Chapter 3
Principles and State of the Art
of Agent-Based Migration Modelling

Martin Hinsch and Jakub Bijak

Migration as an individual behaviour as well as a macro-level phenomenon happens as part of hugely complex social systems. Understanding migration and its consequences therefore necessitates adopting a careful analytical approach using appropriate tools, such as agent-based models. Still, any model can only be specific to the question it attempts to answer. This chapter provides a general discussion of the key tenets related to modelling complex systems, followed by a review of the current state of the art in the simulation modelling of migration. The subsequent focus of the discussion on the key principles for modelling migration processes, and the context in which they occur, allows for identifying the main knowledge gaps in the existing approaches and for providing practical advice for modellers. In this chapter, we also introduce a model of migration route formation, which is subsequently used as a running example throughout this book.

3.1 The Role of Models in Studying Complex Systems

Before focusing specifically on modelling human migration, it might be helpful to briefly discuss the role that models can play in analysing complex social phenomena in general. In a wider sense, models can have various purposes (Edmonds et al., 2019; Epstein, 2008); however, here we are specifically interested in the application of models to the study of complex systems. Such systems, that is, systems of many components with non-linear interactions, are notoriously difficult to analyse. Even under best experimental conditions, emergent effects can make it nearly impossible to deduce causal relationships between the behaviour and interactions of the components and the global behaviour of the system (Johnson, 2010). This issue is greatly exacerbated in those systems that are not amenable to experimentation under controlled conditions because they can neither be easily replicated nor manipulated, such as for instance large-scale weather, a species' evolutionary history, or most

medium- to large-scale social systems. In these cases, modelling can be an extremely useful – and sometimes the only – way to understand the system in question.

3.1.1 What Can a Model Do?

As argued in Chap. 2, whether a model is constructed by following inductive or abductive principles or indeed a mixture of both, and whether it is a computer simulation or a mathematical model, at its heart, it ends up being a deduction engine. It is a tool to – rigorously and automatically – infer the consequences of a set of assumptions, thereby augmenting the limited capacity of human reasoning (Godfrey-Smith, 2009; Johnson, 2010). At the most general level, we can distinguish two epistemologically distinct ways in which such a tool can be used in the context of studying complex systems: proof of causality and extrapolation.

Proof of Causality. Understanding causality in complex systems can be challenging since the links between micro- and macro-behaviour or between assumptions and dynamics tend to be opaque. A model can be used in this situation to infer specific chains of causality. By modelling a set of micro-processes or assumptions we can demonstrate – rigorously, assuming no technical mistakes have been made – which behaviour they produce.

   The ability of agent-based models to link the micro- and macro-level processes and phenomena can be used to directly validate or disprove the logical consistency of a pre-existing hypothesis of the form *'(macro-level) phenomenon X is caused by (micro-level) mechanism Y'*. Alternatively, by iterating over several different (micro-level) mechanisms, the (minimum) set of assumptions required to produce a specific behaviour can be discovered (see Grimm et al., 2005; Strevens, 2016; Weisberg, 2007). It is important to note, however, that any such proof of causality can only demonstrate logical consistency of a hypothesis. Empirical research is required to prove the occurrence of the mechanism in question in a given real-world situation.

   In a classical example, the famous Schelling (1971) separation model demonstrates that the observed segregation between population groups in many cities can be caused by relatively minor preferences at the individual level. Similarly, the series of 'SugarScape' models by Epstein and Axtell (1996) show that a number of population-level economic phenomena can be the result of basic interactions between very simple agents.

Extrapolation. For many complex systems, we are interested in their behaviour under conditions that are not directly empirically accessible, such as future behaviour or the reaction to specific changes in circumstances. Assuming that we already have a good understanding of a system, we can use a model to replicate the mechanisms responsible for the aspects of the system we are interested in, and use it to extrapolate the system's behaviour.

   Different types of complex models of the physics of the Earth's atmosphere, for example, can be used to predict changes in local weather over the range of days on

one hand, as well as the development of the global climate in reaction to human influence on the other.

3.1.2   Not 'the Model of', but 'a Model to'

At this point it is important to note that everyday use of language tends to obscure what we really do when building a model. We tend to talk about real world systems in terms of discrete nouns, such as 'the weather', 'this population', or 'international migration'. This has two effects: first, it implies that these are things or objects rather than observable properties of dynamic, complex processes. Second, it suggests that these phenomena are easy to define with clear borders. This leads to a – surprisingly widespread – 'naive theory of modelling' where we have a 'thing' (or an 'object' of modelling) that we can build a canonical, 'best' 'model of', in the same way we can draw an image of an object.

   In reality, however, for both types of inference described above, how we build our model is strictly defined by the problem we use it to solve: either by the set of assumptions and behaviours we attempt to link, or by the specific set of observables we want to extrapolate. That means that for a given empirical 'object' (such as 'the weather'), we might build substantially different models depending on what aspect of that 'object' we are actually interested in. In short, which model we build is determined by the question we ask (Edmonds et al., 2019).

   As an illustration, let us assume that we want to model a specific stretch of river. Things we might possibly be interested in could be – just to pick a few arbitrary examples – the likelihood of flooding in adjacent areas, sustainable levels of fishing or the decay rate of industrial chemicals. We could attempt to build a generic river model that could be used in all three cases, but that would entail vastly more effort than necessary for each of the single cases. To understand flooding risk, for example, population dynamics of the various animal species in the river are irrelevant. Not only that, building unnecessary complexity into the model is in fact actively harmful as it introduces more sources of error (Romanowska, 2015). It is therefore prudent to keep the model as simple as possible. Thus, even though we will in all three cases build a model 'of the river', the overlap between the models will be limited.

3.1.3   Complications

The main foundational task in modelling therefore consists in defining and delineating the system. First, the system needs to be defined horizontally – that is, which part of the world do we consider peripheral and which parts should be part of the model? Second, it needs also to be specified vertically – which details do we consider important? This can be quite challenging as there is fundamentally no

straightforward way to determine which processes are relevant for the model output (Barth et al., 2012; Poile & Safayeni, 2016).

Defining the system can become less of a challenge, as long as we are working in the context of a proof-of-causality modelling effort, since finding which assumptions produce a specific kind of behaviour is precisely the aim of this type of modelling. However, as soon as we intend to use our model to extrapolate system behaviour, trying to include all processes that might affect the dynamics we are interested in, while leaving out those that only unnecessarily complicate the model, becomes a difficult task. As a further complication, we are in practice constrained by various additional factors, such as availability of data, complexity of implementation, and computational and analytical tractability of the simulation (Silverman, 2018). Even with a clear-cut question in mind, designing a suitable model is therefore still as much an art as a science.

3.2 Complex Social Phenomena and Agent-Based Models

Almost all social phenomena – including migration – involve at least two levels of aggregation. At the macroscopic level of the social aggregate – such as a city, social group, region, country or population – we can observe conspicuous patterns or regularities: large numbers of people travel on similar routes, a population separates into distinct political factions, or neighbourhoods in a city are more homogeneous than expected by chance. The mechanisms producing these patterns, however, lie in the interactions between the components of these aggregates – usually individuals, but also groups, institutions, and so on, as well as between the different levels of aggregation.

In order to understand or predict the aggregate patterns we can therefore try to analyse regularities in the behaviour of the aggregate (which can be done with some success, see e.g. Ahmed et al., 2016), or we can try to derive the aggregate behaviour from the behaviour of the components. The latter is the guiding principle behind agent-based modelling/models (ABM): instead of attempting to model the dynamics of a social group as such, the behaviour of the agents making up the group and their interactions are modelled. Group-level phenomena are then expected to emerge naturally from these lower-level mechanisms.

Which modelling paradigm is best suited to a given problem depends to a large degree on the problem itself; however, a few general observations concerning the suitability of ABMs for a given problem can be made. If we want to build an explanatory model, it is immediately clear that agent-based models are a useful – or in many cases the only reasonable – approach. Even for predictive modelling, however, such models have become very popular in the last decades. The advantages and disadvantages of this method have been discussed at length elsewhere (Bryson et al., 2007; Lomnicki, 1999; Peck, 2012; Poile & Safayeni, 2016; Silverman, 2018), but to sum up the most important points: agent-based models are computationally expensive, not easy to implement (well), difficult to parameterise, and are

dependent on arbitrary assumptions. On the other hand, they provide unrivalled flexibility in terms of which mechanisms and assumptions to make part of the model, and describe the system on a level that is more accessible to domain experts and non-modellers than aggregate methods. Most importantly, as soon as interactions or differences between people are assumed to be an essential part of a given system's behaviour, it is often much more straightforward to model these directly and explicitly than to attempt to find aggregate solutions.

3.2.1   Modelling Migration

Migration is a prime example of a complex social phenomenon. It is ubiquitous, as well as being one of the crucial processes driving demographic change. Migration can have substantial impacts in all countries involved in the process – origin, transit and destination – in terms of demography, economy, politics and culture. As a political topic, it has also both been important and contentious. Migration complexity and the agency of migrants are some of the important reasons behind the ineffectiveness of migration policies and the reasons why they bring about unintended consequences (Castles, 2004). In recent years, migration has also found increased relevance and focus in the context of the 'digital revolution' (see e.g. Leurs & Smets, 2018; Sánchez-Querubín & Rogers, 2018).

Given the importance and implications of migration processes, there are strong scientific as well as practical incentives for a better understanding of their complexity. However, as argued in Chap. 2, while there is substantial empirical research on migration, existing theoretical studies are sparser and still largely focused on voluntary, economically motivated migration (Arango, 2000; Massey et al., 1993), with forced and asylum migration lagging behind.

3.2.2   Uncertainty

To make things even more difficult, for most of the research questions relevant to the migration processes we are unable to exclude that differences as well as interactions between individuals are an essential part of the dynamics we are interested in. At least as a starting point, this commits us to agent-based modelling as the default architecture.

In the context of migration modelling, the agent-based methodology presents two major challenges. First, as mentioned earlier, many of the processes involved in our target system are not well defined. We therefore have to be careful to take the uncertainty resulting from this lack of definition into account. This is no easy task for a simple model, but even less so for a complicated agent-based model. Second, agent-based models tend to be computationally expensive, which reduces the range of parameter values that can be tested, and thus ultimately the level of detail of any results, including through the lens of sensitivity analysis.

Moreover, in the context of migration modelling, the situation is further complicated by the fact that empirical data on many processes are quite sparse, if they exist, or of poor quality, as further exemplified in Chap. 4. For example, there may be strong anecdotal or journalistic evidence that smugglers play an important role not only in transporting migrants across the Mediterranean, but also in helping them, for instance, along the Balkan route (Kingsley, 2016). Empirically it is, however, extremely difficult to assess the prevalence of smuggling on these routes since all parties involved – smugglers, migrants, as well as law enforcement agencies – have a vested interest in understating these numbers. As another example, it is obvious that borders and border patrols are an extremely important factor in determining how many migrants arrive in which EU country. While numbers on border apprehensions exist (as for example reported by Frontex, 2018), it is unclear how these numbers map to actual border crossings, in particular taking into account repeat attempts.

As a result, we have very little hard knowledge concerning the underlying migration processes. How likely is it for migrants to be caught at the border? How much do migrants usually know about border controls? How do they use that knowledge in deciding where to go? What do migrants do if they fail to cross a border? In the light of these – and many other – grey areas in describing migration processes in detail, any modelling endeavour has to put a strong emphasis on the different guises of the associated uncertainty. In particular, we need to test not only for numeric uncertainty resulting from the intrinsic stochasticity of the modelled processes, but also for uncertainty resulting from our lack of knowledge of the processes themselves (Poile & Safayeni, 2016). While migration uncertainty and unpredictability is well acknowledged (Bijak, 2010; Castles, 2004; Williams & Baláž, 2011), simulation models still need to incorporate it in a more formal and systematic manner.

3.3 Agent-Based Models of Migration: Introducing the Routes and Rumours Model

For a long time, theoretical migration research has been dominated by statistical or equation-based flow models in the economic tradition (Greenwood, 2005). However, the rise of agent-based modelling in the social sciences in the last decades has left its mark on migration research as well. A full review of migration-related ABM studies is outside the scope of this book (but see for example Klabunde & Willekens, 2016 or McAlpine et al., 2021). Instead, we present a number of key aspects of ABMs in general and migration models in particular, and discuss how they have been approached in the existing literature.

Throughout the book we also present a running example taken from our own modelling efforts related to a model of migrant route formation linked to information spread (Routes and Rumours), different elements of which are described in successive boxes throughout this book. We attempt to clarify the points made in the main text by applying them to our example in turn. Insofar as relevant for this chapter, the documentation of the model can be found in Appendix A.

3.3.1 Research Questions

A key dimension along which to distinguish existing modelling efforts is the purpose for which the respective models have been built. The majority of ABMs of migration are built with a concrete real-world scenario in mind, often with a specific focus on one aspect of the situation: Hailegiorgis et al. (2018) for example aimed to predict how climate change might affect emigration from rural communities (among other aspects) in Ethiopia. They used data specific to that situation (including local geography) for their model. Entwisle et al. (2016) studied the effect of different climate change scenarios on migration in north Thailand using a very detailed model that includes data on local weather patterns and agriculture. Frydenlund et al. (2018) attempted to predict where people displaced by conflict in the Democratic Republic of Congo will migrate to. Their model, among other features, includes local geographical and elevation data.

Many of these very concrete models, however, while being calibrated to a specific situation are meant to provide more general insights. Suleimenova and Groen (2020), for example, modelled the effect of policy decisions on the number of arrivals in refugee camps in South Sudan. Their study was intended to provide direct support to humanitarian efforts in the area. At the same time, it serves as a showcase for a new modelling approach that the authors have developed.

A minority of studies eschew data and specific scenarios, and instead focus on more general theoretical questions. Collins and Frydenlund (2016), for example, investigated the effect of group formation on the travel speed of refugees using a purely theoretical model without any relation to specific real-world situations. In a similar vein, Reichlová (2005) explored the consequences of including safety and social needs in a migration model. Although her study was explicitly motivated by real-world phenomena, the model itself and the question behind it are purely theoretical.

Finally, some models are built without a specific domain question in mind. In these cases, the authors often explore methodological issues or put their model forth as a framework to be used by more applied studies down the line (e.g. Groen, 2016; Lin et al., 2016; Suleimenova et al., 2017). Others simply explore the dynamics arising from a set of assumptions without further reference to real-world phenomena (e.g. Silveira et al., 2006, or Hafızoğlu & Sen, 2012).

The research question underpinning the Routes and Rumours model is defined in Box 3.1.

3.3.2 Space and Topology

Migration is an inherently spatial process. Spatial distance between countries of origin and destination has long been part of macroscopic, so-called gravity models of migration (Greenwood, 2005). Agent-based models, however, make it possible to model spatial aspects of migration much more explicitly.

Box 3.1: Routes and Rumours: Defining the Question

The starting point for the Routes and Rumours model that serves as our running example was the observation, first, that very little theoretical work has been done on the migration journey itself and second, that on that journey what little information migrants have on the local conditions often is based on hearsay from other migrants (Dekker et al., 2018; Wall et al., 2017). From there, we decided to investigate the effect of the availability and transmission of information on the emergence of migration routes. In the first instance, we did not attempt to describe a *specific real-world situation*, however, but wanted to use our model to better understand the *general mechanisms* behind the interaction between information and route formation.

Our model was therefore at this point purely theoretical. Our working hypothesis was that routes – which clearly emerge in the real world – are a result more of self-organisation than optimisation and would therefore be difficult to predict, if prediction was at all possible.

How relevant space is in a given model is determined by the phenomena that a modeller is interested in. In a situation where the net flow of migration between a small number of countries or locations is being investigated, for example, spatial relationships beyond mutual distances is often not taken into account (e.g. Heiland, 2003; Lin et al., 2016, but see e.g. Ahmed et al., 2016 for a non-agent-based model that includes geographic information). There are also some models that include a spatial component but use the relative spatial position of agents solely as a simple representation of social distance (e.g. Klabunde, 2011; Reichlová, 2005).

If actual spatial detail is required, spatial information is usually represented either by a square grid or a graph. While a grid-based approach has the advantage of being straightforward to implement and understand, it does tend to be computationally heavier. Which structure works best, however, ultimately often depends on the requirements of the model and the availability of data.

Fully theoretical models tend to use simple grid-based spatial structure (Silveira et al., 2006; Collins & Frydenlund, 2016; but see Naqvi & Rehm, 2014). Similarly, spatial models built to simulate a specific scenario but without using real-world geographical data (e.g. Sokolowski et al., 2014; Werth & Moss, 2007) will often resort to this solution for convenience. While Hailegiorgis et al. (2018) used detailed rasterised data for their model, most models employing real-world data seem to be built on much simpler graph structures representing networks of, for example, cities (Groen, 2016), districts (Hassani-Mahmooei & Parris, 2012), or even entire countries (Lin et al., 2016).

Finally, in some cases, a completely different approach is used. Naivinit et al. (2010) used a grid structure but with hexagonal instead of square cells. Similarly, although the description of their model is not very detailed, it appears that Frydenlund et al. (2018) did not implement a discretised spatial representation at all, but directly used polygonal data extracted from a geographical information system (GIS). For the Routes and Rumours model, the spatial structure of the simulated world is summarised in Box 3.2.

Box 3.2: Space in the Routes and Rumours Model

Since we intended to study the emergence of migration routes, we had to take spatial structures into account. An initial version of the model showed, however, that a naive grid-based approach was too computationally costly. We settled therefore on representing cities and transport links as vertices and edges of a graph, respectively. Such a representation is sparser than a full grid, but nevertheless reflects the main topological features of the modelled landscape, which are the spatial connections between different settlements through transport links. An example topology is shown in Fig. 3.1 below.
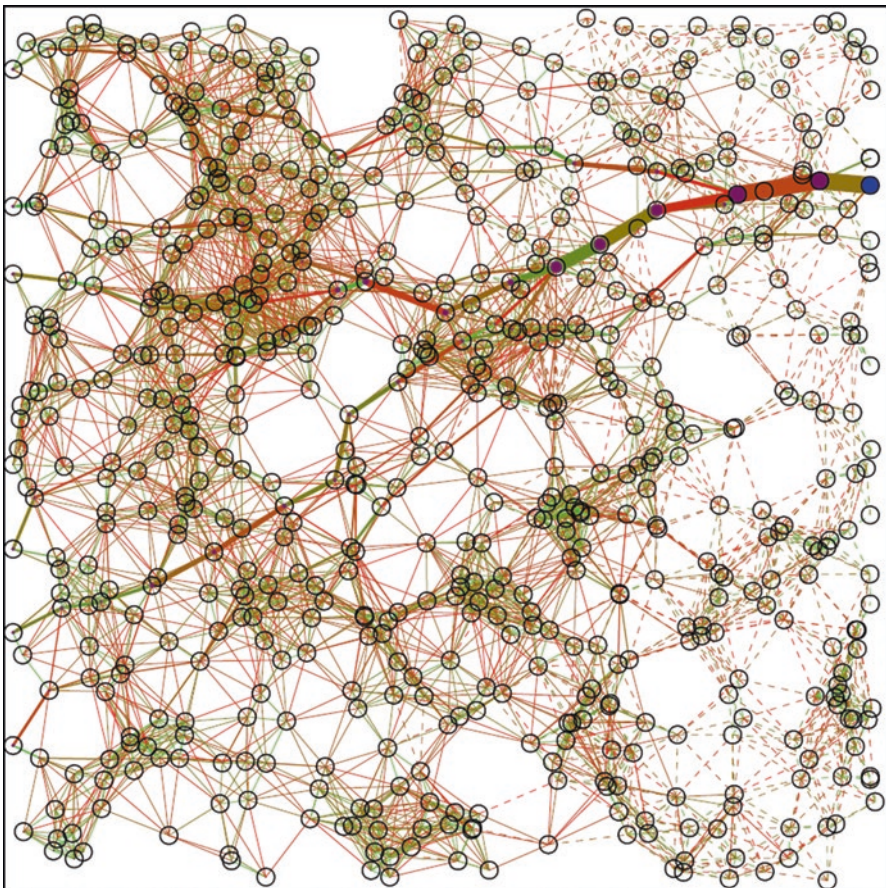

Fig. 3.1 An example topology of the world in the Routes and Rumours model: Settlements are depicted with circles, and links with lines, their thickness corresponding to traffic intensity

3.3.3   Decision-Making Mechanisms

Decision making is an essential part of most models of human migration, or indeed of most other forms of human behaviour (Klabunde & Willekens, 2016). However, which of the many different types of decisions involved a given model makes explicit varies, and is primarily a function of the question the model is used to answer.

Traditionally, modelling studies on migration were primarily invested in understanding under which conditions people decide to migrate and where they will go (Massey et al., 1993). Consequently, the two types of decisions most often included in migration models – agent-based or not – are first, whether to leave and migrate in the first place, and second, which destination to choose when migrating.

In a common type of model, the main focus lies on the conditions in the area or country of origin. In this case, migration is just one of several ways in which individuals can react to changes in local conditions, and the fate of migrants is usually not tracked beyond the decision to leave unless return migration is included (e.g. Entwisle et al., 2016). Examples of such models include Naivinit et al. (2010), Smajgl and Bohensky (2013) and Hailegiorgis et al. (2018).

Unless they are focused on a pair of countries or locations (such as the USA and Mexico, e.g. Klabunde, 2011 and Simon et al., 2016; or East and West Germany, Heiland, 2003), models that simulate the entire migration process usually include the decision to leave as well as a decision where to go. For models of internal migration this is often implemented as a detailed, spatially explicit choice of location (e.g. Frydenlund et al., 2018; Hébert et al., 2018; or Groen et al., 2020). In models of international migration, the decision is usually presented as a choice between different possible countries of destination (e.g. Reichlová, 2005 or Lin et al., 2016).

In addition, a few studies extend the scope of the analysis beyond the simple decisions to leave and where to go. As mentioned before, some models let migrants decide whether to return to their country of origin (e.g. Klabunde, 2014; Simon, 2019). Others include the option to attempt to reach the destination using illegal means (Simon et al., 2016). Finally, there are a few rare modelling studies that focus on entirely different aspects of migration, and consequently model different decisions, such as whether to join a group while travelling (Collins & Frydenlund, 2016). The way decisions are implemented also varies a lot between different studies. In some cases, the decision model is based on an established paradigm such as utility maximisation (e.g. Heiland, 2003; Klabunde, 2011; Silveira et al., 2006). In others, the model is specifically intended as a test case to study the effects of decision making, such as the inclusion of social norms in an economic model (Werth & Moss, 2007), using the theory of motivation (Reichlová, 2005) or the Theory of Planned Behaviour (Klabunde et al., 2015; Smith et al., 2010). Often, however, there does not seem to be a clear justification for the behaviour rules built into the model.

Even in models specifically aimed at prediction within a given real-world scenario, empirical validation of decision rules does not seem to be very common. If it happens, it is usually limited to calibrating the model with regression data linking

migration decisions to individuals' circumstances (e.g. Entwisle et al., 2016; Klabunde, 2014; Smith, 2014). Direct validation of decision processes using, for example, survey-based information (Simon et al., 2016), is rare. For further reading on decision making in migration models we recommend the review by Klabunde and Willekens (2016).

In our case, the way the decisions about the subsequent stages of the journey are being made in the Routes and Rumours model is summarised in Box 3.3.

> **Box 3.3: Decisions in the Routes and Rumours Model**
>
> Since we were primarily interested in the journey itself, we assumed in our running example that individuals have already made the decision to leave their home country, but are not yet at a point where the decision as to which destination country to travel to matters. Instead, we focused on the decisions that determine the route a migrant travels, that is which city to head for next and how to get there.
>
> In principle, agents attempt to reach their destination as quickly as possible. However, in our model the shortest path is not necessarily optimal. The quality of a route is affected by *friction*, an aggregate measure of distance and ease of travel but also the risk a specific leg of the journey entails, as well as the general quality (a stand in for e.g. availability of resources and shelter or permissiveness of local law enforcement) of waypoints. For most components of that decision, we did not have any data to draw on, so we resorted to a simple *ad hoc* model of decision making. For the effect of risk, however, we were able to incorporate data from a psychological survey (see Chap. 6).

3.3.4   Social Interactions and Information Exchange

By definition, macroscopic models have difficulty in capturing the interactions between individuals. This turns out to be a methodological issue once it becomes clear that network effects play an important role in determining the dynamics of international migration (Gurak & Caces, 1992; Massey et al., 1993). To a certain degree, and in some cases, these network effects and other interactions between individuals can be approximated at a macroscopic level (e.g. Ahmed et al., 2016; Massey et al., 1993). However, modelling interactions between individuals is substantially more straightforward in agent-based models, even though there are examples of such models of migration that either do not include any interactions between individuals at all, or only indirect interactions via some global state (e.g. Hébert et al., 2018; Heiland, 2003; Lin et al., 2016).

The simplest forms of interaction take place in movement models where proximity (Frydenlund et al., 2018) or group membership (Collins & Frydenlund, 2016) affect an agent's trajectory. If more complicated interactions are taken into account, then most often this takes the form of social networks that affect an individual's willingness and/or ability to migrate. In the simplest form, this is done by using

space as a proxy for social distance (see Sect. 3.3.2) and defining an individual's 'social network' as all individuals within a specific distance in that space (e.g. Reichlová, 2005; Silveira et al., 2006). More elaborate models explicitly set up links between individuals and/or households (Simon, 2019; Smith et al., 2010; Werth & Moss, 2007), which in some cases are assumed to change over time (e.g. Klabunde, 2011; Barbosa et al., 2013).

The effects that networks are assumed to have on individuals vary and in many cases more than one effect is built into models. Most commonly, networks directly affect individuals' migration decisions either by providing social utility (e.g. Reichlová, 2005; Silveira et al., 2006; Simon, 2019) or social norms (Smith et al., 2010; Barbosa et al., 2013). Another common function is the transmission of information on the risk or benefits of migration (Barbosa et al., 2013; Klabunde, 2011; Simon et al., 2018). Direct economic benefits of networks are only taken into account in a few cases (Klabunde, 2011; Simon, 2019; Werth & Moss, 2007).

Apart from social networks, a few other types of interaction occur in agent-based models of migration. In some studies, agents make their migration decisions without any direct influence from others but interact with them in other ways, such as economically (Naivinit et al., 2010; Naqvi & Rehm, 2014) or by learning (Hailegiorgis et al., 2018), which affects their economic status and thus the likelihood of migrating.

Information and exchange of that information between migrants are the main processes we assumed to be relevant for the emergence of migration routes, and consequently had to be a core part of our model. The information dynamics within the model, as well as the mechanism for the update of agents' beliefs, are summarised in Box 3.4.

3.4   A Note on Model Implementation

A significant hurdle to the broader adoption of agent-based modelling – in particular, in the social sciences – is the specialist skill required to build these kinds of models. There are ways to lower that hurdle, such as specialised software packages (Railsback et al., 2006) or domain-specific languages (discussed in Chap. 7), however all of these come at the cost of reduced flexibility and at times very low efficiency (Reinhardt et al., 2019).

In order to leverage the full potential of agent-based modelling it is therefore often still helpful to implement these models from scratch in a general-purpose language. There is a vast array of languages and methods from which to choose. Traditionally, these fall on a spectrum marked by a trade-off between speed and convenience. At one end, we have fast, yet difficult and unwieldy 'systems-programming' style languages such as C, C++, Fortran or Rust, and at the other much simpler and more convenient, but slow languages such as Python or R. Unfortunately, the fast end of this spectrum tends to be only accessible to experienced programmers, and even then involves trading off convenience and productivity for speed.

Box 3.4: Information Dynamics and Beliefs Update in the Routes and Rumours Model

Agents in our model start out knowing very little about the area they are travelling through, but accumulate knowledge either by exploring locally or by exchanging information with agents they meet or are in contact with. This information is not only necessarily incomplete most of the time, but may also not be accurate. Through exchange it is even possible that incorrect information spreads in the population.

For each property of the environment – say, risk associated with a transport link – an agent has an estimate as well as a confidence value. Collecting information improves the estimate and increases the confidence. During information exchange with other agents, however, confidence can even decrease if both agents have very different opinions.

Our model of information exchange therefore had to fulfil a number of conditions: (a) knowledge can be wrong and/or incomplete, (b) knowledge can be exchanged between individuals, yet, crucially the exchange does not depend on objective, but only on subjective reliability of the information, and (c) agents therefore need an estimate of how certain they are that their information is correct.

Since existing models of belief dynamics do not fulfil all of these criteria, we designed a new (sub-) model of information exchange.

Formally, we used a mass action approach to model the interaction between the certainty $t \in (0, 1)$ and doubt $d = 1 - t$ components of two agents' beliefs. During interactions we assumed that these components interact independently in a way that agents can be convinced (doubt transforming to certainty through the interaction with certainty), converted (certainty of one belief is changed to certainty of a different belief through the interaction with certainty) or confused (certainty is changed to doubt by interacting with certainty if the beliefs differ sufficiently).

For two agents A and B we calculated difference in belief as

$$\delta_v = \frac{|v_A - v_B|}{v_A + v_B}.$$

The new value for doubt is then:

$$d_A' = d_A d_B + \left(1 - c_i\right) d_A t_B + c_u \delta_v t_A t_B,$$

and the new value estimate:

$$v_A' = \frac{t_A d_B v_A + c_i d_A t_B v_B + t_A t_B \left(1 - c_u \delta_v\right)\left(\left(1 - c_e\right) v_A + c_e v_B\right)}{\left(1 - d_A'\right)},$$

where c_i, c_e and c_u are parameters determining the amount of convincing, conversion and confusion.

Julia, a new language developed by a group from MIT (Bezanson et al., 2014), has recently started to challenge this trade-off. It has been designed with a focus on technical computing and the express goal of combining the accessibility of a dynamically typed scripting language like Python or R with the efficiency of a statically typed language like C++ or Rust. A combination of different techniques is used to achieve this goal. In order to keep the language easily accessible, it employs a straightforward syntax (borrowing heavily from MatLab) and dynamic typing with optional type annotations. Runtime efficiency is accomplished by combining strong type inference with just-in-time compilation based on the LLVM platform (Lattner & Adve, 2004). Following a few relatively straightforward guidelines, it is therefore possible to write code in Julia that is nearly as fast as C, C++ or Fortran while being substantially simpler and more readable.

Beyond simplicity and efficiency, however, Julia offers additional benefits. Similar to languages such as R or Python, it comes with interactive execution environments, such as a REPL (read-eval-print loop) and a notebook interface that can greatly speed up prototyping. It also has a powerful macro system built in that has, for example, been used to enable near-mathematical notation for differential equations and computer algebra. Some specific notes related to the Julia implementation are summarised in Box 3.5.

Box 3.5: Specific Notes on Implementation of the Routes and Rumours Model in Julia

We implemented the Routes and Rumours model in Julia from the outset. Beyond the noted combination of simplicity and efficiency, there were a few additional areas where development of the model benefitted substantially from the choice of language:

- Defining and inputting model parameters tends to be cumbersome and error-prone in static languages. Usually the addition of a parameter requires several changes at different places in the code. Using Julia's meta-programming facilities, it was straightforward to have all uses of a model parameter (definition, description, default values, input and output) generated from a single point of definition.
- Similarly, collection and output of data from the model often leads to either inefficient or scattered and fragile code. Using macros, we implemented a simple declarative interface that allows for the definition of data output in one place and mostly separate from the model code.
- As a minor benefit, we were able to use the same language to interactively analyse and graph the data generated by the simulations as for the simulation itself.
- As discussed in Chap. 7, we used Julia's macro system to implement an abstraction of event-based scheduling that is nearly as convenient as a dedicated external domain-specific language.
- Adding dynamically loadable, yet efficient, scenario modules to the model turned out to be close to trivial (see Chap. 8).

3.5   Knowledge Gaps in Existing Migration Models

As we can see, ABMs have become firmly established as a method available for migration modelling. Their application ranges from purely theoretical models to efforts to predict aspects of migration calibrated to a specific real-world situation. A variety of different topics have been tackled such as the effects of climate change on migration via agriculture, the spread of migration experiences through social networks, the formation of groups by travelling migrants, or how the local threat of violence affects numbers of arrivals in refugee camps. Methodologically, these models vary considerably as well, including for example GIS-based spatial representation, decision models based on the theory of planned behaviour, or a spatially explicit ecological model that predicts agricultural yields.

On the other hand, some notable counter-examples notwithstanding, many models in this field still tend to be simple, not at all or poorly calibrated, narrow in focus and littered with *ad hoc* assumptions. In many cases, this is despite best efforts on the part of the authors. Not only is agent-based modelling in general a very 'data hungry' method, but in addition – as further discussed in Chap. 4 and in Sect. 3.2 in this chapter – migration is a phenomenon that is inherently difficult to access empirically.

While macroscopic data on e.g. number of arrivals, countries of origin or demographic composition are sometimes reasonably accessible, microscopic data, in particular on individual decision making, can be nearly impossible to obtain (Klabunde & Willekens, 2016). Consequently, decision making – arguably the most important part of a model concerned with an aspect of human behaviour – is in most models at best calibrated with regression data (but see Simon et al., 2016 for a notable exception) and often neither calibrated, nor in other ways justified (e.g. Hébert et al., 2018).

Unfortunately, even calibration or validation against easier to obtain macroscopic data is not a given. Even some predictive studies restrict themselves to the most basic forms of validation, for example by simply showing model outcomes next to real data (e.g. Groen et al., 2020; Lin et al., 2016; Suleimenova & Groen, 2020). For a purely theoretical model, a lack of empirical reference is not necessarily a cause for concern. But if it is the express goal of a study to be applicable to a concrete real-world situation, then a certain effort towards understanding the amount as well as the causes of uncertainty in the model results should be expected. As some authors, who go to great lengths to include the available data and to calibrate the model against it, demonstrate, high-quality modelling efforts do exist (e.g. Naivinit et al., 2010; Simon et al., 2018; Hailegiorgis et al., 2018).

Another point to note is the relative paucity of theoretical studies attempting to find general mechanisms – as opposed to generating predictions of a specific situation – in the tradition of Schelling (1971) or Epstein and Axtell (1996). Of the existing examples, some stand in the tradition of abstract modelling approaches employed in physics, so that it is difficult to assess the generality of their results (Hafızoğlu & Sen, 2012; Silveira et al., 2006). All these issues additionally reinforce the need for

the model-based research programme, advocated in Chap. 2, going beyond the state of the art in agent-based modelling, and including other approaches and sources of empirical information. As argued before, such efforts should be ideally guided by the principles of classical inductive reasoning.

Generally, however, we can see that formal modelling can open up new areas for migration studies. Many questions remain untouched, providing promising areas for future research. On the whole, as argued above, the primary focus of any modelling exercise should not be aimed at a precise description, explanation or prediction of migration processes, which is an impossible task, but at identifying gaps in data and knowledge. Furthermore, for any given migration system, there is no canonical model. As argued before, the models need to be built for specific purposes, and with particular research questions in mind. Of course, many such questions still have direct practical, policy or scientific relevance. Examples of such questions may include:

- What is the uncertainty of migration across a range of time horizons? What can be a reasonable horizon for attempts at predicting migration, under a reasonable description of uncertainty?
- How are the observed flows of migration likely to be formed, who might be migrating, and who would stay behind? What is the role of historical trends, migrant networks, or other drivers?
- What drives the emergence of migration routes, policies and political impacts of migration? Are migration policies only exogenous variables, or are they endogenous, driven by migration flows?
- More generally, does migration lead to feedback effects, for example through the impacts on societies, policies or markets, and how is it mediated by the level of integration of migrants?
- What are the root causes of migration, and how does migration interact with other aspects of social life? To what extent are various actors (migrants, institutions, intermediaries…) involved?
- How are migration decisions formed and put into action? Do cognitive components dominate, or are emotions highly involved as well? Does it vary between different migration types?

The specific questions, which can be driven by policy or scientific needs, will determine the model architecture and data requirements. Next, we discuss a way of assessing the data requirements of the model through formal analysis.