The desperation threshold: a model to explain decisions in poverty

Du Buisson de Courson, B.

1 Introduction

> "Money is better than poverty, if only for financial reasons."
>
> – *Woody Allen, Without Feathers (1975)*

1.1 Poverty is unpleasant

Explaining a joke is bad taste: it generally ruins it. But allow me, just this once, to lead into my dissertation by explaining why the quote above is, objectively, funny. First, it seems to make a new and profound statement, but just states something that is so obvious that it is never verbalized: poverty is unpleasant. Second, the quote is such a blatant understatement. Poverty is not merely a matter of being unable to afford organic vegetables. According to the Cambridge Dictionary, poverty is, instead, 'the state of not having enough money to buy basic things'. The 'basic' adjective is eloquent. Metaphorically, 'basic things' are things that form the basis of a normal life. When they can not be bought, the metaphor implies, the structure risks collapsing.

Intuitively, we thus see poverty as a point where the lack of money has catastrophic effects on life. Implicitly, we think of poverty as a 'cliff', a point where losing little resources causes an abrupt decline in wellbeing. This intuition of a discontinuity is reflected in common language: low-income people are said to 'live on the edge', to be in 'tight' situations, to 'fall into poverty', or to be 'trapped in poverty'. These phrases suggest that poverty is not a matter of degree, but a distinct category. This is also reflected in measures: while wealth is often measured by counting money (e.g. through the median income), poverty is measured as a rate, by counting people below a certain 'poverty line', or those who cannot satisfy certain needs (Blasco, 2023).

In turn, when people with little economic resources behave in a certain way – and they do behave in striking ways (Pepper & Nettle, 2017) – academics often attribute it to a fear of the poverty cliff. After all, if being poor is that awful, it is reasonable to expect that people adapt their decisions to avoid it. More precisely, decisions typical of people on low incomes are often explained as attempts to reduce the risk of not meeting certain basic needs. Examples are countless and span all fields of social sciences. Let us consider a few classical examples. When subsistence populations are observed to help each other and share, as in Kropotkin's classical ethnography of Siberian populations during winters, it is seen as a form of 'collective insurance' against starvation (Kropotkin, 1902). When people in situations of poverty refuse to specialize in one profession, and instead keep one foot in agriculture and one in industry and earn a lower income as a result, this is seen as a "risk-spreading" technique to be protected from bad harvests (A. V. Banerjee & Duflo, 2007). When they show reluctance to innovation, for instance in agriculture, this is explained by the lack of a buffer: if the innovation does not work, one could not bear the losses (Schultz, 1964). When people save, it is seen as a way to build a 'buffer', to avoid catastrophic consequences in case something bad happens in the future (Collins et al., 2009). This principle is even used to explain institutional arrangements in subsistence societies: Scott (1977)

claims that such societies develop a 'subsistence ethic': "elites had a positive moral obligation to provide for the maintenance needs of their subjects in time of dearth" (p33).

Yet, not all people show carefulness or solidarity in situations of poverty. Excessive borrowing, gambling, migration, revolts, prostitution and crime are all more frequent in low-income populations (Dobbie & Skiba, 2013; Gurr, 2015; Hsieh & Pugh, 1993; Monroe, 2005; Wardle et al., 2014). And the thing is, we spontaneously understand these behaviours as 'desperate' actions, that stem from the very same cliff: a person needs a minimal amount of resources, and if he lacks resources, he might have no choice but to take risks to get them. Examples of such reasoning pervade not only in common conversations, but also in social sciences. The most classical example is Merton's 'strain theory' (Merton, 1938): when someone can not achieve culturally valued goals (e.g. wealth and status) with legitimate means, they resort to 'innovation', that is, to alternative, potentially illegitimate, means, such as crime. It is often used in criminology, with a clearer link with poverty: in an ethnography of offenders, Rossmo & Summers (2022) observed that "[the person] needs a specific amount of money such that anything less has limited value; for example...a debt with an impending payment due or [the requirement to] come up with the rent in order to avoid eviction. For these individuals, it is a matter of all or nothing" (p.7). Put otherwise, the offenders have too much to gain in the crime. Another, closely related way to frame this reasoning is to say that in dire poverty, people have too little to lose: they are miserable anyway, their situation could not be much worse, so they might as well try something. The phrase 'little to lose' is often used to justify risky behaviours in situations of poverty: for instance, Holdsworth et al. (2020) explains gambling among homeless populations this way: "With little to lose and a desire to change life circumstances, gambling may appear to be desirable for the hope it provides to win money." (p.14).

We face a double paradox. First, at an empirical level, how can people be both risk takers and risk avoiders in situations of poverty? Second, at a theoretical level: how can the concept of a poverty cliff be used to explain both risk taking and risk avoidance? This ambivalence has been pointed out before by Banerjee in an essay entitled "The two poverties": he observed that "There are at least two distinct and, prima facie, inconsistent views of poverty in these models" (p. 59), conceptualising poverty respectively as having too much to lose and too little to lose. This does not just question the coherence of the concept, it also questions its scientific value: if any empirical pattern can be accommodated to the idea of 'poverty as a cliff', then it is unfalsifiable and therefore unscientific.

As we will see, the paradoxes disappear when the situation is properly formalised. In Fig. 1.1, I translate the representation of poverty that I have verbally sketched into a 'utility function', that represents the satisfaction of the person depending on its level of resources. Importantly, the function features a cliff, that I henceforth call the 'desperation threshold' (DT): at some point, losing little resources has catastrophic consequences on satisfaction, as basic needs can no longer be met. Below the cliff, utility levels out: people have 'little left to lose'. The consequences are visible on the figure. Above the DT, you have more to lose than to gain, so you avoid risk. This makes sense: you do not dance on a tightrope. Below the threshold, you have more to gain than to lose. This also makes sense: if you are falling from the tightrope, you have to try something, get out of your way. Thus, the paradoxes disappear: poverty can imply both 'too much to lose' and 'little to lose' situations. Depending on whether they barely make ends meet or not, people in situations of poverty should thus either take, or avoid, risks. The central aim of my PhD is to formalise this 'desperation threshold model'

(DTM), looking at which social phenomena it can help explain, both at the individual and the population level, and test empirically whether it is a reasonable explanation.

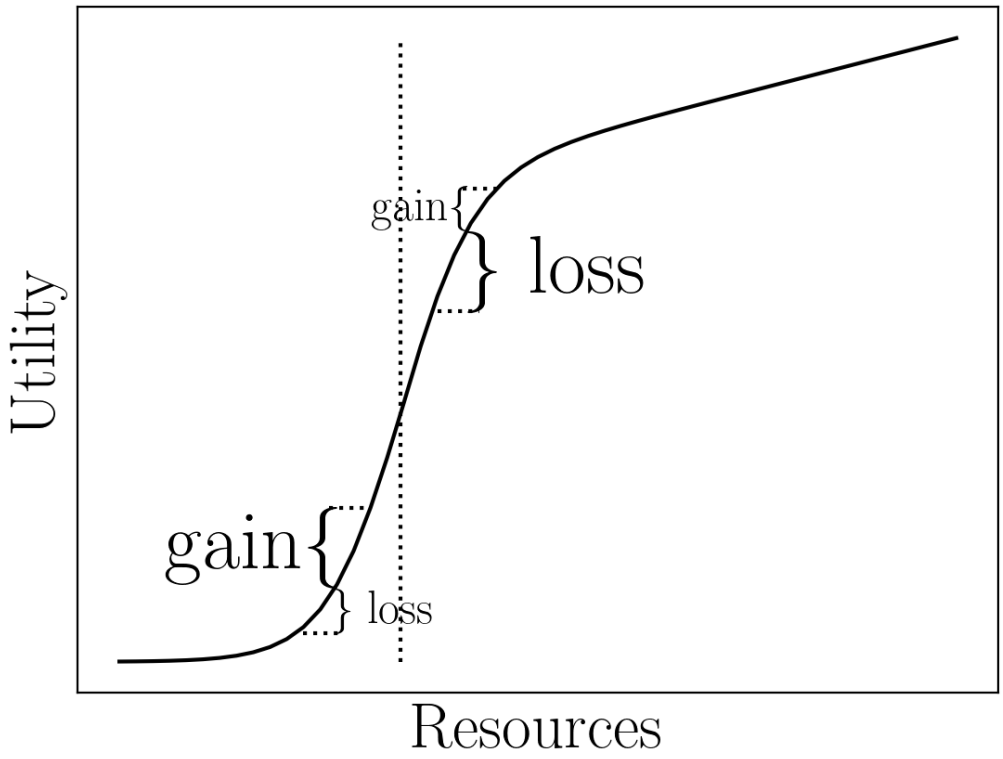

Figure 1.1: Graphical representation of the desperation threshold model. Here, I use the following function: $U(x) = \frac{1}{1+e^{-x}} + x\mathbb{1}_{x>0}$. The first hand is a sigmoid, representing the urge to meet basic needs. The second hand creates a linear increase above the threshold, ensuring that the agent also 'wants' more resources once his basic needs are secured. The dotted line is placed at the threshold location (here, 0), where one has precisely as much to gain as to lose.

I coined the phrase 'desperation threshold' with Daniel Nettle in 2019, before I started my PhD. At the time, we were not sure why. It just sounded right. We stuck with it, and until now, we have had no regrets: it still corresponds to what we want to say, and people spontaneously understand it, even without grasping

the underlying mathematics. After six years of scratching my head on this idea, I finally understood why 'desperation' was the right word. 'Desperation' comes from the latin dēspērāre, which means literally 'losing hope'. This meaning is still present in the older word 'despair', which shares the same etymology, and that the Cambridge dictionary defines it as "the feeling that there is no hope and that you can do nothing to improve a difficult or worrying situation". It is usually associated with passivity, and with a sense of resignation. 'Desperation', however, came to refer to opposite behaviours: the same dictionary defines it as "the feeling that you have when you are in such a bad situation that you are willing to take risks in order to change it".

Why would one try something if one is losing hope? Our model, I hope, illustrates why. Like many words (e.g. awesome or astonished), desperation has likely undergone 'semantic bleaching': it lost some of its intensity over time, and started to be used to describe still extremely bad situations, but where there still is a glimmer of hope. Likewise, the DT marks the point at which meeting basic needs becomes unlikely. It is a point where one is not just worried, but actually expecting a catastrophe as a plausible outcome. Resignation results in a sure catastrophe, while taking a risk maintains a possibility of making it. If the situation further degrades, then the catastrophe becomes surer, and one needs to take even bigger risks to keep hope. At some point, no hope is left, and 'desperation' turns into 'despair'. Thus, our model not only reconciles Banerjee's "two poverties", it also reconciles the modern sense of desperation with its etymological sense: economic misery can bring about caution (above the DT), risk taking (below the DT) and resignation (well below the DT).

In this introduction, I state the goals and the means of my dissertation. In the first section, I discuss the value of formal modeling in social sciences. In the second section, I present and defend my particular use of modeling, insisting on my commitment to rational choice and methodological individualism. In the third section, I trace the intellectual history of the desperation threshold across the social sciences. Finally, I give an outline of the chapters to come.

1.2 Why model

If I have leaned heavily on colloquial phrases and etymologies, it is because I am convinced that the core predictions of the desperation threshold model are already present in the collective unconscious. It is not exactly breaking news that extreme situations give rise to extreme behaviours. So, why shall we bother mathematicising the whole thing? Why translate something so intuitive into a language of Greek letters and obscure symbols?

Many people – including academics – are confused about the role of formal models in social sciences (Tiokhin, 2021). In the papers that form this dissertation, I did not have the space to make general epistemological points to avoid confusion. I hope this introduction can serve as the right place to situate the work epistemologically — so that readers understand how to approach the chapters that follow.

1.2.1 "Models do not investigate nature"

A naive reader might think that having a model will help us make predictions, just like weather forecasts are more accurate than folk sayings. At parties, I used to boastfully present myself as a PhD student in criminology who models crime in situations of poverty. Most people spontaneously assumed that I was building tools to

predict who will commit a crime, based on what is left on people's bank accounts. At the risk of being even more distressing (or, depending on your political awareness, slightly reassured), the models that I will present do not make quantitative predictions. Or, at least, not useful ones.

In fact, except in Chapter 4, no data has been directly involved in my dissertation. The models are built on assumptions: that people have particular options, that they make decisions by optimizing certain criteria, and that they interact in specific patterns. As they are quite abstract, the assumptions are not validated by empirical data: they are minimal axioms that simply sound reasonable and intellectually exciting.

The results of a model depend entirely on the assumptions. This is the source of two frequent critiques. First, a model is tautological. Second, you can obtain any result by tweaking the assumptions. Both critiques are entirely accurate. Unless all the assumptions are true – which is not realistic in social sciences – the results of a model do not teach us anything about the real world. As Kokko (2007) puts it "models do not investigate nature" (p. 7). This point is fundamental and must be hammered: most readers wrongly understand formal models as proving something about reality. For instance, the first chapter finds that in our model, inequality produces more crime. About a hundred papers have cited it, most of them to claim that inequality does produce crime. I wish they heard me: models are not about the real world, but the world of ideas. They only teach us that a set of conclusions logically follow from a set of assumptions. Of course

1.2.2 We are all modelers

So, why model, then? In response to this very question, J. Epstein (2008) responds that every social scientist with even the slightest theoretical ambition is a modeler. "Anyone who ventures a projection, or imagines how a social dynamic—an epidemic, war, or migration—would unfold is running some model" (p. 1). The choice is not whether to model, but whether to state the model verbally or formally. Verbal models are simpler to elaborate and to communicate, but they have critical limitations that formal models help overcome.

1.2.3 Verbal models are vague

The first flaw of verbal models is their vagueness: they leave many details unspoken. At best, their assumptions are implicit. At worst, assumptions are ambiguous and open to multiple interpretations. This gives the model a dangerous room for maneuver, and it can then be stretched to fit a wide range of empirical patterns. The problem is not just practical, but fundamental: social sciences, and psychology in particular, currently go through a 'theory crisis' (Eronen & Bringmann, 2021; Muthukrishna & Henrich, 2019; Scheel et al., 2021), as it becomes increasingly evident that there is little or no cumulative progress in theoretical work. Two pitfalls strike me. First, many projects pretend to test theories, when they are actually only based on vague frameworks. Daniel Nettle calls this tendency "Calling It Theory for Effect": to exaggerate their contributions, authors often dress their confused thoughts as a theory. This plague does not spare criminology (Ducate et al., 2024; Walters & Mandracchia, 2017): the most influential theories in the field have been deemed "too imprecise and discursive to yield meaningful tests of theories" (Ducate et al., 2024, p. 5). It has recently been proposed that authors might make their verbal models vague on purpose, due to the incentive structure in academia. Slippery models are moving targets: they are hard to refute, as they can be

interpreted in so many ways, and shape-shift to match data.

In contrast, formal models are written in formal languages – mathematics or computer code – that, by design, do not tolerate ambiguity. Unlike humans, computers cannot interpret vague instructions; they require explicit, fully specified instructions to function. Thus, formal modeling is, first and foremost, an exercise in making theories entirely explicit. It is also an exercise in intellectual humility: it often makes it visible that the assumptions are simplistic, the scope narrow, and the model itself a caricature. Yet this transparency is necessary for cumulative progress in theoretical social science.

Formal models are also a better guide for empirical social science. Testing a verbal model is a treacherous task: its predictions can be bent in many ways, and it is often hard to see whether the data support the model. In contrast, the predictions made by formal models make can not be taken seriously quantitatively, but they have the merit of being explicit. In Chapter 4, I use the DTM to derive clear and falsifiable predictions on the effect of resources on risk taking.

1.2.4 Verbal models are fallible

The second flaw of verbal models is that they can be actively misleading. Some theories 'sound' true, but turn out to be logically inconsistent upon further inspection. Chapter 3 gives an example: many authors argue that in situations of high need ('I need to eat today'), people should prefer to obtain resources sooner. Through a cumbersome mathematical model, we show that this is not necessarily true, and that one needs further conditions to justify it.

Formal logic is the only way to prove that a set of conclusions follow from a set of conclusions. As Kokko (2007) puts it, formal models "investigate the validity of our own thinking, i.e. whether the logic behind an argument is correct." (p. 7). In some cases, the result is quite obvious, and verbal models are in practice a good replacement for formal models. In other cases, we face problems that are too complex to be grappled by simple intuition. During my PhD, I often found that my intuitions were wrong—or that the models produced results I hadn't anticipated. In such cases, modeling serves as a cognitive aid, a "telescope of the mind, multiplying human powers of analysis and insight just as a telescope does our powers of vision" (Bouchaud, 2009, p. 32). This is the way I feel about mathematics: it is simply empowering.

1.2.5 Poverty and risk taking need formalisation

Risk taking in poverty actually exemplifies the risk of vague theories, and the merits of specification. As discussed earlier, the intuition behind the desperation threshold can be used to predict both more risk taking and less risk taking in situations of poverty. Put otherwise, the verbal model is flexible, because it is underspecified.

There is a central ambiguity in the verbal argument: even if we agree on a utility cliff at some point, what happens below that cliff? Does utility continue to fall, or does it level off? Metaphorically, is poverty a utility pit, or an abyss? I have realised that the first option – assuming that there is a 'rock bottom' on utility – is

necessary to capture the intuition of 'desperation'. This assumption represents the idea that at some point, people have 'little to lose', and is the only way to generate risk taking below the threshold.

Once we formally specify what we mean by a desperation threshold (Fig. 1), surprising conclusions emerge. Whereas verbal models yield monotonic predictions—poverty either increases or decreases risk taking—the specified model reveals a non-monotonic relationship. Poverty has no consistent directional effect, but a polarizing one: individuals avoid risk just above the threshold and embrace it below. This result offers a guide to test our model empirically: unlike most previous studies, we do not fit the data with linear function, but with a broken-stick pattern (Chapter 4).

1.3 What is a good model?

To sum up, a formal model is an explicit and internally consistent representation of a candidate explanation for a phenomenon. This definition applies generally to formal modeling in social sciences. But there are countless ways to explain a phenomenon that do not necessarily contradict each other. Take crime rates in deprived neighbourhoods, the simplest explanation would be: people commit crime because they are in poverty. A slightly more elaborate would be: poverty is frustrating, so people commit crime. These are explanations, just not particularly satisfying ones.

What kind of explanation does this dissertation aim to provide? I am not a philosopher, and the project did not begin with a clear epistemological framework. That said, I believe I have implicitly followed principles that have been articulated by philosophers. What follows is a kind of retrospective epistemology, a post-hoc attempt to clarify and justify the approach I have taken.

1.3.1 Abductive reasoning

Scientific reasoning is often described as either deductive or inductive. Deductive reasoning starts with assumptions and derives predictions from them. This is the case, historically, of neoclassical economics. From axioms of rationality and market equilibrium, economists derive all sorts of consequences, from the effect of a minimum wage to the effect of trade tariffs. In inductive reasoning, scientists accumulate findings, and a general law emerges. This has, for a large part, been the case of experimental psychology.

Both approaches have important limitations in social sciences. As humans are quite messy, induction tends to never produce a clear theory. Discussing the state of psychological research, Muthukrishna & Henrich (2019) cite Poincaré to illustrate the problem: 'an accumulation of facts is no more a science than a heap of stones is a house' (p. 4). Deductive social sciences have also reached a dead end. For instance, neoclassical economics yielded elegant predictions on every aspect of the economy, but little predictive power. Many predictions are out of touch with reality: at equilibrium, no firm makes any profit and there is no involuntary unemployment. In the last decades, the field has pivoted toward applied research, at the expense of theory (Backhouse & Cherrier, 2017). More broadly, I believe that a purely deductive social science is unattainable, as we lack proper axioms about human behaviour.

Formally, the models in this dissertation are deductive: I prove that the DT and other assumptions have several implications. However, my thinking process is not a deduction, but rather an 'abduction', also called 'inference to the best explanation' (P. Lipton, 2017). In each chapter, I start from macro-level empirical patterns, and argue that the DT is capable of explaining them. In other words, I do not claim that the DT was a priori an obvious assumption. Otherwise, neoclassical economists would likely have thought of it. My claim

is that the DT is able to account for many social phenomena, which makes it a posteriori plausible. Abduction is not unusual in theoretical social sciences (Muthukrishna & Henrich, 2019), but it is rarely acknowledged. Some of the results of my dissertation were also made deductively, in particular, Chapter 4 where I investigate the consequences of the DT on time discounting. In those cases, I followed the thread of the DT and looked for interesting consequences. However, I also framed those cases as abductions, starting from known empirical phenomena. This choice reflects my conviction that the value of the model is not predictive, but explanatory: the DT is only interesting insofar it helps us to understand existing findings.

Abduction has a strong shortcoming: it does not show that the model is the true explanation for a phenomenon. Again, models do not investigate nature. Rather, abductive models humbly show that an assumption is a 'candidate explanation' (J. M. Epstein, 2012) for these phenomena: the DT is sufficient to explain crime in deprived populations, but whether it is the mechanism at play remains to be proven. This is just a step in the scientific process, but a crucial, and rarely made step.

1.3.2 State-dependent rational choice

I now turn to concrete theoretical frameworks that have guided my work. The central one is rational choice: in all four chapters, I have aimed to explain social phenomena as the result of cost-benefit analysis. The DT itself is formulated as a utility function. I also use game theoretical arguments in Chapter 2 and 4.

Rational choice theory is, of course, widely discredited empirically (Kahneman, 2011). However, I would argue that its bad name is due to its use in a deductive fashion, with the aim of making quantitative predictions. Here, I instead use it in an abductive fashion (see above). Instead of assuming rationality, I start from empirical phenomena, and wonder what rationality can explain. In Boudon's terms, I look for 'good reasons' behind people's actions.

Rational choice has also been criticised for its flexibility and lack of falsifiability: for a very wide range of behaviours, one can build an aptly shaped utility function accounting for the findings. In these cases, rational choice theory is almost tautological. The desperation threshold, however, is not an arbitrarily shaped utility function. It aligns with the intuition of 'basic needs', and resonates with experiences of people in situations of poverty. It also has plausible evolutionary roots. Rubin & Paul (1979) argue that there is a "minimum income needed in order to support a mate and offspring" (p. 593). In other words, reproductive value is plausibly very small below some level of resources, and can of course not be negative.

I should nonetheless stress that my use of rational choice theory is not a commitment to its realism. Nor do I claim that rationality accounts for most social phenomena. However, when applicable, I find rational accounts particularly satisfying: they are 'self-sufficient' (Boudon, 2003). Most explanations beg further questions – for instance, emotional explanations beg the question of why the emotion occurs in such contexts. In contrast, rational explanations can bring the scientific inquiry to a close: people behave this way because they have good reasons to do so, period.

I believe rational choice theory is especially valuable for analysing decision making in contexts of poverty. Since Malthus (1798), a long-standing tradition in the social sciences has portrayed the poor as irresponsible. Today, one of the most influential paradigms — 'scarcity theory' — attributes behaviour under poverty to

what Vohs (2013) calls the 'Poor's Poor Mental Power.' According to scarcity theory, people in poverty take risks or act impulsively because they fail to take into account the negative consequences of their actions, as poverty disrupts their decision making (Shah et al., 2012). Despite its limited empirical support (O'Donnell et al., 2021), scarcity theory has had immense influence.

More broadly, behaviours in situations of poverty are almost always framed as bad decisions. This framing is particularly evident in discussions of risk taking. When people in poverty take risks, they are accused of endangering themselves and those around them. When they avoid risk, they are said to refuse profitable opportunities, thus perpetuating their poverty (Haushofer & Fehr, 2014).

In this dissertation, I take the opposite track: following the work of Daly & Wilson (2001) and of Pepper & Nettle (2017), I analyse such behaviours in poverty as attempts to manage in a bad situation. Putting it more politically, I try to see the world through the eyes of a person in poverty, and to take their decisions as seriously as we take the decisions of the well-off. My interest in rationality is not absolute, but state-dependent: we are not interested in what behaviour is optimal in general, but only in how the context – level of resource and the social environment – shapes the optimal decision. However, I do not study the reverse causality – the role of behaviour in generating poverty. This reverse causality clearly exists, but is beyond the scope of the PhD.

1.3.3 Methodological individualism

A good explanation is not a simple paraphrase. Saying that poverty causes crime is just restating an empirical phenomenon, without adding much light. This amounts to saying that a car advances because fuel makes it go forward. An explanation is more satisfying when it 'opens the black box', that is, proposes a mechanism by which the micro-level entities of the system generate the macro-level ones. This 'reductionism' is not always feasible: to explain global conflicts, molecules are of little help. It is also sometimes not desirable: to explain the existence of poverty, the individual scale might be a distraction (Brady, 2023).

In the case of the phenomena that my dissertation studies, however, I believe individual decisions cannot be ignored. Groups do not have an independent will, and in cases like violent crime rates, it is not plausible that a centralised decision making is happening. Of course, the mapping between individual decisions and macro-level outcomes is not trivial. Structural factors (economic inequality, the penal system) play a role, and strategic interactions between individuals generate non-trivial effects. In Chapter 2, I use an agent-based model to investigate such effects. I found an interesting and unexpected hysteresis phenomenon: the level of violence in a group depends not only on the current economic conditions, but also on the past. Put otherwise, the group has properties of its own – but those properties are maintained through individual decisions.

1.3.4 Toy models

Last, the models I will present belong to the class of 'toy models', also called 'conceptual' (Kokko, 2007) or 'explanatory' (J. Epstein, 2008) models. Toy models are models that do not aim at being as realistic as possible, but to be as simple as possible while remaining interesting, non-trivial and explanatory. Kokko (2007) defends this approach by drawing a comparison with maps. The first reason that maps are simplified versions of reality is that reality is too complex. Just like the fractal-shaped coast of Brittany is impossible to cartography, we could never measure everything that is relevant in social reality, as too many variables (culture, landscape, economies, family structures...) are relevant. And even if we had data, the models could not

handle the complexity of computations, as all the variables interact with one another. The second reason is that a realistic map would be terribly unhelpful to orient: 'if lost in a forest, [a hiker] would not become any wiser by looking at a too vastly detailed map than by staring at the original forest' (p. 4). What the hiker needs is a highly simplified version of the forest, that highlights the paths and the contour lines. The same way, a perfect social simulation like The Matrix would be less useful to a sociologist than Schelling (1971) model of segregation (see below).

Toy models shall not be used to make useful quantitative predictions. At best, they make qualitative predictions: for instance, in deprived neighbourhoods, we expect more crime than in richer ones. However, just like 'scale models' are good at showing children how catapults work, toy models are very good at illuminating the core dynamics of a proposed explanation. Ideally, they should find the minimal conditions for a phenomenon to emerge. This justifies our unapologetic use of rational choice theory: integrating bounded rationality or biases would vastly increase the realism of our models, but in our cases, such assumptions were not needed to obtain the results.

1.3.5 What for?

If the models that I present here make no quantitative predictions, are they just elegant mind games? Well, not entirely. In most cases, explaining a phenomenon is more intellectually satisfying than being able to predict the consequences of an intervention, but less useful. In some cases, however, an explanatory model can be more useful than an empirically-informed model. This argument is usually known as the 'Lucas critique'. Lucas (1976) argued that statistical models that rely solely on historical correlation often fail to predict the effect of structural changes, because they ignore how people adapt their expectations and behavior in response to new rules. An explanatory model, however, can be used to approach unprecedented situations or new rules. Understanding the system helps to generalize to situations that do not exist in the empirical record. Harford (2015) gives the following example: "Fort Knox has never been robbed, so we can save money by sacking the guards." This is of course absurd, hence the conclusion: "You can't look just at the empirical data, you need also to think about incentives" (p. 182). A model that opens the black box and proposes a rudimentary cost-benefits analysis for the guards' behaviour would – correctly – predict that sacking the guards would result in a Fort Knox robbery. Lucas applied this reasoning to macro-economic models: providing micro-foundations does not help to better fit historical data, but it can plausibly help to better predict the effects of policy changes.

I believe the Lucas critique applies to the DTM in several places. Proposing rational motivations for risk taking in situations of poverty does not help to predict the crime rate in Paris next year. However, the flexibility and the generality of my models can make them suitable instruments to approach the effects of radical public policies. For instance, crime rates are often observed to be insensitive to the severity of punishment (Nagin, 2013), and the DTM proposes an explanation for it (de Courson & Nettle, 2021). However, no one has ever observed the effect of a complete abolition of the criminal justice system. The DTM can make a qualitative prediction: below a certain threshold of severity, crime rates would explode. More importantly, the DTM can inform us on the behavioural consequences of an unprecedentedly protective welfare, that would guarantee that every citizen can meet its basic needs. I discuss the consequences of the DTM in that regard at the end of each chapter.

1.4 The desperation threshold: an old, simple, but radical idea

With a clearer view of the aim and value of this scientific enterprise, I now turn back to introducing concretely what this dissertation consists of. The aim of this dissertation is to propose a new way to explain certain behaviours in situations of poverty and social outcomes in deprived or unequal populations. The model I propose is simple and intuitive: in poverty, people try first and foremost to make ends meet. This idea is so simple that I am of course not the first to propose it. Actually, I found instances in many disciplines – instances that in most cases ignored each other, even though they proposed models that are formally equivalent. Most of the instances, however, failed to grasp the full consequences and radicality of the model. In particular, they rarely thought about the two aspects of the threshold at once, that is, the fact that it produces risk aversion above and risk taking below.

As I mentioned in the first page, a threshold is implicitly present in many verbal models of decisions in poverty, for instance in Merton's strain theory. To our knowledge, the oldest explicit presentation comes from the Soviet economist Chayanov (1926), who considered that "in a natural economy, human economic activity is dominated by the requirement of satisfying the needs of each production unit" (p. 4). Chayanov put this principle at the center of his analysis of the "peasant economy", and studied the many ways that peasants minimized the risk of not meeting their needs: for instance by growing crops that one can directly consume, by working insanely hard or by selling cattle in bad years. However, Chayanov was not interested in what happens when ruin is likely. Interestingly, his principle came to be known as the 'safety first' principle, which connotes prudence and not risk taking.

This principle was later formally modeled in several disciplines, independently. Let us review, by disciplines, how models analogous to the DTM have been proposed.

1.4.1 Agricultural economics

The first, historically, is agricultural economics. Roumasset (1971) built a formal model to explain the choice of technique of peasants in the Philippines. He assumes that agents have 'lexicographic preferences': they first try to make sure that they are above a 'disaster level', and only if the probability of falling below is very small, optimise average production. When one gets closer to the threshold, agents switch to safer actions, but then 'reswitch' to risk taking when they dip below the threshold. This result seems not to have been understood at the time: in reviews of Roumasset's later book, Lundahl (1977) concludes that "This result runs contrary to all conventional wisdom" and "casts some doubt on Roumasset's methods" (p. 392), while M. Lipton (1977) argued that "it is not clear how it withstands more usual, and in farming more plausible, assumptions" (p. 826).

The idea was simultaneously developed by Kunreuther (1971) who also found this result puzzling: "It is thus conceivable that extremely rich farmers and poor farmers will follow a similar cropping pattern [i.e. risk-taking] but for entirely different reasons" (p. 7). Masson (1974) proposed a model closer to the DTM, representing 'disaster level' as a jump in the utility function. Again, the model produced the same result, but Masson discarded it as largely irrelevant empirically: "[being below the threshold] should be rare because these imply that even to remain at the current income level would be disastrous. Most people with discontinuities to the right of the zero point would in fact not remain in farming and thus would not show up in the

sample" (p. 562). Overall, the literature on agricultural economics has focused almost entirely on risk aversion above the threshold, and scarcely on risk seeking below. For instance, McCloskey (1976) used this principle to explain scattering in English open fields as risk minimization, but did not explore the fact that when times are dire enough, the model prediction reverses.

1.4.2 Development economics

After a hiatus, the principle was rediscovered in development economics. It was first explored by A. V. Banerjee & Newman (1994), who pointed out that if poverty is best represented as having 'nothing to lose' – formally, if individuals in poverty are close to a lower bound in their utility function – then they should take risks. In this case, an individual who received a loan would have no reason to repay it as long as he has a chance to get away with it. No one would therefore be willing to loan money to individuals in poverty, and this 'credit constraint' might generate poverty traps.

Lybbert & Barrett (2007) took the opposite point of view. They showed that if a poverty trap exists, then individuals around it should alter their risk taking. The poverty trap can be modelled as a point at which wealth dynamics 'bifurcate': above, individuals tend to accumulate wealth, whereas below, they tend to fall into poverty. Individuals should therefore avoid risks just above the bifurcation level: any risk taken increases one's chance of ending up in poverty. Below the threshold, they should on the contrary take risks, to get a chance to break out of poverty. While the justifications are different, the result emerges the same way as in our model: around the poverty trap, the long-term utility of wealth is sigmoid-shaped (Lybbert & Barrett, 2007). Unlike Banerjee, they made a connection with the agriculture economics literature, presenting their model as "conceptually indistinguishable from the safety-first, lexicographic preferences popular in the 1970s" (p. 414).

1.4.3 Evolutionary biology and psychology

Meanwhile, Caraco et al. (1980) and Stephens (1981) independently introduced the idea of a 'starvation threshold' in behavioural ecology. This resulted in a large empirical literature in animal behaviour, known under the name 'energy budget rule', or 'risk-sensitive foraging'. This literature stayed very separate, even though the model is equivalent: in both cases, a sigmoid-shaped utility function generates risk-taking below a critical level of resources, and risk aversion above it. The principle then diffused to evolutionary approaches of human behaviour, in anthropology (Kuznar, 2001; Mace & Houston, 1989; Winterhalder et al., 1999) and psychology (Mishra & Lalumière, 2010; Nettle, 2009; Pietras & Hackenberg, 2001; Rode et al., 1999). These fields interpreted the threshold in a looser way, not in terms of starvation, but of 'subsistence' in anthropology, and 'need' in psychology. Both in biology and psychology, however, authors focused on the right side of the threshold: they merely predicted – and tested – an increase in risk taking in dire situations.

It is remarkable that different fields have independently converged on analogous ideas, sharing a desperation threshold as a common denominator. It highlights how general the mechanism is, and how he can apply across various domains. Though, it is unfortunate that these models have rarely crossed disciplinary boundaries, and that the same idea had to be rediscovered multiple times. It is also striking that most authors – with the exception of Travis & Lybbert – had interest in only one side of the threshold: they looked at behaviour

either above or below the threshold, but rarely both at once. With this simplification, the DTM makes a linear prediction that is easily testable: poverty either increases or decreases risk taking. This simplification is very problematic: it makes the DTM non-falsifiable, since both a positive and a negative correlation between resources and risk taking can be predicted.

Despite the many intellectual forerunners, I believe however that most authors have not grasped the radicality of the desperation threshold. The DTM predicts that poverty has a complex effect on risk taking: people take less risk up to a point, then much more risk. The DTM predicts a 'behavioural switch' at the threshold: a little loss in resources can trigger a massive change in behaviour. At the population level, it predicts a polarization: a deprived population will be a mixture of people avoiding risk, and people taking risks. Chapter 4 clarifies these predictions and presents an empirical test on survey data.

1.5 Dissertation outline

The four chapters of my thesis are connected by the desperation threshold. They are part of a larger project, some parts of which are not included in this dissertation, as I took part in them before starting my PhD. The first (de Courson & Nettle, 2021) is a modeling paper, where we showed that the DT could explain "Why do inequality and deprivation produce high crime and low trust?". The second, led by Setayesh Radkani, was a test of these predictions in experimental games (Radkani et al., 2023). We show that when faced with an artificial threshold, participants broadly conform to the model's prediction.

1.5.1 Chapter 2: Why is violence high and persistent in deprived communities? A formal model

The second chapter in this dissertation stems from these two projects. It extends the previous model, allowing for the possibility to use violence to deter exploitation. I show that under these conditions, the DT can account not only for the effect of poverty on property crime, but also on violent crime. The effect of the DT is indirect: some individuals are desperate and therefore take risks, which triggers a need for protection among the rest of the population. The model also reproduces two important stylised facts about violence. First, the fact that homicide rates are so variable over time and space. Second, the model features hysteresis: violence can persist even if the economic conditions triggering it have disappeared – which can explain violence 'neighbourhood effects'.

This chapter is co-authored with Willem Frankenhuis, Daniel Nettle and Jean-Louis van Gelder. It has been published in journal Proceedings of the Royal Society B. The full citation is:

de Courson, B., Frankenhuis, W. E., Nettle, D., & Van Gelder, J. L. (2023). Why is violence high and persistent in deprived communities? A formal model. Proceedings of the Royal Society B, 290(1993), 20222095.

1.5.2 Chapter 3: Why does poverty increase time discounting? Present needs and uncertain future

In this chapter, I extend the desperation threshold model to the study of time discounting. I argue that current models of time discounting in situations of poverty are not sufficient to explain empirical findings, and that the verbal theory that one should prioritize present needs in situations of urgent needs has not been appropriately modeled. I present four different scenarios of the desperation threshold, varying the consequences of the threshold on future utility. My model predicts high discounting around the threshold, and high patience at intermediate levels of resources. In contrast with the risk-centered DTM, the effect is predicted to occur on both sides of the DT. I show that, unlike existing accounts, our explanation does not depend on assuming a future improvement, but predicts a U-shaped effect of future expectations.

This chapter is co-authored with Willem Frankenhuis and Daniel Nettle.

1.5.3 Chapter 4: Poverty is associated with both risk avoidance and risk taking: empirical evidence for the desperation threshold model from the UK and France

This chapter provides an empirical test of the desperation threshold model. I derive and preregister distinctive predictions of the model on the relationship between poverty and risk taking, and test them against data from the UK and France. I obtain the predicted V-shape with our subjective resource measure, but not the objective one. In line with the model, I observe that risk taking is unambiguously polarized among people with low income: variance in risk taking is higher, and both extreme risk avoidance and extreme risk taking are more frequent. I cast out alternative explanations, by showing that income is not correlated with coherence in the risk task, and that there is no polarization on the similarly phrased time discounting task.

This chapter is co-authored with Willem Frankenhuis and Daniel Nettle. It has been published in journal Proceedings of the Royal Society B. The full citation is:

de Courson, B., Frankenhuis, W. E., & Nettle, D. (2025). Poverty is associated with both risk avoidance and risk taking: empirical evidence for the desperation threshold model from the UK and France. Proceedings B, 292(2040), 20242071.

1.5.4 Chapter 5: Explaining the paradoxical effects of poverty on decision making: The Desperation Threshold Model

In this chapter, I expose the desperation threshold model in its most general form and highlight the minimal necessary assumptions. I trace back the different versions of this model that have been previously proposed in economics, biology and social sciences, and that have ignored each other until recently. I review a wide range of evidence from different methods and disciplines, to evaluate the model validity. I discuss the consequences of the DTM at the population level, to explain the effect of inequality, welfare states and deterrence. I spell out the critical open issues of the DTM: whether the threshold is objective or subjective, absolute or relative, and whether it is falsifiable. I conclude by setting an agenda for DTM research in the future.

This chapter is co-authored with Willem Frankenhuis and Daniel Nettle. Daniel Nettle and I have contributed equally. It is forthcoming *Behavioral and Brain Sciences*. The full citation is:

de Courson, B., Frankenhuis, W. & Nettle, D. (2026). Explaining the paradoxical effects of poverty on decision making: The Desperation Threshold Model. *Behavioral and Brain Sciences*.