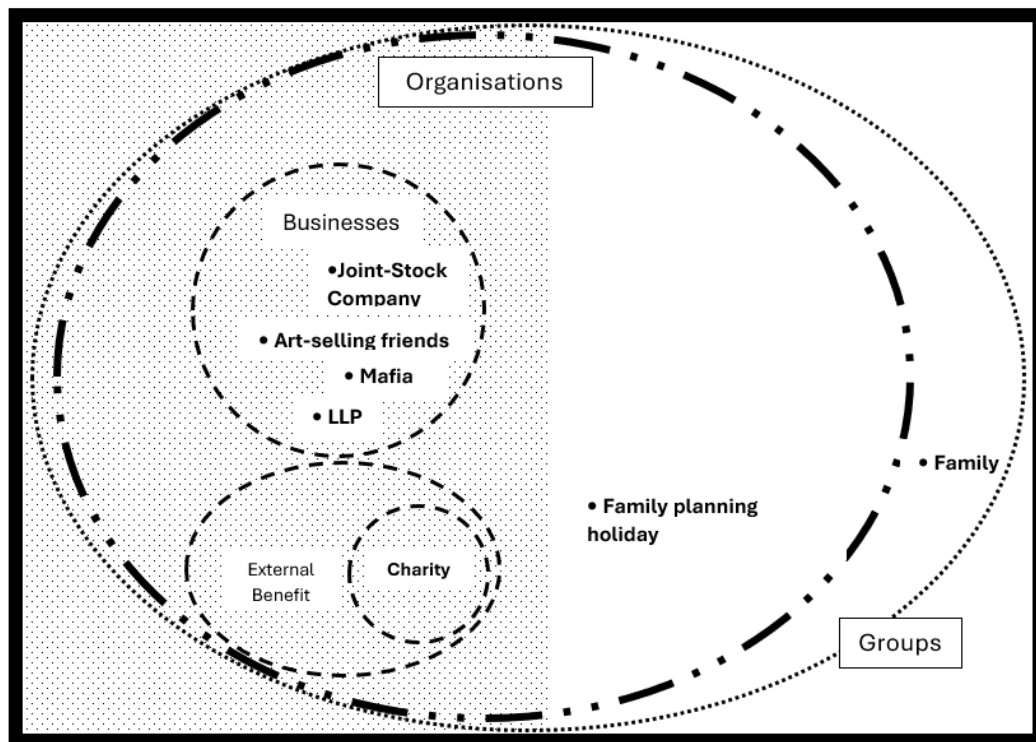


Figure 3 A diagram illustrating the classification of businesses laid out in this thesis. The hatched area represents those operating in an economic environment.

Why, though, think that this is what a business is? I think that we ought to accept this description of businesses because it is explanatorily useful. It captures a clear social kind. Those entities which it captures seem to fit many of our intuitive and everyday uses of business. Those that are excluded are not clearly the same sort of entity. So, even if this definition is somewhat stipulative, it does capture an important kind to which we ought to pay attention. It does so without embroiling us in the debates involved in taking legalistic definitions and catches a general class rather than narrowing our focus as corporation-based accounts do.

The classification is explanatorily useful because it explains what we pre-theoretically take to go wrong when we see organisations shift their activities in certain ways. Imagine that HelpUK is a surplus-producing organisation. It produces this surplus by providing services to a healthcare provider. Initially, it uses this surplus to give grants to individuals who have a specific medical condition. On my classification this means that HelpUK is not a business. This aligns with our intuitions, HelpUK seems more like a charity (though I will not attempt to define that as well). Suppose that HelpUK increases its staff's pay until most of its surplus is used funding these salaries. Even if it were able to maintain grant making at stable levels or increase it (by increasing the overall surplus), something seems to have changed about the nature of the organisation. The classification explains what has gone wrong. HelpUK has become a business. Similar observations extend to various spheres where the rise of "business-like" entities may be considered problematic such as education and state provision of services and some of the effects of financialization.¹⁷ Its changes in activity have shifted it into an important social classification: it is acting as a business would.

The classification I have proposed makes clear what we are picking up on (without all the technicality) in our everyday discussions. It is useful and satisfies much of what we seem to mean in everyday talk by business. It is neutral on questions of legal status. I have claimed that there are such entities, but I leave it open as to whether this is what the law refers to when it talks of such entities. Absent a better definition, we should accept it. A business is a surplus-making organisation operating in an economic environment which distributes that surplus internally.

5.0 Economic pressures

Economic environments will put particular pressures on an organisation, and these pressures will increase the importance for that organisation of producing a surplus.

¹⁷ It is worth noting that there is a sense in which the word "corporate" (or "business-like") is used to describe the culture or atmosphere of a working environment. A charity whose staff are purely wagedworkers may be appropriately described like this (though perhaps there is nothing inappropriate about it being like that). This I think is simply to compare a different (and contingent) descriptive feature which is prominent in contemporary businesses.

Let us return to the example that I have used throughout this Section: Alan, Barbara and Cheryl's organisation which sells artworks. This organisation has costs – the art materials, the petrol for the car driven to the craft fairs, licensing costs for accountancy software, etc. – which need to be met. To continue to exist the organisation will therefore at least need to “strike even”, making its running costs equal to these expenses.

Volatility and competitive practices will produce further pressures, which may mean that just striking even is not an option if the business is to continue to exist.

Businesses with larger surpluses may be able to increase their internal investment finding greater efficiencies or producing higher quality products, giving them a competitive advantage. Alan, Barbara and Cheryl's business could be outcompeted, for example, by some other friends who produce a larger initial surplus and use it to finance higher quality art materials. Equally, if they have only struck even, they may be more open to volatility and risk which arises in the market. For example, Alan may become unable to afford petrol costs without payment if fuel prices unexpectedly rise, which would require an increase in the surplus.

In economic environments, many organisations will find that they need to do more than just strike even to continue to exist. This economic pressure may cause problems for businesses' trustworthiness, and we will return to it in the next chapter.

6.0 Should we distribute a surplus internally?

If businesses are entities classified in the way that I have suggested, then we may wonder if we ought to distribute a surplus internally. I leave the overall justification of this an open question. There are likely to be justified situations in which business can come about; and I assume this in what follows. As I will argue in the next chapter, this need not prevent an organisation from being robustly trustworthy and so settling the issue further is beyond what is needed for this thesis.

What this account does show is that there are different ways to understand how businesses come about. Which of these is correct (and they may both be) will alter what is needed to establish whether businesses are justified. I leave this open, but

that this account helps to clarify what justifications are needed may help us to answer some of the outstanding ethical questions about business.

6.1 Actual use

The first potential explanation of how businesses come about is that businesses are how they are because certain individuals with power in them use them in this way. On this view, once an organisation that produces a surplus has come about certain actual individuals may act on it in such a way that the surplus is distributed internally. For example, in HelpUK, the HR Director may think that by increasing staff salaries she will increase productivity in a way that will further help their external beneficiaries. She may be mistaken (staff may simply take the extra payments with no change in how they work) or fail (by implementing the salary increases in a way that does not encourage greater productivity).

An actual use explanation is plausible. It helps to explain, for example, what happens in cases of so-called “managerial capture”. These are cases in joint-stock (or other investor-return based) companies where executives use the organisation to increase their personal rewards at the expense of either internal investment (including wider staff salary raises) or shareholder/investor returns. The underlying entity in such cases has usually remained identical. It is the uses to which individuals capable of exercising its control put it that has changed.

6.2 Systemic function

The second explanation relates to businesses’ potential systemic function. Given that organisations can produce surpluses and that such surpluses can go to internal beneficiaries, the existence of a business would be a way to incentivise participation in a risk-filled environment. As we have seen they can produce surpluses because they reduce transaction costs. That may change how rational it is for an individual to engage in the risk-filled environment. That would happen were the potential reward from the surplus to be sufficiently high. On this understanding, a business would be an organisation with a specific function: to provide incentives to agents to undertake risky interactions in an economic sphere.

This could then be developed into a functional explanation which explains how businesses come to be. For the system to function, it requires entities of this kind. We would then expect them to persist so long as they continue to provide sufficient

incentives to individuals to participate in them. This explanation again seems plausible. We might think of it as the “entrepreneurial” explanation of business. Businesses function to reward individuals for risk-taking, and they serve the need for individuals to maximise their overall payoffs.

Which of these explanations (for how businesses come about) is right may change the overall conditions that would justify businesses’ existence. The former would require a justification in terms of what an individual may do; the latter may have to appeal to the justifiability of the overall system. These may affect how justified, overall, we consider businesses to be. That is, however, a question that I will park. Whichever they come about, in the next chapter we will see if businesses can be trustworthy.

6.0 Conclusion

In this chapter, I have argued that a business is a surplus-making organisation which distributes that surplus to those internal to it. This classification makes sense of many standard uses of the term “business” and helps us to clarify what goes wrong in a situation where concern for internal beneficiaries begins to trump concern for external beneficiaries: we notice a shift in the type of entity with which we are dealing. I have left open whether, ultimately, we can justify the existence of such entities, or what further purpose they may serve.

We now have a clearer picture of what kind of organisations are businesses. They face pressures because they operate in an economic environment, especially when such an environment is competitive and volatile (as markets are). In the next chapter, we will consider whether this makes it more difficult for them to be trustworthy than it was for the organisations considered in Chapter 11.

13 | A Trustworthy Business

The aim of this chapter is to answer the question with which we started this thesis, “Can a business be trustworthy” with a qualified yes. Much of the argumentative work has already been done in Chapters 10 and 11, as businesses are a form of organisation. In this chapter, I focus on specific problems businesses face because of the important role that money plays for them. §1 lays out three potential ways that financial concerns can undermine trustworthiness and provides details on how these come about because of the given and adopted salience of money. I argue that, ultimately, they can be overcome, and so businesses can be trustworthy. It also offers a case study into Reuters which illustrates these problems. §2 provides a picture of a trustworthy business.

*

Finally, we can turn to the question with which we started: what would a trustworthy business look like? The key features of a trustworthy business are those of a trustworthy organisation, which I laid out in Chapter 11. A trustworthy organisation will be realistic about what it can do, appropriately committed to acting on its reliance-related obligations and understand itself and its actions as socially situated. It will be careful about when it offers its reliability to others and understand that it cannot always refuse to do so. It will ensure that it can do what it offers its reliability to do, and it will ensure that it can do so not just briefly or intermittently but consistently. A trustworthy business is an organisation and so it will do this too.

This chapter focuses on a specific problem for businesses. As we have just seen, “business” is a social kind: organisations characterised by the fact that they are active in an economic environment, which therefore face pressures to increase their financial surplus. Businesses, as we saw, are therefore not just profit-making organisations. They are also not just “corporations”, a specific form of organisation which has become prominent in business over the last couple of centuries (Davoudi, McKenna and Olegario, 2018).

This need to succeed financially is distinctive of businesses, as opposed to other organisations. It poses specific challenges for businesses achieving trustworthiness.

I argue that businesses will, by their nature, find that economic concerns have a high salience. This can result in businesses failing to be trustworthy. This is not, however, a necessary feature of business – not even of corporations. I end the chapter with a sketch of a trustworthy business.

1.0 The problem with money

An organisation is a practical entity: it exists to carry out some action(s).

Organisations can be “businesses” either because they are put to such a use or because the system that they operate in causes them to function like this. Either way, a feature of a business’ activity will be that it must make money else the business will cease.

I will consider the obstacle to being trustworthy caused by three ways that financial considerations can distort a business’ reasoning about what trustworthiness requires:

1. making money can become a default reason in internal deliberation.
2. making money can overpower the importance of other considerations if a business is close to collapse.
3. making money can be adopted as a value to the exclusion of others.

The reason that (1) to (3) are problems for trustworthiness is that when making money comes to be considered in these ways, the importance that reliance-related obligations ought to have can become distorted. As I have argued, the reliance-related obligations we need to be aware of to be trustworthy are *pro tanto* reasons to exclude other reasons and so ought to be acted on absent, say, a moral reason to act differently. The importance of making money is overblown in (1), (2) and (3) distorting this reasoning. In (1) and (2), making money can become unjustifiably more important than other relevant reasons, including those that the business must be responsive to if it is to be trustworthy. In (3), it excludes consideration of them. If a business cannot appropriately reason about what is required of it to be trustworthy, then it will not be trustworthy. So, a business affected by (1) to (3) can fail to be trustworthy.

I will argue that none of (1) to (3) create insurmountable obstacles to trustworthiness. So, (1) to (3) do not show that a business *cannot* be trustworthy. They only show that to be trustworthy, they must avoid these distorted forms of reasoning. In practice, (1) to (3) may happen frequently as features of actual businesses.

In this chapter, I focus on the challenges to trustworthiness for businesses. That should not be interpreted as implying that I think many businesses fail to meet these challenges. That is an empirical question beyond the limits of this study. In everything I say here, it is worth bearing in mind that I do think that financial considerations and trustworthiness can be pursued together. In certain circumstances, they may be complementary. The importance of financial considerations for business will be high. They can incorporate this into how they signal their reliability, and often the fact that they do have financial concerns will be part of what sets the reasonable limits of an interpretation of those offers. Prospective trustors ought to recognise that money does matter. The problems arise when financial considerations take such a weight that they distort the businesses reasoning about its reliance-responsive obligations. Before turning to the theoretical unpacking of these problems, I want to introduce a case study which illustrates them.

1.1 Case Study: Reuters Trust Principles

As I discussed in Chapters 10 and 11, an important feature of trustworthy organisations will be that their decision-making structures embed principles which promote appropriate reasoning about their reliance-responsive obligations. This will often be done in the form of constitutional policies. These make demands of the kind of reliability another individual must offer the organisation, and in doing so obligate her to stick to these principles. This can prevent members from altering these organisational values and so help to ensure greater consistency over time.

Businesses are no different. They can place such constitutional requirements on their members; that is, staff and investors. Reuters, best known for its international news wire, has a historical decision-making structure which closely aligns with my theoretical suggestions – unless otherwise referenced, this information comes from (Thomson Reuters, 2018, Parts 02-04.). In 1941 Reuters was under financial

pressure and its then shareholders needed to raise funds. They wanted to do so without reducing the trustworthiness of the business. To ensure this they set up a specific decision-making structure. Investors who bought shares in the company were required to accept an ‘Agreement of Trust’ which ‘imposed obligations on Reuters and its employees’. This was a constitutional document, in my terms, which contained the following principles (Thomson Reuters, 2018, Part 02):

No single interest, group or faction should come to dominate Reuters;

There should be no interference in the integrity, independence or freedom from bias of Reuters; and

Reuters should devote the appropriate resources to maintaining its leading position in the international news business.

These are not exactly aligned with my account, but they are broadly in its spirit. Assuming that most people Reuters interacted with relied on it for objective news reporting, then adopting a constitutional principle of freedom from control by a specific interest and ‘integrity, independence or freedom from bias’ seems a good way to secure this kind of trustworthiness. Reuters endured financial strain again after the 1940s; its future attempts to maintain these constitutional principles illustrate the wider issues which businesses face if they wish to be trustworthy.

Close to financial collapse in the 1970s and 1980s, Reuters owners publicly listed the company in 1984. The board wanted to protect what had become known as the ‘Reuters Trust Principles’. To ensure this, a new decision-making structure was adopted. This gave special voting rights to a company called the ‘Reuters Founders Share Company’. It could restrict an individual’s share of the company to 15% and it had the power to “veto” decisions which would alter the business’ constitution. If, for example, a shareholding went over 15%, then the Founders Share Company could limit the voting rights of this shareholder and facilitate the sale of any excess shares. It was also tasked with overseeing that the company as a whole respected the Trust Principles.

Again, this structure demonstrates how a decision-making structure and the adoption of constitutional principles can safeguard a consistency in reasoning and

so in its values and actions by binding present and future members. On my account this is likely to have contributed to trustworthiness.

Since 2008, Reuters merged with the Thomson Corporation, which was itself majority owned by a holding company – Woodbridge – that manages the interests of the Thomson family of media moguls. This happened because Reuters again came under financial pressures. This time, the trustees changed Reuters’ constitution to allow for an individual stake to exceed 15%; despite opposition from the unions representing Reuters’ employees, which contested the sale by appealing to the Trust Principles (Edgecliffe-Johnson, 2007a). Following my account, this illustrates the value of a constitutional policies, which allow members to appeal to other members by pointing out the principles which are understood as being part of their joint obligation.

This sale required the new organisation, Thomson Reuters, to adopt a modified and expanded version of the Trust Principles (Edgecliffe-Johnson, 2007b). It required new investors to accept the Principles and included the provision that the Founder Share Company would regain its rights should Woodbridge reduce its investment below a 35% threshold.

Since then, Thomson Reuters has sold a 55% stake in its ‘Financial & Risk business’ “Refinitiv” to one of the world’s largest private equity companies, Blackstone.

Though this is meant to include certain ‘protections’ and rights for the Founders Share Company, what these are is less clear. This was viewed as a financial necessity for its ailing news business, as it provided it with a guaranteed \$325 million a year (Edgecliffe-Johnson et al., 2019). The new company appears to not be bound by the trust principles. It has ‘pulled’ stories supplied by Reuters (which remain under Thomson Reuters’ control) from the information terminals they supply, including stories covering the anniversary of the suppression of protestors in Tiananmen Square (Edgecliffe-Johnson et al., 2019).¹

¹ This part of the case also shows the complexities of trusting business in the modern world, where ownership is often opaque. As I said at the end of Chapter 11, a lack of external clarity about what your organisation is will not help it to be trustworthy. It becomes difficult to signal your reliability clearly. It also provides further reason to support a trustworthiness first

As R.S. Downie (1998, p.117) puts it, '[w]hile a philosopher has no special competence in sifting facts, he at least can hope to clarify the theoretical issues.' I cannot determine whether the Trust Principles have actually worked and whether the most recent changes have undermined them. The case does, however, demonstrate the tensions that can arise for businesses between financial concerns and principles.

First, it is notable that the imposition of a charter or other constitutional document can be implemented and accepted by investors. Businesses can ensure consistency over time and bind investors to the obligations of membership. Decision-making structures can be put in place, and accepted, which help to foster and guarantee this long-term commitment to principles. Constitutional policies are an option for business.

Second, businesses' adherence to these principles will be strained when they face financial pressures. Trustworthiness will be increased where businesses are honest about the financial pressures that they face and the limits that this must place on their offers of reliability. There is a reasonable level of financial concern for a business. The trouble is that this reasonable level can be exceeded. Let us turn to how that comes about.

1.2 The given value of money

1.2.1 Money as a default reason

I distinguished in Chapter 10 between two forms of organisational value. Given values arise from their relevance to the action that is being pursued. Adopted values are explicitly chosen by the organisation.

I think that given values can encourage the use of *default* reasons. These are reasons for which we tend to act because they are common reasons when we deliberate on some matter. So, we easily come to *habitually reason* using these default reasons, and this leads us to miss or fail to respond appropriately to other relevant reasons which we have, but which we are unused to considering in this context. I think that

approach, where we can hopefully gain the benefits of having trustworthy entities even if it remains difficult to precisely place trust.

the given value of money for businesses can lead to financial considerations becoming default reasons within them.

Money matters for businesses: they require it to continue. This means that when it considers what to do, a business and its members will often find reasons related to securing the money that they require to be relevant. The ways this can happen are clear. In most internal deliberation within a business, reasons related to money will be accepted because they are reasons that are common to a range of sub-activities in which the organisation is engaged, just as they are reasons relevant to its overall activity. This means that it is often easier in such deliberation to reach for that readily available money-related reason. Such reasoning can then become habitual within a business, leading it to neglect other relevant reasons including the demands of its trustworthiness. The problem is not that the importance of making money is overvalued, but that the organisation simply fails to consider other relevant reasons.

I think this “default” to financial reasons is common, but it is not damaging to a business’ potential to be trustworthy. It is negligent reasoning on the part of the business and its members and a failure to consider something that it ought to consider, chiefly out of laziness in deliberation. It can, however, be easily rectified. Policies can be adopted and role-requirements implemented which require the consideration of non-money related reasons. Control will need to be exercised to avoid slipping back into default reasoning, but there is no philosophical reason to think that such control cannot be brought to bear on members. If we accept, as I have suggested, that members may have individual moral duties to bring it about that an organisation is trustworthy then part of fulfilling that duty may be for individuals to avoid slipping into deliberating just on default financial reasons within their roles.

A further problem may arise in that businesses may favour structures that allocate significant resources to considering financial concerns. Following my account of organisational attention, they will then pay more attention to it, potentially missing other relevant reasons. This may compound – and in turn be compounded by – default reasoning about financial considerations. A business with this kind of attentional deficit may fail to notice reasons relevant for its trustworthiness, and so

it will fail to be trustworthy. This is, again, not an insurmountable problem.

Businesses can – and ought – to avoid having such limited attention.

1.2.2 Trustworthiness at breaking point

A different problem can come about because of how important making money is if a business is to continue existing. The importance of financial reasons can increase for an organisation if it has a choice between prioritising such reasons and its own dissolution. For example, financial reasons will be particularly important for a company approaching bankruptcy.

Again, this does not make trustworthiness impossible. I noted in Chapter 11 that realism about what it could do was important for a trustworthy organisation, just as realism about our own competencies and motivations is required to be a trustworthy individual. There are, therefore, reasonable limitations around what might be required of a trustworthy business, and these include some level of financial concern which it ought to take steps to clearly signal to others.

Where a business already has reliance-related obligations and changes how it acts because of financial pressures (which may have come about unexpectedly) so that it fails to fulfil those obligations, then it will fail to be trustworthy. There may be no choice in the matter, but a business facing its closure should prioritise fulfilling those obligations which will mitigate the highest levels of others' vulnerability to it. This will at least ensure it is as trustworthy as possible.

I do not, however, think that this will always be the correct route. It may sometimes be more appropriate for an organisation including a business to dissolve itself, than to give in to financial pressures. This is not, however, a way to be trustworthy. A dissolved business cannot act; and so it will fail to fulfil its reliance-related obligations which are left unspent when it ceases to exist. Dissolution may still sometimes be the better route because it may better lessen others' potential vulnerability to the organisation, than if the organisation were to continue and behave in non-trustworthy ways.

This may be preferable to continuing to exist but having to behave untrustworthily. A trustworthy business ought not always to favour the principle that “[its] future ... takes precedence over [its] principles” (as expressed by Pehr Gyllenhammer, chair

of the Reuters Founders Share Company, at the time of its merger with Thomson Corporation). Sometimes principles should (again, *contra* Mr Gyllenhammer's expressed views) not "get in the way of what shareholders [think is] a good and attractive deal" (Edgecliffe-Johnson, 2007b).

I think that a business should favour its dissolution in two circumstances related to trustworthiness. First, (absent some comparable benefit from its existence) if it cannot achieve trustworthiness in the future. This may happen if a business operates in such a volatile financial environment that it cannot guarantee its long-term behaviour, making robust trustworthiness impossible. A business may also be incapable of achieving trustworthiness if its operations fundamentally prevent it from being reliance-responsive trustworthy in an appropriate way. The highest level of vulnerability to fossil fuel companies, for example, is arguably that of those who will suffer most heavily from the effects of fossil fuel burning. It is unclear that a fossil fuel company can reasonably avoid offering its reliability to them while respecting its moral duties. Yet, they are unlikely to be able to be reliable in a way that mitigates this vulnerability because their main operation – extracting and processing fossil fuels – contributes to it. Members may even have something like a "de-coordination" duty to work for the dissolution of such organisations.

Second, a company should favour dissolution if it is likely to act in a way that lets down (some of) those who currently trust it. That is, if its continued existence will lead to it being untrustworthy. This is because it is likely to produce more harm by acting against what would be required to fulfil its reliance-related obligations, than it would if it simply stopped acting. This is a particular risk given trust's evidence resistance, which means that those who trust the organisation may not notice its swing towards untrustworthiness. If individuals trust Reuters for impartial news on financial risk and this is undermined by changes to its business structure necessary for its continued existence, then it would cause significant harm to these trusting individuals were Reuters to undergo these changes rather than if they had to look elsewhere to trust. It is ethically better to avoid these further harms. The harm is likely to be lessened if the trust simply had to end than if it were to become

misplaced. This will be the case if the likely harms outweigh any potential trustworthiness the business may be able to achieve in the future.

This is the organisational equivalent of an individual requirement to clearly signal if the limits of her reliability are reached. For example, a scout leader who finds worsening weather conditions mean that she can no longer keep the expedition safe, is more trustworthy if she takes them back to the bus.

Whether dissolution is justified will depend on other factors. For example, the internal obligations and potential harm the dissolution of the business may do to its members, and the extent of the harm that failing to meet the obligation may do to others.²

There will be more, and less, trustworthy ways to go about such dissolution. Businesses ought to clearly signal to others that they must withdraw their offer of reliability. They may seek to lessen others' vulnerability by demonstrating alternative trustworthy actors if these are available.

Overall, the fact that financial considerations are important for business and can become overly important is not fatal to their ability to be trustworthy and even robustly trustworthy. They create constraints, but constraints that limit the business only if they are allowed to ensnare its entire reasoning. Financial considerations just need to be valued appropriately, alongside the requirements of trustworthiness which will usually outweigh them.

1.3 The adopted value of money

When money is adopted as a chief or the only value of a business, we have a problem of adopted value.³ Where this excludes consideration of how non-members are reliant on the organisation and the associated obligations, it runs counter to trustworthiness. It stops a business from reasoning in an appropriate way to be trustworthy.

² On this potential harm and the costs of reorganisation see French (1994). I leave the ethical question of how this reorganisation's costs ought to be shared for future work.

³ This section draws on discussions with Philip Pettit following a talk he gave at a workshop on 'Philosophical Perspectives on Trust and Reputation' at the Oxford University Said Business School. These have helped to clarify the thoughts presented here.

Again, I am not here to arbitrate how often businesses do adopt this value in a way that is distorting of their reasoning. A traditional view, in which the business seeks long-term, stable profits over an extended time-period may be compatible with trustworthiness. Trustworthiness may even assist with this, if it attracts customers and other businesses to collaborate with you. What matters is that this will not always be the case. The adoption of money as a value can distort the salience of reliance-related obligations. For example, this happens whenever short-term profit is pursued, to the exclusion of other considerations.

This sort of value may be given to financial reasons quite frequently. It need not be. That it would stop businesses being trustworthy is a further reason to question whether we ought to allow it to be dominant.⁴ The facile argument that it is what businesses must do, should be rejected because of the reasons given in the previous sub-section. It is *a* thing that businesses must do; it is a further step to think it is *all* they must do.

A possible way of arguing that financial reasons ought to dominate a company would be something along the following lines. Market conditions are such that a business can only continue through its own growth, given the effects of competition and volatility. The law of the market is grow or perish. So, financial reasons must dominate if a business is to continue. To be persuasive, this argument would need some justification for the exclusion of all other usually important considerations. Potential arguments include that property rights somehow trump moral and other values or that the social good provided by such conditions provides some higher moral value (most likely because of some link between growth and improved general welfare). Neither seem easy to defend. The former seems to give implausibly strong weight to property rights. The latter may be defensible from a consequentialist perspective, but it rests on a contingent claim about what the market produces and

⁴ Other reasons usually refer to social purposes. As explained in the last chapter, I have avoided a social purpose account of businesses. I am open to there being a social purpose to business, but I am remaining neutral as to whether there should be or what it amounts to. The requirement to be trustworthy is independent of any social purpose they may or may not have.

one which – given the harms we know markets can cause – we have good reason to doubt. So, I do not think that there is a good argument for this strong conclusion.

We may though weaken it. We may argue that it is justifiable for money making to be adopted as the dominant value, in certain circumstances. On my account, we would talk of a decision-making structure being adopted in which certain members are allowed to set the goal for the business. An example of this kind of decision-making structure would be a joint-stock company: shareholders determine what the business is for. In such a decision-making structure, it is open to those who set the business' goal to set its sole goal as making money. As this is the company's goal, it is what the members have obligated themselves to work towards and so – it seems – how they must act. (Alternatively, we could look to a Friedmanian argument that there are some special obligations, e.g., fiduciary duties, for employees to give special consideration to money making.)

The weaker conclusion is also not justified. It does not seem tenable to argue that such considerations ought to trump the demands of trustworthiness. As I previously argued, reliance-related obligations are still likely to come about for the business and it ought to appropriately consider them.

The best reason why financial considerations should not dominate is that, as I have argued, the joint obligation underlying businesses as organisation will often be constrained by a moral duty to make it trustworthy. Those within such businesses ought to seek to make them trustworthy. Doing so requires not adopting – or reforming – a decision-making structure which is an impediment to trustworthiness as those just described are. These moral considerations are binding on shareholders as much as any other member. They ought not to be setting money-making as the only goal. (Similar things can be said about the strength of fiduciary duties against moral duties.)

This argument then says nothing about the *possibility* of a business being trustworthy; it simply notes that there are ways that this can be undermined, but these ways are themselves objectionable. That they happen is then regrettable, and where money is adopted as the only goal this will often mean that the business is not trustworthy, we ought to try to get rid of such a decision-making structure.

If those arguments fail and we must accept that at least some forms of business must only be concerned with financial reward, then they will not be trustworthy. We should then ask whether we want that type of business. If corporations cannot be trustworthy, then that is (another) reason to ban such forms of organisation or to regulate them in such a way that their non-harmful reliability is ensured. I will return to this in the Conclusion.

Ultimately, I think there is little reason to think that even corporations have such limited concerns. Shareholders can invest with other values, and as far as their values set the limits of the business' values there is no reason to exclude a wish to be trustworthy. This again illustrates the advantages of binding investors to constitutional principles.

The problems that financial considerations pose to businesses can be solved. These financial concerns are what distinguishes businesses from other organisations, and that they can be solved means that businesses can be trustworthy.

2.0 A Trustworthy Business

There is nothing in the nature of business that suggests trustworthiness is impossible for them. A trustworthy business will share the features of a trustworthy organisation. It will be *realistic* about what it can do, and this will require it to have a realistic appraisal of its financial constraints. It will consider this when offering its reliability to others and will make sure that it does so without obfuscation or overpromising. It will appropriately respond to the reliance-related obligations that it then acquires. It will look out for its reasoning being distorted by default financial reasons or the adoption of money-making as its sole goal. It will make sure that it has measures to counteract this, including appropriate requirements on its members – including its investors. It will best do this by adopting constitutional policies which embed the appropriate values and limit how its members can reasonably deliberate within their roles. It will then ensure that it acts as trustworthiness requires.

To become a trustworthy business, a business may have to transform its self-understanding, so that it views and reasons about its actions as potential sources of reliance for non-members, in ways that it has not previously done. A decision about

whether to continue with or withdraw an investment in a certain area should take into account how it has offered its reliability and may have caused others to become reliant on it. It will think about the suppliers it has who need its custom, the hospitality premises and shops that are reliant on its members' custom, etc. It will recognise these as reasons which may have brought about reliance-related obligations, and it will recall statements that it and its members have made about "being here for the long-haul", "being committed to this place", and "bringing 500 jobs into the area, supporting new employment and helping to drive the local economy". If it is trustworthy, it ought to have been more careful and because it has not been it will recognise what it is obligated to do. It will understand that these cannot just be shrugged off with an apology and a statement that "times change". It will recognise that it ought to do what is required of it; and this will count in its decision about whether to continue investing. It will recognise that standardly the demands of trustworthiness ought to be very important reasons for it: that it is obligated to act in these ways. If it does so, it will be trustworthy.

This socially situated view of a business is certainly drastically different from a traditional image that still often dominates our cultural idea of business. Steps to understand corporate social responsibility have gone some way to realising it, as have movements to express and explain the social purpose of business. If we want trustworthy businesses, then we need to go further. We need businesses to recognise the enormous role that they play as objects of reliance for many different individuals. The socially situated view develops the traditional view of stakeholders (Goodpaster, 1991; Agle *et al.*, 2008), by elaborating the kinds of impacts on those affected by a business's actions that a business ought to take into account, and by providing a framework for considering them. Business must act to reform their internal decision-making so that the demands of trustworthiness are acted on, to the reasonable exclusion of financial concerns.

Our lives go better when we can trust more agents. Reliance-responsive trustworthiness opens options to us which we would otherwise lack; it allows us to do more, and to pursue goals we otherwise could not. This thesis has shown that this form of trustworthiness could be had by organisations and by businesses. Given their power and their prevalence, such a world would come with significant benefits.

14 | Conclusion

'[I]n business the earning of profit is something more than in incident of success. It is an essential condition of success; because the continued absence of profit itself spells failure. But while loss spells failure, large profits do not connote success. Success must be sought in business also in excellence of performance; and in business, excellence of performance manifests itself, among other things, in the advancing of methods and processes; in the improvement of products; in more perfect organization, eliminating friction as well as waste; in bettering the condition of the workingmen, developing their faculties and promoting their happiness; and in the establishment of right relations with customers and with the community.

In the field of modern business, so rich in opportunity for the exercise of man's finest and most varied mental faculties and moral qualities, mere money-making cannot be regarded as the legitimate end. Neither can mere growth in bulk or power be admitted as a worthy ambition. Nor can a man nobly mindful of his serious responsibilities to society, view business as a game; since with the conduct of business human happiness or misery is inextricably interwoven.'

Louis Brandeis (1914)

We set out from some observations about our modern societies. I pointed to the shift in social relations from those between individuals to those between an individual and an institution. Increasingly we interact with businesses, and as a result we are highly reliant on them. We may not always think about it, but – regardless of whether we do – we stand in a bare reliance relation to them, which means that much of what we can do, much of the success of our actions and our plans for life, depend on how businesses behave. This means that businesses are 'inextricably interwoven' with our 'happiness or misery'.

I argued that when we face these kinds of vulnerabilities in our interactions – these 'social vulnerabilities' – then we look out for features that agents have which reduce

them. These features are what our trust-concepts help us navigate.

“Trustworthiness” picks out these features; “trust” the attitudes which help us to track them.

Some of these vulnerabilities are linked to our needs for love or for group membership, but those sorts of vulnerabilities ought not to dominate our interactions with business. What matters is our reliance: the way that businesses’ actions determine the success of what we do.

I have argued that it therefore should matter to us whether a business can be trustworthy. It should matter regardless of whether we actually do trust business. We should care because other people are trusting businesses on our behalf. Regulators, the world over, do so. However we feel about business, whether they can be trustworthy matters.

So, working out if a business could be trustworthy required working out what we meant by trustworthiness. I argued that the best way to ensure that others act in ways that mitigate our vulnerability to our reliance on them is for them to be certain about what they can do, disposed to respond to reliance-related obligations and to appropriately reason about those obligations. I introduced the reliance-related obligation as something that comes about because others offer us their reliability as a reason for action, and in doing so obligate themselves to ensure their reliability because of norms which govern this practice of exchanging reasons.

Being trustworthy in this way requires agents who have a realistic view of their capabilities, who have the control to make sure that they act in certain ways, the long-term disposition to do so, and who can appropriately recognise the importance of these reliance-related obligations. Many individual people do this. What about businesses? Unfortunately, it was not clear what a business was. So, we set out to work this out as well.

I argued that a business is a form of organisation; and one which produces more than it needs (a ‘surplus’) and redistributes this ‘surplus’ to itself or its members, and which operates in an economic environment.

I offered a new account of what an organisation is. An organisation is a joint obligation between its members. A joint obligation is a special form of reliance-responsive obligation; it comes about when we offer our reliability to each other to be responsive to a decision-making structure, whether that is the handing out of some roles with their requirements or being bound by some decision-making procedure, like a vote. This helps us to understand why organisational roles are reason-giving for us.

I then argued that such organisations can be agents because the outputs of the decision-making structures amount to intentions which cause action and because they do so on the basis of reasoning and beliefs about the world. I argued that these agents can have control, because they can ensure that their decisions get turned into actions and can limit the actions that individuals subject to their decision-making structure may take. They can have and appropriately reason about reliance-related obligations. They can do this by embedding appropriate values into their decision-making structure. They can do so consistently, especially if they put in place ‘constitutional policies’ which bind future members to accepting these values. If they can do this, then an organisation can be trustworthy.

The special nature of businesses and the environment in which they operate means that they face pressures to continue to make such a surplus; if they fail to do so, then they will cease to exist. That this is ‘an essential condition of success’ does not mean that ‘mere money-making [can] be regarded as a legitimate end’. I argued that it undermined trustworthiness when things end up just getting done for financial reasons, or it is decided to only pursue profit, or the business faces an existential financial threat, and these financial concerns cause it to disregard the demands of its trustworthiness. These problems, we saw, are not insurmountable. They need to be guarded against, but it is not necessary that businesses reason in this way. In fact, they should not be this way.

A business can be trustworthy; it needs to think about others’ reliance on it before it acts. It will (i) dedicate sufficient resources to ensuring that it understands what these bare reliances are, and (ii) ensure that appropriate mechanisms are in place to enable such awareness and to gain appropriate information. It will need to be

careful about how it signals its reliability to others, which limit or expand the scope of its reliance-related obligations, and so what is required of it if it is to be trustworthy. It will be careful about when it offers its reliability to others and understand that it cannot always refuse to do so. It will be *realistic* about what it can do and will have a realistic appraisal of its financial constraints. It will use this to ensure that it can bring about what it claims it can reliably do, and it will ensure that it can do so not just briefly or intermittently but consistently. It will have to make sure that it looks out for the dominance of financial reasoning, and that it has measures in place to appropriately mitigate for this. It will ensure this by adopting constitutional policies which embed the appropriate values and limit how its members can reason within its decision-making structure. Finally, it will ensure that it has appropriate control over its members to ensure that they act in the ways its trustworthiness demands.

I also set out to answer a question about a change in policy, which has seen regulators increasingly pursuing trust-focused form of regulation of business. My argument shows that in some instances, our regulatory turn may be justified, and we may even want to trust businesses (though we do not have to). The mismatch between how a trustworthy business must be and how businesses are will mean that some trust-focused regulation is unjustified: it aims to allow regulators and the public to trust, but its aim is misguided and does not let them trust *well*. This is because the regulation misunderstands what trustworthiness would require. This thesis has tried to illuminate what should be looked for.

We may accept that to create a trustworthy business is demanding, and decide it is too demanding. There may also be situations in which it is impossible. I think that when this happens, this thesis has shown that we need to change regulatory tack. The main vulnerability we face to businesses is our reliance on them. We can reduce this vulnerability through trustworthiness, but we can also do it by forcing them to be reliable. External imposition of reliability is an option here if we cannot reduce our vulnerability through trustworthiness. Often, that will be the most sensible route. Though it has its disadvantages. Reliability is less responsive, and external regulation is costly, even when there is political will to make it happen.

To think that this is all we can do is to be too pessimistic about business. A pessimism which leads to us missing opportunities to improve our world. The pessimist may often have the justification of experience, but she will then only have what she has experienced. I think Louis Brandeis was right when he wrote the words which opened this conclusion. I think we should focus more on the wide variety of ways we can make businesses better. This thesis has been the start of a contribution to that.

We can make businesses trustworthy. We really ought to. It is difficult to avoid the creation of reliance-related obligations, and businesses ought to fulfil those obligations they create. Making that trustworthiness more robust may be more demanding, but those who work in business should recognise the duties they each have to improve the world in which we live among which is the duty to make their organisations more trustworthy. The duty to build trustworthy businesses is among our ‘serious responsibilities to society’.

Doing so means working to shift some of those elements of our cultural views of business, which contribute to their failure to be trustworthy. The most insidious of these is, I think, the view that a business is simply a tool to make money, whether for investors or for employees. This distorts the importance of financial gain for a business and the importance that its trustworthiness ought to have. It views businesses as if they are a socially isolated tool, to be used by whoever has them as she personally sees fit. That is not what businesses are. Businesses too are connected to us in a web of reliance. When working in them, we need to recognise the social situatedness they have. We need to try to remember that the considerations most readily at hand – which will often be financial ones – are not the only considerations that matter. We can embed the organisational salience of different issues by placing them into role requirements, by producing decision-making procedures which care about them, by allocating sufficient resources for the business to attend to them, and by producing constitutional policies which bind future members to these values. This happens in business when we adopt codes of conduct or professional ethics, for example.

We may need to build the institutional support for such trustworthiness to grow. Regulators may have a role to play. Especially in creating conditions where non-trustworthy businesses cannot gain unscrupulous advantages which damage those businesses trying to be trustworthy. That may require a tightening of regulation on non-trustworthy businesses. My thesis also suggests that, as they increasingly do, regulators should be more concerned about governance and organisational structures, and that they should consider how they can protect organisational structures which help to ensure trustworthiness. This may require legal structures which better allow for the protection of constitutional policies and greater regulation of investors.

That, however, will not fulfil our duties to build trustworthy organisations. We can help to cultivate them. We – as workers, managers and investors – can accept this duty, and change the businesses and organisations of which we are members by organising together to make them trustworthy, as they can be.

The power, influence and near ubiquity that businesses have in our societies means we are heavily reliant on them. Their unreliability can harm us all. This thesis has shown that we can do better. It has shown that trustworthy businesses are possible. It has argued that we have duties to create them. If we choose to fulfil these duties, we will live in a society in which we can trust that bit more easily, worry that bit less, and in which each of us can pursue what we want to do more securely, simply and successfully.

References

- Agle, B.R. *et al.* (2008) 'Dialogue: Toward Superior Stakeholder Theory', *Business Ethics Quarterly*, 18(2), pp. 153–190.
- Alfano, M. and Huijts, N. (2020) 'Trust in Institutions and Governance', in *The Routledge Handbook of Trust and Philosophy*. Routledge.
- Alonso, F.M. (2009) 'Shared Intention, Reliance, and Interpersonal Obligations', *Ethics*, 119(3), pp. 444–475.
- Alonso, F.M. (2014) 'What Is Reliance?', *Canadian Journal of Philosophy*, 44(2), pp. 163–183. Available at: <https://doi.org/10.1080/00455091.2014.919722>.
- Anderson, J.M. and Waggoner, I. (2014) 'Recent History: A Shift to Reforming Corporations from Within', in *The Changing Role of Criminal Law in Controlling Corporate Behavior*. RAND Corporation, pp. 41–68. Available at: <https://www.jstor.org/stable/10.7249/j.ctt1287mfw.10> (Accessed: 9 September 2021).
- Anscombe, G.E.M. (2000) *Intention*. 2nd ed. Cambridge, MA.; London: Harvard University Press.
- Baier, A. (1986) 'Trust and Antitrust', *Ethics*, 96(2), pp. 231–260.
- Baier, A. (1991) 'Trust', *The Tanner Lectures on Human Values* [Preprint]. Available at: https://tannerlectures.utah.edu/_documents/a-to-z/b/baier92.pdf (Accessed: 13 July 2021).
- Bauer, P.C. and Freitag, M. (2017) 'Measuring Trust', in E.M. Uslaner (ed.) *The Oxford Handbook of Social and Political Trust*. Oxford University Press, pp. 15–36. Available at: <https://doi.org/10.1093/oxfordhb/9780190274801.001.0001>.
- Becker, L.C. (1996) 'Trust as Noncognitive Security about Motives', *Ethics*, 107(1), pp. 43–61.
- Bennett, M. (2021) 'Demoralizing Trust', *Ethics*, 131(3), pp. 511–538. Available at: <https://doi.org/10.1086/712564>.
- Bennett, M. (2024) 'Trusting groups', *Philosophical Psychology*, 37(1), pp. 196–215. Available at: <https://doi.org/10.1080/09515089.2023.2179478>.
- Berman, Y. and Hovland, T. (2024) *The Impact of Austerity on Mortality and Life Expectancy*. London: International Inequalities Institute, LSE. Available at: https://eprints.lse.ac.uk/123915/1/WP_139.pdf (Accessed: 24 November 2024).
- Björnsson, G. and Hess, K. (2017) 'Corporate Crocodile Tears? On the Reactive Attitudes of Corporate Agents', *Philosophy and Phenomenological Research*, 94(2), pp. 273–298. Available at: <https://doi.org/10.1111/phpr.12260>.
- Blackman, R. (2023) 'Explaining Role-Based Reasons', in A. Barber and S. Cordell (eds) *The Ethics of Social Roles*. Oxford: Oxford University Press, pp. 156–174. Available at: <https://doi.org/10.1093/oso/9780192843562.003.0008>.
- Braithwaite, J. (2011) 'The Regulatory State?', in R. Goodin (ed.) *The Oxford Handbook of Political Science*. Oxford University Press, p. 0. Available at: <https://doi.org/10.1093/oxfordhb/9780199604456.013.0011>.
- Brandeis, L.D. (1914) *BUSINESS--A PROFESSION* by Louis D. Brandeis — Louis D. Brandeis School of Law Library, *Business - A Profession*. Available at: <https://louisville.edu/law/library/special-collections/the-louis-d.-brandeis-collection/business-a-profession-by-louis-d.-brandeis> (Accessed: 1 November 2024).

- Bratman, M.E. (1992) 'Practical Reasoning and Acceptance in a Context', *Mind*, 101(401), pp. 1–15.
- Bratman, M.E. (1999) *Intention, Plans, and Practical Reason*. Stanford, CA: Center for the Study of Language and Information (David Hume Series).
- Bratman, M.E. (2014) *Shared Agency: A Planning Theory of Acting Together*. Oxford: Oxford University Press. Available at: <https://doi.org/10.1093/acprof:oso/9780199897933.003.0005>.
- Bratman, M.E. (2022) *Shared and Institutional Agency: Toward a Planning Theory of Human Practical Organization*. New York, NY: Oxford University Press. Available at: <https://doi.org/10.1093/oso/9780197580899.001.0001>.
- Bruning, L. (2023) 'Virtuous Chameleons: Social Roles, Integrity, and the Value of Compartmentalization', in A. Barber and S. Cordell (eds) *The Ethics of Social Roles*. Oxford: Oxford University Press, pp. 298–320. Available at: <https://doi.org/10.1093/oso/9780192843562.003.0015>.
- Bruno, D. (2020) 'Trust-Based Theories of Promising', *The Philosophical Quarterly*, 70(280), pp. 443–463. Available at: <https://doi.org/10.1093/pq/pqz086>.
- Business Roundtable (2019) *Business Roundtable Redefines the Purpose of a Corporation to Promote 'An Economy That Serves All Americans'* | *Business Roundtable*. Available at: <https://www.businessroundtable.org/business-roundtable-redefines-the-purpose-of-a-corporation-to-promote-an-economy-that-serves-all-americans> (Accessed: 15 July 2021).
- Cartwright, N. and Hardie, J. (2012) *Evidence-Based Policy: A Practical Guide to Doing It Better*. New York, NY; Oxford: Oxford University Press.
- Casey, Baroness (2023) *Baroness Casey Review: Final Report: An independent review into the standards of behaviour and internal culture of the Metropolitan Police Service*. London: Baroness Casey Review. Available at: <https://www.met.police.uk/SysSiteAssets/media/downloads/met/about-us/baroness-casey-review/update-march-2023/baroness-casey-review-march-2023a.pdf>.
- Coase, R.H. (1937) 'The Nature of the Firm', *Economica*, 4(16), pp. 386–405. Available at: <https://doi.org/10.1111/j.1468-0335.1937.tb00002.x>.
- Collins, S. (2013) 'Collectives' Duties and Collectivization Duties', *Australasian Journal of Philosophy*, 91(2), pp. 231–248. Available at: <https://doi.org/10.1080/00048402.2012.717533>.
- Collins, S. (2017) 'Duties of Group Agents and Group Members: Duties of Group Agents and Group Members', *Journal of Social Philosophy*, 48(1), pp. 38–57. Available at: <https://doi.org/10.1111/josp.12181>.
- Collins, S. (2019) *Group Duties: Their Existence and Their Implications for Individuals*. Oxford: Oxford University Press.
- Collins, S. (2022) 'I, Volkswagen', *The Philosophical Quarterly*, 72(2), pp. 283–304. Available at: <https://doi.org/10.1093/pq/pqab032>.
- Cooper, D. (1998) 'Responsibility and the "System"', in *Individual and Collective Responsibility*. 2nd edn. Rochester: Schenkman Books, pp. 133–150.
- Coppola, F. (2019) *The Case for People's Quantitative Easing*. Medford, MA: Polity ('The Case for' Series).
- Davidson, D. (2006) *The Essential Davidson*. Oxford: Clarendon Press.

Davoudi, L., McKenna, C. and Olegario, R. (2018) 'The Historical Role of the Corporation in Society', *Journal of the British Academy*, 6(s1), pp. 17–47. Available at: <https://doi.org/10.5871/jba/006s1.017>.

D'Cruz, J. (2020) 'Trust and Distrust', in J. Simon (ed.) *The Routledge Handbook of Trust and Philosophy*. New York: Routledge, pp. 41–51. Available at: <https://doi.org/10.4324/9781315542294>.

Debus, D. (2016) 'Shaping Our Mental Lives: On the Possibility of Mental Self-Regulation', *Proceedings of the Aristotelian Society*, 116(3), pp. 341–365. Available at: <https://doi.org/10.1093/arisoc/aow019>.

Dewey, J. (1926) 'The Historic Background of Corporate Legal Personality', *The Yale Law Journal*, 35(6), pp. 655–673. Available at: <https://doi.org/10.2307/788782>.

Domenicucci, J. and Holton, R. (2017) 'Trust as a Two-Place Relation', in P. Faulkner and T. Simpson (eds) *The Philosophy of Trust*. Oxford: Oxford University Press, pp. 149–160. Available at: <https://doi.org/10.1093/acprof:oso/9780198732549.003.0009>.

Downie, R.S. (1998) 'Responsibility and Social Roles', in *Individual and Collective Responsibility*. 2nd edn. Rochester: Schenkman Books, pp. 117–132.

Edgecliffe-Johnson, A. (2007a) 'Reuters Unions Question Thomson Bid', *Financial Times*, 14 May. Available at: <https://www.ft.com/content/bb75e45e-0265-11dc-ac32-000b5df10621> (Accessed: 27 November 2024).

Edgecliffe-Johnson (2007b) 'Thomson Accepts Reuters Voting Code', *Financial Times*, 16 May. Available at: <https://www.ft.com/content/6460287a-03e4-11dc-a931-000b5df10621> (Accessed: 27 November 2024).

Edgecliffe-Johnson, A. *et al.* (2019) 'Thomson Reuters Brushes Off News Wire Suitors', *Financial Times*, 1 November. Available at: <https://www.ft.com/content/ed533f68-f711-11e9-a79c-bc9acae3b654> (Accessed: 27 November 2024).

Ewin, R.E. (1991) 'The Moral Status of the Corporation', *Journal of Business Ethics*, 10(10), pp. 749–756.

Faulkner, P. (2015) 'The Attitude of Trust Is Basic', *Analysis*, 75(3), pp. 424–429.

Faulkner, P. (2017) 'The Problem of Trust', in T. Simpson and P. Faulkner (eds) *The Philosophy of Trust*. Oxford University Press. Available at: <https://doi.org/10.1093/acprof:oso/9780198732549.001.0001>.

Faulkner, P. (2018) 'Finding Trust in Government', *Journal of Social Philosophy*, 49(4), pp. 626–644. Available at: <https://doi.org/10.1111/josp.12262>.

Financial Reporting Council (2024) *UK Corporate Governance Code 2024*. London: Financial Reporting Council.

Fischer, J.M. and Ravizza, M. (1998) *Responsibility and Control: A Theory of Moral Responsibility*. New York, NY; Cambridge: Cambridge University Press (Cambridge Studies in Philosophy and Law).

Fleming, P. (2016) 'Good Advice', *Philosophical Papers*, 45(1–2), pp. 181–207. Available at: <https://doi.org/10.1080/05568641.2016.1187484>.

French, P.A. (1979) 'The Corporation as a Moral Person', *American Philosophical Quarterly*, 16(3), pp. 207–215.

French, P.A. (1982) 'Crowds and Corporations', *American Philosophical Quarterly*, 19(3), pp. 271–277.

French, P.A. (1994) 'The Practical and Ethical Costs of Corporate Reengineering', *Business & Professional Ethics Journal*, 13(3), pp. 3–18.

French, P.A. (ed.) (1998a) *Individual and Collective Responsibility*. 2nd edn. Rochester: Schenkman Books.

French, P.A. (1998b) 'Types of Collective', in *Individual and Collective Responsibility*. 2nd edn. Rochester: Schenkman Books, pp. 33–50.

Friedman, M. (1970) 'A Friedman Doctrine-- The Social Responsibility of Business Is to Increase Its Profits', *The New York Times*, 13 September. Available at: <https://www.nytimes.com/1970/09/13/archives/a-friedman-doctrine-the-social-responsibility-of-business-is-to.html> (Accessed: 9 October 2024).

Friedman, M. and Friedman, R.D. (2002) *Capitalism and Freedom*. 40th Anniversary Ed. Chicago: University of Chicago Press.

Friedrich, D. and Southwood, N. (2010) 'Promises and Trust', in H. Sheinman (ed.) *Promises and Agreements: Philosophical Essays*. New York, NY; Oxford: Oxford University Press.

Frost-Arnold, K. (2014) 'The Cognitive Attitude of Rational Trust', *Synthese*, 191(9), pp. 1957–1974. Available at: <https://doi.org/10.1007/s11229-012-0151-6>.

Gilbert, M. (1990) 'Walking Together: A Paradigmatic Social Phenomenon', *Midwest Studies In Philosophy*, XV(1), pp. 1–14. Available at: <https://doi.org/10.1111/j.1475-4975.1990.tb00202.x>.

Gilbert, M. (1992) *On Social Facts*. Princeton Paperback Print. Princeton, NJ: Princeton University Press.

Gilbert, M. (2000) *Sociality and Responsibility: New Essays in Plural Subject Theory*. Lanham, MD: Rowman & Littlefield Publishers.

Gilbert, M. (2005) 'A Theoretical Framework for the Understanding of Teams', in N. Gold (ed.) *Teamwork: multi-disciplinary perspectives*. Houndmills, Basingstoke; New York, NY: Palgrave Macmillan, pp. 22–32. Available at: <https://doi.org/10.1057/9780230523203>.

Gilbert, M. (2006) 'Who's to Blame? Collective Moral Responsibility and Its Implications for Group Members', *Midwest Studies In Philosophy*, XXX(1), pp. 94–114. Available at: <https://doi.org/10.1111/j.1475-4975.2006.00130.x>.

Gilbert, M. (2008) *A Theory of Political Obligation: Membership, Commitment, and the Bonds of Society*. 1. publ. in paperback. Oxford: Oxford University Press.

Gilson, E.C. (2014) *The Ethics of Vulnerability: A Feminist Analysis of Social Life and Practice*. New York: Routledge (Routledge Studies in Ethics and Moral Theory, 26). Available at: <https://doi.org/10.4324/9780203078136>.

Gold, N. and Sugden, R. (2007) 'Collective Intentions and Team Agency', *The Journal of Philosophy*, 104(3), pp. 109–137.

Goodpaster, K.E. (1991) 'Business Ethics and Stakeholder Analysis', *Business Ethics Quarterly*, 1(1), pp. 53–73.

Govier, T. (1992a) 'Distrust as a Practical Problem', *Journal of Social Philosophy*, 23(1), pp. 52–63. Available at: <https://doi.org/10.1111/j.1467-9833.1992.tb00484.x>.

- Govier, T. (1992b) 'Trust, Distrust, and Feminist Theory', *Hypatia*, 7(1), pp. 16–33.
- Govier, T. (1993) 'Self-Trust, Autonomy, and Self-Esteem', *Hypatia*, 8(1), pp. 99–120.
- Hampshire, S. (1992) *Innocence and Experience*. London: Penguin Books.
- Hardin, R. (2006) *Trust*. Cambridge: Polity (Key concepts in the social sciences).
- Hardin, R. (2013) 'Government without trust', *Journal of Trust Research*, 3(1), pp. 32–52. Available at: <https://doi.org/10.1080/21515581.2013.771502>.
- Haslanger, S. (2000) 'Gender and Race: (What) Are They? (What) Do We Want Them to Be?', *Noûs*, 34(1), pp. 31–55.
- Hawley, K. (2014) 'Trust, Distrust and Commitment', *Noûs*, 48(1), pp. 1–20. Available at: <https://doi.org/10.1111/nous.12000>.
- Hawley, K. (2017) 'Trustworthy Groups and Organizations', in P. Faulkner and T. Simpson (eds) *The Philosophy of Trust*. Oxford: Oxford University Press. Available at: <https://doi.org/10.1093/acprof:oso/9780198732549.001.0001>.
- Hawley, K. (2019) *How to Be Trustworthy*. Oxford: Oxford University Press. Available at: <https://oxford.universitypressscholarship.com/view/10.1093/oso/9780198843900.001.0001/oso-9780198843900> (Accessed: 13 July 2021).
- Heape, A. (2020) 'Promises, Obligation, and Reliance', *Philosophy and Phenomenological Research*, 104, pp. 150–170. Available at: <https://doi.org/10.1111/phpr.12752>.
- Hedahl, M. (2013) 'The Collective Fallacy: The Possibility of Irreducibly Collective Action Without Corresponding Collective Moral Responsibility', *Philosophy of the Social Sciences*, 43(3), pp. 283–300. Available at: <https://doi.org/10.1177/0048393113489266>.
- Hess, K.M. (2020) 'Assembling the Elephant: Attending to the Metaphysics of Corporate Agents', in S. Bazargan-Forward and D. Tollefsen (eds) *The Routledge Handbook of Collective Responsibility*. New York, NY; London: Routledge (Routledge handbooks in philosophy). Available at: <https://doi.org/10.4324/9781315107608>.
- Hieronymi, P. (2008) 'The Reasons of Trust', *Australasian Journal of Philosophy*, 86(2), pp. 213–236. Available at: <https://doi.org/10.1080/00048400801886496>.
- Hieronymi, P. (2022) *Freedom, Resentment, and the Metaphysics of Morals*. Princeton, NJ; Oxford: Princeton University Press.
- Hinchman, E. (2017) 'On the Risks of Resting Assured: An Assurance Theory of Trust', in P. Faulkner and T.W. Simpson (eds) *The Philosophy of Trust*. Oxford: Oxford University Press. Available at: <https://doi.org/10.1093/acprof:oso/9780198732549.001.0001>.
- Hinchman, E. (2020) 'Trust and Will', in J. Simon (ed.) *The Routledge Handbook of Trust and Philosophy*. New York: Routledge. Available at: <https://doi.org/10.4324/9781315542294>.
- Holden, D., Jones, R. and Johns, R. (1982) *The House of Saud*. London: Pan Books.
- Holton, R. (1994) 'Deciding to Trust, Coming to Believe', *Australasian Journal of Philosophy*, 72(1), pp. 63–76. Available at: <https://doi.org/10.1080/00048409412345881>.

- Home Office (2012) *Definition of Policing by Consent*. London: Home Office. Available at: <https://www.gov.uk/government/publications/policing-by-consent/definition-of-policing-by-consent> (Accessed: 25 November 2024).
- Hood, C. (2006) 'Transparency in Historical Perspective', in C. Hood and D. Heald (eds) *Transparency: The Key to Better Governance?* London: British Academy, Available at: <https://doi.org/10.5871/bacad/9780197263839.003.0001>.
- Horsburgh, H.J.N. (1960) 'The Ethics of Trust', *The Philosophical Quarterly* (1950-), 10(41), pp. 343–354. Available at: <https://doi.org/10.2307/2216409>.
- House of Lords Economics Affairs Committee (2021) *Quantitative Easing: A Dangerous Addiction?* HL Paper 42. London: House of Lords.
- Hsieh, N.-H. et al. (2018) 'The Social Purpose of Corporations', *Journal of the British Academy*, 6(s1), pp. 49–73. Available at: <https://doi.org/10.5871/jba/006s1.049>.
- Hugo, V. (2013) *Les Misérables: Volume One*. Ware: Wordsworth Editions Ltd (Wordsworth Classics, v. Volume I).
- Jackson, R.L. (2000) 'The Sense and Sensibility of Betrayal: Discovering the Meaning of Treachery through Jane Austen', *Humanitas*, XIII(2), pp. 72–89.
- Jaffer, S., Knaudt, S. and Morris, N. (2014) 'Failures of Regulation and Governance', in N. Morris and D. Vines (eds) *Capital Failure: Rebuilding Trust in Financial Services*. Oxford: Oxford University Press, pp. 100–126. Available at: <https://doi.org/10.1093/acprof:oso/9780198712220.003.0005>.
- Jaffer, S., Morris, N. and Vines, D. (2014) 'Why Trustworthiness Is Important', in N. Morris and D. Vines (eds) *Capital Failure: Rebuilding Trust in Financial Services*. Oxford University Press. Available at: <https://doi.org/10.1093/acprof:oso/9780198712220.001.0001>.
- Jaspers, K. (1946) *Die Schuldfrage: ein Beitrag zur deutschen Frage*. Zürich, Switzerland: Artemis.
- Jones, K. (1996) 'Trust as an Affective Attitude', *Ethics*, 107(1), pp. 4–25.
- Jones, K. (2004) 'Trust and Terror', in P. DesAutels and M.U. Walker (eds) *Moral Psychology*. Lanham, MD: Rowman & Littlefield, pp. 3–18.
- Jones, K. (2012) 'Trustworthiness', *Ethics*, 123, pp. 61–85. Available at: [https://doi.org/0014-1704/2012/12301-0004\\$10.00](https://doi.org/0014-1704/2012/12301-0004$10.00).
- Jones, K. (2019) 'Trust, Distrust, and Affective Looping', *Philosophical Studies*, 176(4), pp. 955–968. Available at: <https://doi.org/10.1007/s11098-018-1221-5>.
- Keeley, B. and Love, P. (2010) *From Crisis to Recovery: The Causes, Course and Consequences of the Great Recession*. Paris: OECD (OECD insights).
- Kelp, C. and Simion, M. (2023) 'What is Trustworthiness?', *Noûs*, 57(3), pp. 667–683. Available at: <https://doi.org/10.1111/nous.12448>.
- de Kenessey, B. (2020) 'Promises as Proposals in Joint Practical Deliberation', *Noûs*, 54(1), pp. 204–232. Available at: <https://doi.org/10.1111/nous.12269>.
- Kirby, N., Kirton, A. and Crean, A. (2018) 'Do Corporations Have a Duty to Be Trustworthy?', *Journal of the British Academy*, 6(s1), pp. 75–129. Available at: <https://doi.org/10.5871/jba/006s1.075>.

- Kirton, A. (2020) 'Matters of Trust as Matters of Attachment Security', *International Journal of Philosophical Studies*, 28(5), pp. 583–602. Available at: <https://doi.org/10.1080/09672559.2020.1802971>.
- Krishnamurthy, M. (2015) '(White) Tyranny and the Democratic Value of Distrust', *The Monist*, 98(4), pp. 391–406. Available at: <https://doi.org/10.1093/monist/onv020>.
- Kutz, C. (2000) *Complicity: Ethics and Law for a Collective Age*. Cambridge; New York, NY: Cambridge University Press (Cambridge Studies in Philosophy and Law).
- Lahno, B. (2001) 'On the Emotional Character of Trust', *Ethical Theory and Moral Practice*, 4, pp. 171–189.
- Lahno, B. (2004) 'Three Aspects of Interpersonal Trust', *Analyse & Kritik*, 26, pp. 30–47.
- Lahno, B. (2017) 'Trust and Collective Agency', in P. Faulkner and T. Simpson (eds) *The Philosophy of Trust*. Oxford: Oxford University Press, pp. 129–148. Available at: <https://doi.org/10.1093/acprof:oso/9780198732549.001.0001>.
- Lawford-Smith, H. (2019) *Not in Their Name: Are Citizens Culpable for Their States' Actions?* Oxford; New York, NY: Oxford University Press (New Topics in Applied Philosophy).
- Lenard, P.T. (2015) 'The Political Philosophy of Trust and Distrust in Democracies and Beyond', *The Monist*, 98(4), pp. 353–359. Available at: <https://doi.org/10.1093/monist/onv017>.
- Les misérables* (2019). Srab Films, Rectangle Productions, Lyly Films.
- Lewis, H.D. (1948) 'Collective Responsibility', *Philosophy*, 23(84), pp. 3–18.
- List, C. and Pettit, P. (2013) *Group Agency: The Possibility, Design, and Status of Corporate Agents*. First publ. in paperback. Oxford: Oxford University Press.
- Logan, N. (2019) 'Corporate Personhood and the Corporate Responsibility to Race', *Journal of Business Ethics*, 154(4), pp. 977–988. Available at: <https://doi.org/10.1007/s10551-018-3893-3>.
- Ludwig, K. (2017) *From Plural to Institutional Agency: Collective Action II*. Oxford: Oxford University Press. Available at: <https://doi.org/10.1093/oso/9780198789994.001.0001>.
- Luhmann, N. (1997) *Trust and Power: Two Works*. Facsimile of publ. from Chichester: Wiley, 1979. Translated by G. Poggi. Ann Arbor, Mich: UMI Books on Demand.
- Mackenzie, C., Rogers, W. and Dodds, S. (2014) *Vulnerability: New Essays in Ethics and Feminist Philosophy*. Oxford: Oxford University Press (Studies in Feminist Philosophy).
- Manson, N.C. and O'Neill, O. (2007) *Rethinking Informed Consent in Bioethics*. Cambridge New York: Cambridge University Press.
- Manyika, J. et al. (2021) *A New Look at How Corporations Impact the Economy and Households*. McKinsey Global Institute. Available at: <https://www.mckinsey.com/capabilities/strategy-and-corporate-finance/our-insights/a-new-look-at-how-corporations-impact-the-economy-and-households#/> (Accessed: 27 November 2024).
- Marder, M. (2020) 'Betrayal: A Philosophy', *Research in Phenomenology*, 50(1), pp. 79–98. Available at: <https://doi.org/10.1163/15691640-12341440>.

McGeer, V. (2002) 'Developing Trust', *Philosophical Explorations*, 5(1), pp. 21–38. Available at: <https://doi.org/10.1080/10002002018538720>.

McGeer, V. (2008) 'Trust, Hope and Empowerment', *Australasian Journal of Philosophy*, 86(2), pp. 237–254. Available at: <https://doi.org/10.1080/00048400801886413>.

McGeer, V. and Pettit, P. (2002) 'The Self-Regulating Mind', *Language & Communication*, 22(3), pp. 281–299. Available at: [https://doi.org/10.1016/S0271-5309\(02\)00008-3](https://doi.org/10.1016/S0271-5309(02)00008-3).

McGeer, V. and Pettit, P. (2017) 'The Empowering Theory of Trust', in P. Faulkner and T. Simpson (eds) *The Philosophy of Trust*. Oxford: Oxford University Press, pp. 14–34. Available at: <https://doi.org/10.1093/acprof:oso/9780198732549.003.0002>.

McKenna, M. (2006) 'Collective Responsibility and an Agent Meaning Theory', *Midwest Studies in Philosophy*, XXX(1), pp. 16–34. Available at: <https://doi.org/10.1111/j.1475-4975.2006.00126.x>.

McLeod, C. (2020) 'Trust', in E.N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy*. Fall 2020. Metaphysics Research Lab, Stanford University. Available at: <https://plato.stanford.edu/archives/fall2020/entries/rust/> (Accessed: 12 July 2021).

McMahon, C. (1995) 'The Ontological and Moral Status of Organizations', *Business Ethics Quarterly*, 5(3), pp. 541–554. Available at: <https://doi.org/10.2307/3857398>.

Miller, S. (1995) 'Intentions, Ends and Joint Actions', *Philosophical Papers*, 24(1), pp. 51–66.

Miller, S. and Mäkelä, P. (2005) 'The Collectivist Approach to Collective Moral Responsibility', *Metaphilosophy*, 36(5), pp. 634–651. Available at: <https://doi.org/10.1111/j.1467-9973.2005.00397.x>.

Mole, C. (2024) 'Attention', in E.N. Zalta and U. Nodelman (eds) *The Stanford Encyclopedia of Philosophy*. Fall 2024. Metaphysics Research Lab, Stanford University. Available at: <https://plato.stanford.edu/archives/fall2024/entries/attention/> (Accessed: 20 September 2024).

Morgan, B. and Yeung, K. (2007) *An Introduction to Law and Regulation: Text and Materials*. Cambridge, UK; New York: Cambridge University Press (The law in context series).

Nguyen, C.T. (2020) 'Layers of Agency', in *Games: Agency as Art*. Oxford: Oxford University Press, pp. 52–73. Available at: <https://doi.org/10.1093/oso/9780190052089.003.0003>.

Nguyen, C.T. (2022a) 'Transparency is Surveillance', *Philosophy and Phenomenological Research*, 105(2), pp. 331–361. Available at: <https://doi.org/10.1111/phpr.12823>.

Nguyen, C.T. (2022b) 'Trust as an Unquestioning Attitude', in T.S. Gendler, J. Hawthorne, and J. Chung (eds) *Oxford Studies in Epistemology*. Oxford: Oxford University Press, pp. 214–244. Available at: <https://doi.org/10.1093/oso/9780192868978.003.0007>.

Nickel, P.J. (2007) 'Trust and Obligation-Ascription', *Ethical Theory and Moral Practice*, 10(3), pp. 309–319. Available at: <https://doi.org/10.1007/s10677-007-9069-3>.

O'Brien, J. (2014) 'Professional Obligation, Ethical Awareness, and Capital Market Regulation', in N. Morris and D. Vines (eds) *Capital Failure: Rebuilding Trust in Financial Services*. Oxford University Press, Available at: <https://doi.org/10.1093/acprof:oso/9780198712220.003.0010>.

O'Neil, C. (2012) 'Lying, Trust, and Gratitude', *Philosophy & Public Affairs*, 40(4), pp. 301–333. Available at: <https://doi.org/10.1111/papa.12003>.

O'Neill, O. (2002a) *Autonomy and Trust in Bioethics*. Cambridge; New York, NY: Cambridge University Press.

O'Neill, O. (2002b) 'The Reith Lectures - A Question of Trust' (2002) *The Reith Lectures*. London: BBC Radio 4. Available at: <https://www.bbc.co.uk/programmes/p00ghvd8> (Accessed: 11 March 2020).

O'Neill, O. (2014) 'Trust, Trustworthiness, and Accountability', in N. Morris and D. Vines (eds) *Capital Failure: Rebuilding Trust in Financial Services*. Oxford University Press, Available at: <https://doi.org/10.1093/acprof:oso/9780198712220.003.0008>.

Pacherie, E. (2011) 'Framing Joint Action', *Review of Philosophy and Psychology*, 2(2), pp. 173–192. Available at: <https://doi.org/10.1007/s13164-011-0052-5>.

Park, C. *et al.* (2020) 'The Effect of Loneliness on Distinct Health Outcomes: A Comprehensive Review and Meta-Analysis', *Psychiatry Research*, 294, p. 113514. Available at: <https://doi.org/10.1016/j.psychres.2020.113514>.

Pasternak, A. (2021) *Responsible Citizens, Irresponsible States: Should Citizens Pay for Their State's Wrongdoings?* New York, NY: Oxford University Press. Available at: <https://doi.org/10.1093/oso/9780197541036.001.0001>.

Pettit, P. (1995) 'The Cunning of Trust', *Philosophy & Public Affairs*, 24(3), pp. 202–225.

Pettit, P. (2004) 'Trust, Reliance and the Internet', *Analyse & Kritik*, 26, pp. 108–121.

Potter, N.N. (2002) *How Can I Be Trusted? A Virtue Theory of Trustworthiness*. Lanham, MD: Rowman & Littlefield (Feminist Constructions).

Pouryousefi, S. and Tallant, J. (2023) 'Empirical and Philosophical Reflections on Trust', *Journal of the American Philosophical Association*, 9(3), pp. 450–470. Available at: <https://doi.org/10.1017/apa.2022.14>.

Pratchett, T. (1993) *Small Gods: A Discworld Novel*. Corgi Edition. London: Corgi Books.

Preston-Roedder, R. (2017) 'Civic Trust', *Philosopher's Imprint*, 17(4), pp. 1–24.

Raz, J. (2002) *Practical Reason and Norms*. Oxford: Oxford University Press.

Rovane, C. (2004) 'What is an Agent?', *Synthese*, 140(1/2), pp. 181–198.

Rovane, C. (2014) 'Group Agency and Individualism', *Erkenntnis*, 79(S9), pp. 1663–1684.

Scanlon, T. (1998) *What We Owe to Each Other*. Cambridge, Mass: Belknap Press of Harvard University Press.

Scheffler, S. (1995) 'Individual Responsibility in a Global Age', *Social Philosophy and Policy*, 12(1), pp. 219–236. Available at: <https://doi.org/10.1017/S0265052500004623>.

- Schilke, O., Reimann, M. and Cook, K.S. (2021) 'Trust in Social Relations', *Annual Review of Sociology*, 47(1), pp. 239–259. Available at: <https://doi.org/10.1146/annurev-soc-082120-082850>.
- Schlosser, M. (2019) 'Agency', in E.N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy*. Winter 2019. Metaphysics Research Lab, Stanford University. Available at: <https://plato.stanford.edu/archives/win2019/entries/agency/> (Accessed: 22 November 2024).
- Searle, J.R. (1990) 'Collective Intentions and Actions', in P.R. Cohen, J. Morgan, and M. Pollack (eds) *Intentions in Communication*. Cambridge, MA: MIT Press, pp. 401–415.
- Searle, J.R. (1995) *The Construction of Social Reality*. London: Lane, The Penguin Press.
- Shaver, R. (2020) 'Promises as invitations to trust', *Philosophical Studies*, 177(6), pp. 1515–1522. Available at: <https://doi.org/10.1007/s11098-019-01271-7>.
- Shklar, J.N. (1984) *Ordinary Vices*. Cambridge, MA; London: The Belknap Press of Harvard University Press.
- Shoemaker, D. (2019) 'Blameworthy but Unblamable: A Paradox of Corporate Responsibility', *Public Policy*, 17(S), pp. 897–917.
- Silver, K. (2019) 'Can a Corporation be Worthy of Moral Consideration?', *Journal of Business Ethics*, 159(1), pp. 253–265. Available at: <https://doi.org/10.1007/s10551-018-3787-4>.
- Simpson, T.W. (2012) 'What Is Trust?', *Pacific Philosophical Quarterly*, 93(4), pp. 550–569. Available at: <https://doi.org/10.1111/j.1468-0114.2012.01438.x>.
- Simpson, T.W. (2017) 'Trust and Evidence', in P. Faulkner and T.W. Simpson (eds) *The Philosophy of Trust*. Oxford: Oxford University Press, pp. 177–194. Available at: <https://doi.org/10.1093/acprof:oso/9780198732549.003.0011>.
- Simpson, T.W. (2023) *Trust: A Philosophical Study*. Oxford: Oxford University Press. Available at: <https://doi.org/10.1093/oso/9780198855866.003.0001>.
- Simpson, T.W. (2024) 'Trust and the Rule of Law', in R. Chang and A. Srinivasan (eds) *Conversations in Philosophy, Law, and Politics*. Oxford: Oxford University Press, pp. 99–118. Available at: <https://doi.org/10.1093/oso/9780198864523.003.0005>.
- Smith, M.N. (2008) 'Terrorism, Shared Rules and Trust', *Journal of Political Philosophy*, 16(2), pp. 201–219. Available at: <https://doi.org/10.1111/j.1467-9760.2006.00273.x>.
- Smith, M.N. (2010) 'Reliance', *Noûs*, 44(1), pp. 135–157.
- Smith, T.H. (2023) 'My Job and Its Requirements', in A. Barber and S. Cordell (eds) *The Ethics of Social Roles*. Oxford: Oxford University Press, pp. 236–260. Available at: <https://doi.org/10.1093/oso/9780192843562.003.0012>.
- Sparrow, M.K. (2020) *Fundamentals of Regulatory Design*.
- Strawson, P.F. (1974) *Freedom and Resentment, and Other Essays*. London: Methuen.
- Thomas, L. (1989) 'Trust and Survival: Securing a Vision of the Good Society', *Journal of Social Philosophy*, 20(1 & 2), pp. 34–41.

Thomas, L. (1990) 'Opinion | Next Life, I'll Be White', *The New York Times*, 13 August. Available at: <https://www.nytimes.com/1990/08/13/opinion/next-life-ill-be-white.html> (Accessed: 2 July 2023).

Thomson Reuters (2018) *Thomson Reuters Founders Share Company Limited*. Thomson Reuters. Available at: <https://www.thomsonreuters.com/content/dam/ewp-m/documents/thomsonreuters/en/pdf/corporate-responsibility/thomson-reuters-founders-share-company-limited-1218.pdf> (Accessed: 27 November 2024).

Tombs, S. and Whyte, D. (2015) *The Corporate Criminal: Why Corporations Must Be Abolished*. London; New York, NY: Routledge (Key Ideas in Criminology Series).

Topham, G. *et al.* (2015) 'The Volkswagen Emissions Scandal Explained', *The Guardian*, 23 September. Available at: <https://www.theguardian.com/business/ng-interactive/2015/sep/23/volkswagen-emissions-scandal-explained-diesel-cars> (Accessed: 2 December 2024).

Tuomela, R. (2005) 'We-Intentions Revisited', *Philosophical Studies*, 125(3), pp. 327–369. Available at: <https://doi.org/10.1007/s11098-005-7781-1>.

Tuomela, R. (2006) 'Joint Intention, We-Mode and I-Mode', *Midwest Studies in Philosophy*, 30(1), pp. 35–58. Available at: <https://doi.org/10.1111/j.1475-4975.2006.00127.x>.

Velleman, J.D. (1997) 'How to Share an Intention', *Philosophy and Phenomenological Research*, 57(1), pp. 29–50. Available at: <https://doi.org/10.2307/2953776>.

Walsh, D. *et al.* (2022) 'Bearing the Burden of Austerity: How Do Changing Mortality Rates in the Uk Compare Between Men and Women?', *Journal of Epidemiology and Community Health*, 76(12), pp. 1027–1033. Available at: <https://doi.org/10.1136/jech-2022-219645>.

Warren, M. (2017) 'Trust and Democracy', in E.M. Uslaner (ed.) *The Oxford Handbook of Social and Political Trust*. Oxford: Oxford University Press, pp. 75–94. Available at: <https://doi.org/10.1093/oxfordhb/9780190274801.013.5>.

Weber, M. (2015) 'Bureaucracy', in T. Waters and D. Waters (eds. & trans.) *Weber's Rationalism and Modern Society: New Translations on Politics, Bureaucracy, and Social Stratification*. New York, NY: Palgrave Macmillan. Available at: <http://ebookcentral.proquest.com/lib/leeds/detail.action?docID=2057734> (Accessed: 8 December 2024).

Westlund, A.C. (2009) 'Deciding Together', *Philosopher's Imprint*, 9(10). Available at: <http://hdl.handle.net/2027/spo.3521354.0009.010> (Accessed: 2 December 2024).

Williams, B. (2002) *Truth & Truthfulness: An Essay in Genealogy*. Princeton, NJ: Princeton University Press.

Wringe, B. (2016) 'Collective Obligations: Their Existence, Their Explanatory Power, and Their Supervenience on the Obligations of Individuals', *European Journal of Philosophy*, 24(2), pp. 472–497. Available at: <https://doi.org/10.1111/ejop.12076>.