PDE guidance for cognitive animal movement

Hao Wang and Yurij Salaniw

1Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, AB T6G 2G1, Canada

Abstract

The inclusion of cognitive processes, such as perception, learning and memory, are inevitable in mechanistic animal movement modelling. Cognition is the unique feature that distinguishes animal movement from mere particle movement in chemistry or physics. Hence, it is essential to incorporate such knowledge-based processes into animal movement models. We summarize popular deterministic mathematical models derived from first principles and their rigorous analyses. Mathematical rules of thumb will be provided to judge the model rationality. We will briefly present available mathematical techniques and introduce useful measures of success to compare and contrast the possible outcomes. Throughout the review, we propose future mathematical challenges and development in the form of open questions. This topic is timely and cutting-edge in applied mathematics with many intriguing directions to be explored.

Contents

1 Introduction 2

2 Cognition in Animal Movement Models 3

2.1 Perception ........................................ 5

2.2 Implicit Memory ................................... 6

2.2.1 Static Memory ................................ 7

2.2.2 Dynamic Memory .............................. 8

2.3 Explicit Memory .................................. 13

2.3.1 Discrete Time Delays ......................... 13

2.3.2 Distributed Time Delays ....................... 15

2.4 Short and Long Term Memory .................... 17

2.5 Learning ........................................... 18

3 Biological Insights Through Mathematical Analysis 20

3.1 Rules of Thumb ................................... 21

3.1.1 Prerequisites ................................ 21

3.1.2 Considerations ................................. 21

3.2 Measures of Success in Movement Models .... 24

3.2.1 Measures with Explicit Resources .......... 24

3.2.2 Measures with Population Dynamics ....... 25

3.2.3 Measures with No Population Dynamics or Resources ............ 27

3.3 Comparisons through Measures of Success ............ 27

3.4 Emergence of Patterns ............................ 28

4 Concluding Remarks 29

*The corresponding author. Email: salmaniw@ualberta.ca
1 Introduction

The impact that the cognitive processes of organisms have on their movement is undeniable and ecologically important [12]. Cognitive processes, such as perception and memory, are unique features which distinguish the movement of animals from that of purely random movement, such as Brownian motion; similarly, the process of learning is a particularly unique feature which distinguishes merely directed movement, such as in anisotropic media [39] or aggregation of slime mold [27], from truly novel changes in behaviour and space use, a standard hallmark of learning [63]. Without collecting and encoding information about the landscapes in which organisms live, many quintessential movement patterns, such as site fidelity and optimal foraging, would not manifest [17, 65]. The apparent importance of cognition in animal movement processes warrants the development of mathematical models that incorporate these mechanisms [56]. There already exist a number of reviews discussing the biological importance [12, 30], validity of such inclusions, as well as possible mechanisms behind the acquisition and use of available information [60]. An overarching treatment of knowledge-based movement models from a biological perspective is in preparation [49]. To the best of our knowledge, this review is the first effort to explore mathematical challenges in such considerations comprehensively. It is also worth noting that with the dramatic increase in types and quantities of animal movement data, as well as the significant decrease and increase in cost and computational power, respectively, various statistical methods and stochastic modelling efforts (such as individual based models [10]) have grown in their use and application in this field. Such models will not be considered here, but not for their lack of importance or validity. Reviews of such stochastic models with some focus on cognitive processes already exist [56, 61], as well as at least one proposed “standard protocol” to be used when considering such models in lieu of rigorous analytical techniques [21].

In this review, we aim to focus on the existing deterministic mathematical models which incorporate some form (or multiple forms) of cognitive processes. We seek to address not only how to include certain processes from a mathematical standpoint, but also why particular formulations might correspond to their respective cognitive process. In doing so, we encourage a balance between mathematical rigour and maintaining some form of realism. More precisely, we focus on models found in a partial differential equation setting, either as a scalar equation or a system of coupled equations. This allows one to incorporate and investigate the influence of explicit spatial structure on animal space use without appealing to a simpler ordinary differential equation structure, for example, which improves the level of realism. On the other hand, a partial differential equation setting is more tractable than a stochastic or simulated setting, for example, as the latter offers little means of analysis and therefore does not lend itself to the discovery of ecological laws governing animal space use patterns. This ability to perform analysis provides an additional layer of rigour through a potential for concrete and precise mathematical predictions, allowing one to answer some of the most important questions concerning animal movement and space use patterns. Indeed, explaining the spatial distribution of species using environmental factors has been named one of the top five ranked research fronts in ecology [47].
While a deterministic, continuous-in-time-and-space framework may be very difficult to validate and compare with real world data, it is sometimes possible, see [34, 44, 45] for example. However, even when these models cannot be fully integrated with real world data, they still offer meaningful qualitative insights into outcomes in space use, as well as possible predictive power in the mechanisms included. This point is of particular importance in the area of movement ecology since we cannot (easily) directly observe what is happening in the brain of an organism. Even in cases where we can observe certain brain functions or some other proxy [62, 64, 15], it is a more difficult challenge still to make the connection between these observations and the explicit behaviours of the organism. Instead, we can only make inferences on particular mechanisms of decision making based on the observed outcomes of the movements themselves. Hence, the validity of a proposed model or mechanism may at best correspond to its ability to accurately predict more general, qualitative trends in animal space use as observed in the available real world data. While such comparisons may be lacking in precision, they can still provide meaningful insights and yield substantial motivation for future directions of research.

Consequently, our goals are the following. First, we will introduce some of the existing models as they appear within the framework reaction-diffusion-advection equations or systems. This will provide valuable context for less familiar readers, while also providing motivation for more familiar researchers looking to extend these models in a meaningful way. In doing so, we hope to provide a reasonable amount of detail into the motivation behind the inclusion of certain modelling aspects, and how they connect with real world phenomena in an intuitive way. This will allow us to retain the level of realism we desire through a careful consideration of available mechanisms and how they manifest in the existing literature. Second, we will discuss some of the predictions made and insights gained (in a biological sense) from each model in a more mathematical setting. This closes the figurative loop through a final connection between the mathematical constructions and the biological implications of each. As a secondary goal, we will include some generalizations and new model formulations along the way. In doing so, we hope to compare and contrast these existing models so as to motivate and provide scaffolding to future researchers for further exploration of this exciting and still growing area of research, biologically and mathematically.

A conceptual diagram Figure 1 depicts the interplay between spatial and attribute memory in a home range-bound herbivore, as well as the effect of these memories on the animal’s movement. Animals incorporate information about their environment by exploring and moving within their home range. These stimuli may come in the form of (a) food, (b) predators, (c) conspecifics, or even (d) the center of its home range. The animal’s attraction to (or repulsion from) these stimuli is dictated by its attribute memory, which assigns a quality to each landmark or stimulus the animal remembers. The animal’s cognitive map is represented by a 4 × 4 grid (“spatial memory”) with its current location shaded in grey, while the characteristics of common features (“attribute memory”) are depicted as either “good” (green arrow) or “bad” (red arrow). Each of these may change (i.e. learning) as the animal gains a more accurate representation of the landscape through direct experience or through the transfer of knowledge.

The remainder of this paper is organized as follows. In Section 2 we introduce some of the existing prototypical movement models that feature various forms of perception, memory and learning. Extensions and generalizations of these models, including some new formulations, are also included. In Section 3, we discuss some important rules of thumb any modeller should consider, as well as the biological insights we can gain through study of the models introduced in Section 2. This includes a discussion of possible measures one may use to compare and contrast the space use outcomes found in differing model constructions. To guide researchers moving forward, we provide many open questions throughout the manuscript. We then finish with some concluding remarks and words of advice for the direction of future exploration in Section 4.

## 2 Cognition in Animal Movement Models

To begin, we introduce some of the popular modelling efforts which include some form of cognition. This includes basic aspects of cognition such as perception (the ability to gather information through the senses), but can also include much more complicated phenomena such as memory or learning (both implicit and explicit). The order in which these results are presented is intended to, as best as we are able, start with a more simplistic viewpoint before moving towards more complex formulations. The increase in complexity is primarily in the mathematical formulation and analyses involved, however the complexity of the cognitive
function(s) included naturally escalates contemporaneously, as we will see in the forthcoming sections.

In order to provide a strong foundation, we consider the following scalar advection-diffusion equation prototype under a symmetric dispersal kernel:

$$\frac{\partial u}{\partial t}(x,t) = d\Delta(u(x,t)) - \nabla \cdot (u(x,t)\nabla a(x,t))$$ (2.1)

with a more general form derived in Appendix A.1 using a master equation, Taylor expansion approach. The diffusion rate $d$ relates to transition probabilities due to random movement, while the advection rate $a(x,t)$ relates to the bias in movement based on information at a spatial location $x$ at time $t$. In our current context, the advective potential $a(x,t)$ is the most important quantity we consider as it is where virtually all cognitive processes are currently incorporated into these types of models. Heuristically, $a(x,t)$ can be thought of as the attractivity of a point at a given time, and so we may incorporate varying forms of cognition by adjusting this bias in movement through reasonable biological considerations. From a mechanistic point of view, this can be seen most clearly through the derivation of (2.1) where the quantity $a(x,t)$ is obtained through an exponential distribution of pertinent environmental covariates (see the derivation of space use coefficients found in Appendix A.2 and [29] for further discussion of resource selection functions). This should necessarily include mechanisms briefly introduced so far, such as perception, memory, learning, combinations of each, and their relation to other (external) environmental factors. From a mathematical point of view, a negative gradient term corresponds to attraction (moving up the gradient of $a(x,t)$), while a positive sign corresponds to repulsion (moving down the gradient of $a(x,t)$). Notice that this naturally allows the inclusion of attraction towards favorable regions, but also repulsion away from unfavorable regions. In what follows, we essentially derive the mechanisms by which these factors can be included through modification of the weighting function found to bias animal movement as derived in Appendix A.2.
2.1 Perception

We start our exploration with a scalar equation model which includes an animal’s ability to gather information about its landscape via perception, which will motivate our subsequent model formulations. Perception is the ability to see, hear, or otherwise become aware of something through the senses. In some scenarios, this ability to perceive is assumed to be based on purely local information. This may be an appropriate assumption for describing the movement of cells, for example. Moreover, these purely local gradient models (e.g. Keller-Segel models [23, 37]) provide some initial motivation for how knowledge-based movement models could be constructed. However, for larger organisms, perception should certainly not be limited to purely local cues since it is well known that non-local perceptual cues, such as visual, auditory, olfactory or chemosensory cues, play a vital role in informing animal movement [13, 12, 37]. Furthermore, the clarity by which animals can detect these cues may not be uniform across varying distances, let alone across species, within species, or even within individuals [68]. Consequently, there is substantial motivation to include non-local perceptual capabilities, and this should incorporate both distance and quality of detection.

In this first example, the “something” being detected is an assumed “true” resource density function denoted by \( m(x, t) \). It is assumed that the organism has a finite perceptual range or detection scale (the maximum distance at which landscape elements can be identified), as well as some description of how their perception may be altered with distance. Mathematically, this can be described by an integration over space (i.e. a spatial convolution):

\[
h(x, t) := \int_{\Omega} m(y, t)g(y - x)dy,
\]

which is sometimes referred to as a resource perception function [13], however, we will refer to it more generally as a perception function, as some models include perception of other covariates. Here, the kernel \( g(x - y) \) describes the modifications in the forager’s perception with distance, which we refer to as the perceptual kernel or detection function. Notice that this does not explicitly include how the forager is detecting resources. Instead, it is implicitly assumed that there is an ability to detect the desired resources within the forager’s habitat, and this ability may vary with distance. In [13], the authors consider an unbounded, one dimensional spatial domain \( \Omega = \mathbb{R} \) with the following possible perceptual kernels:

\[
g(x - y) := \begin{cases} 
\frac{1}{2R}, & -R \leq x - y \leq R, \\
0, & \text{otherwise}, 
\end{cases}
\]

\[
g(x - y) := \frac{1}{\sqrt{2\pi R}}e^{-(x-y)^2/2R^2},
\]

\[
g(x - y) := \frac{1}{2R}e^{-|x-y|/R}.
\]

The quantity \( R \geq 0 \) is the perceptual range, which is the standard deviation of the forager’s detection function. These particular forms were chosen since the authors were interested in the transition between platykurtic (no tails) and leptokurtic (fat tails) detection functions, each of which can be obtained from the exponential power distribution. In the first case, the so-called top hat detection function, the agent can perceive resources equally a fixed distance away from its current location and cannot see beyond that fixed distance. The subsequent functions, the Gaussian and exponential detection functions, respectively, allow the agent to perceive nearby resources most clearly and decays monotonically as the distance from the observation location increases. In practice, a perceptual kernel could be any function satisfying the following conditions:

i) \( g(x) \) is symmetric about the origin;

ii) \( \int_{\Omega} gdx = 1 \);

iii) \( \lim_{R \to 0^+} g(x) = \delta(x) \);

iv) \( g(x) \) is non-increasing from the origin.
Condition i) assumes that the animal will perceive features equally across all directions. Condition ii) ensures that different detection functions remain comparable, and so the scaling is chosen such that they each integrate to one. In other words, given a perceptual range \( R \), the mean perceptual range is the same for each detection function. Condition iii) assumes that as the perceptual range \( R \) becomes arbitrarily small, the only information collected is purely local. More precisely, it may be instructive to note that as \( R \to 0^+ \), the perception function \( h(x,t) \) converges to \( m(x,t) \), and thus the information used becomes purely local. Finally, condition iv) assumes that an animal's perception does not improve as distance from the stimulus increases. This condition is mathematically convenient and also biologically reasonable, however it may be worth noting that some scenarios may require condition iv) to be violated. One such example is hyperopia, commonly referred to as farsightedness. All kernels introduced in (2.3) satisfy these conditions.

Starting from the prototypical model (2.1), we may use the heuristic of \( a(x,t) \) being the attractivity of a point so that \( a(x,t) = \gamma h(x,t) \) itself becomes the attractive potential for some \( \gamma > 0 \). Here we reserve \( \gamma \) to denote the strength of attraction up (or down) the potential \( a(x,t) \). The model with foraging based on these perceptual capabilities and a constant rate of diffusion is then given by

\[
\frac{\partial u}{\partial t} = a \frac{\partial^2 u}{\partial x^2} - \gamma \frac{\partial}{\partial x} \left( u h(x,t) \right),
\]

subject to the condition that \( \int_\Omega u(x,t) = 1 \). Notice that this ensures the total population remains fixed, which is reasonable since there is no growth or decay of the population included in the model. In fact, this conservation is a consequence of the correct choice in boundary condition. A more general treatment of common boundary conditions and their implications are discussed in further detail in Section 3.1.2.

In [25], a similar formulation is obtained via a moment closure method to obtain a drift-anisotropic diffusion equation with focus on the 2-dimensional spatial case. This is done from the perspective of a velocity-jump random walk [36], sometimes called a “run-and-tumble” model, where an individual’s movement is determined via a sequence of “run” phases and “turning” phases. The authors compare local and non-local gradient sampling with exclusive focus given to a uniform sampling over a circular region of radius \( R \), which is exactly the top-hat detection function found in line (2.3). Their equation describing the motion of agents is identical to the general form found in Appendix A.1. Similar to [13], the authors then investigate the impact of local vs. non-local sampling under certain model types. Note that, while their discussion includes non-local information biasing movement, it does not include any cognitive mechanism influencing movement, and instead focuses exclusively on the detection of environmental cues.

These works motivate a more general consideration of non-local detection and its influence on animal movement. Given a general potential \( a(x,t) \) biasing an animal’s movement in a domain \( \Omega \), we define a general perception function by

\[
\pi_{a,R}(x,t) := \int_\Omega a(y,t)g(x - y)dy,
\]

where \( g(\cdot) \) is any perceptual kernel satisfying the four conditions introduced above. We use the subscripts \( g \) and \( R \) to denote the dependence on the choice of perceptual kernel and perceptual radius, respectively. Notice also that in order to properly define this quantity in a bounded domain, one must restrict the domain of the perceptual kernel \( g(\cdot) \) to be \( \Omega \), the domain considered in the model itself. By integrating over \( \Omega \) only, one essentially “cuts off” the perceptual kernel so that no additional information is gathering from outside the domain considered. Such a technicality is addressed in [41, Eq. 2.14], for example, where the top-hat kernel is used in a bounded domain \( (0, L) \). For a perceptual radius \( R \) and domain \( L \), the resulting perception function is defined in three intervals: \( (0, R) \), \( (R, L-R) \) and \( (L-R, L) \), which is most convenient for numerical analysis. Analytically, however, such an explicit construction is not necessary in principle.

### 2.2 Implicit Memory

Next, we explore how one may incorporate a rudimentary form of memory into an animal movement model. Memory plays a crucial role in the study of animal movement, yet remains a challenging problem for both biologists and mathematicians as memory itself is a complicated process. In our setting, we adhere to the following definition of memory: the process of encoding, storing and retrieving data within the brain.
Memories can be obtained via genetics (e.g., genetic triggers for migration, or inherited avoidance of predators [13]), or could be obtained through direct experience. In this sense, memory is a higher level brain function than perception, as memory involves a secondary process of storing this information that is obtained through perception. We refer interested readers to [12] for an extensive review of the connections between memory and animal movement. Here, we seek to provide only key details applicable to our modelling formulations.

Due to the complexity of memory and the varying classifications that exist, we focus our attention on two forms of memory which are pertinent to animal movement: spatial memory and attribute memory. Spatial memory encodes the spatial relationships or configurations, comparable to a map, whereas attribute memory encodes the particular characteristics of local environmental features. This can be made more clear through a simple example: spatial memory encodes where food is located, whereas attribute memory encodes the quality and quantity of the food. Of course, these two types often interact with each other, such as the storage of attribute information within a spatial structure. This process is sometimes referred to as cognitive mapping. Originally, there was debate on whether or not cognitive maps exist; presently, the debate has shifted to what form these maps actually take, e.g., euclidean vs topological, see [12] and the references therein. Because we cannot (easily) directly observe these processes within the brain, models that include memory offer an alternate avenue to study these complicated agent-environment interactions. The challenge then becomes how to best model these so-called cognitive maps. Most of the models to be introduced only consider the spatial memory aspect. Indeed, it is more difficult to clearly incorporate the quantity of food in absolute terms since shifting a map by a large positive constant yields no difference in the model (i.e., the constant term vanishes since it appears underneath the gradient). On the other hand, satisfaction measures discussed in Section 2.5 may provide a useful avenue to study the effects of attribute memory and the interplay with spatial memory. In the following sections, we discuss two differing perspectives that include a cognitive map, with the first group featuring static cognitive maps, and the second featuring dynamic cognitive maps.

2.2.1 Static Memory

The most obvious way to include memory into an animal movement model is through a simple change of perspective in model (2.1): define the quantity \( a(x,t) \) to be the cognitive map of the animal. This could define desirable regions as well as regions to avoid, e.g., good resource locations or regions known to have predators. In this sense, model (2.4) could directly be interpreted as a memory model if the function \( m(x,t) \) is assumed to be the quantity being recalled. However, this approach may be viewed as naive as it requires the modeller to assume what the cognitive map actually looks like. In some cases, this may be more or less reasonable. For example, in order to study home range behaviour one may simply take \( a(x,t) = a(x) \) to be the distance from a known home range site, such as a den, i.e., \( a(x) = \gamma \| x-x_0 \| \), where \( x_0 \) is the fixed den site location, \( \gamma > 0 \) is the strength of attraction and \( \| \cdot \| \) is the Euclidean norm. In this way, \( \frac{\partial a}{\partial x} \) becomes a unit vector pointing towards the den site. This is precisely the formulation proposed in [35], which features an alternative derivation from a run-and-tumble perspective and an assumed Von-Mises distribution of turning directions; see also [25]. With a constant rate of diffusion, the model then takes the form

\[
\frac{\partial u}{\partial t} = d \Delta u - \gamma \nabla \cdot \left( \frac{x - x_0}{\| x - x_0 \|} u \right). \tag{2.6}
\]

Of course, in this form the memory mechanism being included is very rudimentary as it does not consider other factors that influence movement, such as presence of predators or resources. However, since our general derivation in Appendix A.2 includes the possibility of incorporating a number of influential factors, a central den site may be considered a single covariate of many. This is an example of spatial memory, as opposed to attribute memory.

Alternatively, given a resource density \( m(x,t) \), it may be reasonable to assume that agents have some knowledge of the landscape relative to some other measure. Two such examples are the following: agents that are aware of the average resource density; or agents that are aware of the per-capita resource density. In both scenarios, it is assumed that the agents have additional knowledge beyond mere resource density.
The first scenario has \( a(x, t) = \frac{m(x, t)}{\overline{m}} \) and can be modelled by

\[
\frac{\partial u}{\partial t} = d \Delta u - \gamma \nabla \left( u \nabla \left( \int_{\Omega} \frac{m(y, t)}{\overline{m}} g(x - y) \, dy \right) \right),
\]

where \( \overline{m}(t) = \frac{1}{|\Omega|} \int_{\Omega} m(y, t) \, dy \) is the average resource density at time \( t \).

The second scenario has \( a(x, t) = \frac{m(x, t)}{u(x, t)} \) and can be modelled by

\[
\frac{\partial u}{\partial t} = d \Delta u - \gamma \nabla \left( u \nabla \left( \int_{\Omega} \frac{m(y, t)}{u(y, t)} g(x - y) \, dy \right) \right).
\]

In the first case, we have a non-local equation that remains linear. On the other hand, the second scenario is more complicated in that it is a non-local and nonlinear equation, i.e. \( a(x, t) \neq a(x, t, u) \).

The application of a static cognitive map, while a first step, is far enough removed from reality that the modelling predictions may not be of notable significance. On the other hand, it may be of interest to consider the combination of a static cognitive map (such as a long term den site) and a dynamic cognitive map that captures other, less static, environmental features. We consider other interacting cognitive maps through short and long term memory in Section 2.4.

Figure 2 depicts a static, continuous-in-space cognitive map. The left panel is a sample static cognitive map with smaller peaks and troughs for high and low resources patches, in addition to a single tall peak denoting a den site. The middle panel provides the direction of movement based on the cognitive map, coming from the vector field generated by the advective potential. The right panel features the cognitive map with perception from the perspective of the forager at location \((x, y) = (1, 1)\) with perceptual radius \( R = 1.5 \) with a top-hat detection function.

### 2.2.2 Dynamic Memory

Alternative to a static cognitive map, one may consider a map to be a dynamically changing quantity, continuously updating as the agent moves throughout its environment. This offers more realism than the static cognitive map, as it is understood that memories are continuously formed and reformed as time passes. On the other hand, a dynamic cognitive map increases the mathematical complexity significantly as the description of movement for a single population may require the inclusion of an additional equation to be solved.

#### Existing Models without Population Dynamics

The first form of a dynamic cognitive map we introduce is similar to the well known group of Keller-Segel models, which describe cell aggregation in response to chemical deposits left behind by the cells themselves. However, instead of following or avoiding chemical deposits, it is assumed that the animals follow or avoid areas of high population density, which is assumed to be part of their cognitive mapping process. In general,
we also include perception as introduced previously so that the attractivity $a(x,t)$ includes perception and the population density itself. The perception function is then $a(x,t) = \gamma \pi_{g,R}$ for $\gamma \in \mathbb{R}$. The equation describing motion becomes
\[
\frac{\partial u}{\partial t} = d \Delta u - \gamma \nabla \cdot (u \nabla \pi_{g,R}).
\] (2.9)

In this case, $\gamma > 0$ corresponds to attraction towards high density areas, whereas $\gamma < 0$ corresponds to repulsion from high density areas. Depending on the context, either could be valid: the first case may correspond to phenomena such as group defense strategies [54], while the second may correspond to avoiding high density areas where resources are expected to be less abundant.

More interestingly, perhaps, are scenarios which include interactions between multiple populations in addition to the other cognitive features. Indeed, one of the primary goals of knowledge based movement models is to better understand the dynamics for many interacting populations. For example, one may generalize model (2.9) to include $n$ interacting populations $u^i$, with the perception function for each group based off of the varying population densities of all other groups:
\[
\frac{\partial u^i}{\partial t} = d_i \Delta u^i - \nabla \cdot \left( u^i \nabla \left( \sum_{j=1}^{n} \gamma_{ij} u^j \pi_{g,R} \right) \right).
\] (2.10)

When the detection function $g(\cdot)$ is chosen to be the top-hat function, (2.10) is precisely the form proposed in scenario 1 of [42] in a bounded, one dimensional spatial domain. Similar to the single species model, the sign of $\gamma_{ij}$ determines whether species $u^i$ is attracted to or repelled from high population densities of population $u^j$. Underlying such models is an implicit assumption that each population $u^i$ shares the same information, and so there must be some biological mechanism which allows agents to share this information between themselves. Alternatively, this may be most applicable to very small organisms so that the density function $u^i$ is an appropriate description of how many individuals are found at a certain location in space. Recently, Potts [43] discussed this formulation in relation to step-selection analysis, as well as models’ applicability and the potential for pattern formation to emerge.

**Open Question:** In what context do solutions exist solving problem (2.10)? In [23], a partial answer is established when the detection kernel $g(\cdot)$ is twice continuously differentiable (e.g. the Gaussian detection function), but few results exist for the top-hat detection function (see Appendix B.1 for a more detailed discussion).

In models (2.9)-(2.10), we have a dynamic cognitive map with perception without the necessity of solving an additional equation. In this sense, these self-aggregation/segregation models are self contained. In the following scenario, we consider a case where the cognitive map is dynamic, and is described explicitly by an additional equation. This may be a more appropriate method to describe the cognitive map of larger organisms that have less dense populations, for example. The first example, appearing in e.g. [31, 42], describes the movement of two or more populations in response to marks on the landscape (e.g. urine, faeces, footprints etc.) left by the other population(s). To this end, we may denote by $p^i = p^i(x,t)$ the density of presence of marks that are foreign to population $u^i$. It is assumed that marks grow linearly with respect to the presence of population $u^j$, $j \neq i$, and decay at a constant rate $\mu$. Then, the evolution of marks foreign to population $u^i$ is given by
\[
\frac{\partial p^i}{\partial t} = \sum_{j \neq i} \alpha_{ij} u^j - \mu p^i,
\]
where $\mu > 0$. Similar to previous examples, $\alpha_{ij} > 0$ indicates that population $u^i$ is attracted to population $j$, while $\alpha_{ij} < 0$ denotes repulsion. Notice that self-interaction is not included, however this is not necessary and one may consider cases $i = j$ to capture this additional influence. Then, setting $\alpha_{ii} = 0$ for each $i = 1, \ldots, n$ recovers the original formulation. Notice also that this form of “memory” is somewhat different than our proposed definition, as their map is not stored within the brain, but is instead stored within the environment itself. Despite this, it provides significant motivation for future models.
With this description of each \( p^i \), we then consider the attractive potential for population \( u^i \) to be \( a(x, t) = \gamma_i p^i(x, t) \) for some advection rate \( \gamma_i \in \mathbb{R} \) and so the full model with perception is given by

\[
\begin{align*}
\frac{\partial u^i}{\partial t} &= d_i \Delta u^i - \gamma_i \nabla \cdot (u^i \nabla p^i) , \\
\frac{\partial p^i}{\partial t} &= \sum_{j=1}^{n} \alpha_{ij} u^j - \mu p^i ,
\end{align*}
\]

(2.11)

for \( i = 1, \ldots, n \), where \( n \) is the number of interacting populations. When the detection function \( g(\cdot) \) is chosen to be the top-hat detection function, (2.11) is scenario 2 proposed in [42]. Familiar readers may notice that if \( n = 1 \), (2.11) is very similar to a Keller-Segel system in the limit as \( R \to 0^+ \). The key difference is the lack of diffusion appearing in the equation for \( p^i \): diffusion is, in this setting, has a regularizing effect, and so an absence of diffusion increases the difficulty of analysis (see Section B.1 for further comments on this point).

In the model described above, the “memory” of the foragers is held within the landscape, and so is not exactly memory as we would commonly consider. In the following model, the actual cognitive map within the foragers’ mind is recorded. To achieve this, the main idea is to track direct encounters between agents from different populations, the areas at which these occur referred to as conflict zones. It is assumed that each population remembers an area where a conflict has occurred, and will be more likely to avoid this area in the future. Should they return to a location and experience no conflict, the cognitive map is updated accordingly. It is also assumed that memory decays at some rate proportional to the time since an event has occurred. This can be viewed as a combination of attribute memory and spatial memory, where the conflict is the attribute recorded at some spatial location where the conflict occurs.

Denote by \( k^i(x, t) \) the spatial memory of conflict zones for population \( u^i(x, t) \). For simplicity, we first consider the case of two interacting populations. From our preliminary assumptions, \( k^i \) should grow with respect to interactions between \( u^i \) and \( u^j \), while it should decay proportionally to \( u^i \) and linearly some rate \( \mu > 0 \). The equation describing the evolution of the spatial cognitive map then takes the form

\[
\frac{\partial k^i}{\partial t} = \rho u^i u^2 - (\mu + \beta u^i)k^i.
\]

The quantity \( \rho \geq 0 \) is the rate at which encounters occur; \( \mu \geq 0 \) is the rate at which memory decay with time; \( \beta \geq 0 \) is the rate at which the conflict zone decays due to agents revisiting the site and experiencing no conflict.

We now take a moment to note an important distinction between this model introduced above and the similar form introduced in [41]. In the cases introduced in this review, we focus on describing the cognitive map as a magnitude, describing important areas versus less important areas in relative terms. On the other hand, some works formulate the cognitive map as a probability density, and so the specific form on the dynamic cognitive map is slightly different. In this case, the equation for \( k^i \) above is instead derived to be

\[
\frac{\partial k^i}{\partial t} = \rho u^i u^2(1 - k^i) - (\mu + \beta u^i)k^i.
\]

In this way, outcomes are treated similar to a coin flip: a location is either part of a conflict zone, \((1 - k^i)\), or it is not, \( k^i \). While this is a subtle difference in interpretation, the overall dynamics appear roughly the same. Since the more popular method is to describe the cognitive map as a magnitude, we focus on these cases instead.

**Open Question:** In this case, the long term population dynamics between differing formulations of \( k^i \) appear roughly the same. Is this always the case? What are the differences in outcomes, if any, between these two interpretations?

Then, this can readily be generalized to \( n \) interacting species:

\[
\frac{\partial k^i}{\partial t} = u^i \sum_{j \neq i} \rho_{ij} u^j - (\mu + \beta u^i)k^i,
\]

where \( \rho_{ij} \) now denotes the rate at which encounters occur between populations \( u^i \) and \( u^j \). Notice also that we necessarily have symmetry so that \( \rho_{ij} = \rho_{ji} \), but this assumption may be invalid if we allow the possibility of one member from population \( i \) seeing a member from population \( j \) without \( j \) seeing \( i \), in which case we
may no longer refer to it as a “conflict”. Then, for each $u^i$, the attractive potential is the cognitive map $k^i$ combined with perception, i.e. $a^i(x,t) = \gamma_i k_{g,R}^i(x,t)$, where $\gamma_i \geq 0$ denotes the rate at which species $i$ moves away from all other populations (notice the sign change of the advective term below, recalling that positive advection denotes repulsion). The full model describing the evolution of $n$ interacting species remembering conflict zones with perception is then given as

$$
\begin{align*}
\frac{\partial u^i}{\partial t} &= d_i \Delta u^i + \gamma_i \nabla \cdot \left( u^i \nabla k_{g,R}^i \right), \\
\frac{\partial k^i}{\partial t} &= u^i \sum_{j=1}^n \rho_{ij} u^j - (\mu + \beta u^i)k^i,
\end{align*}
$$

(2.12)

Notice that we now have a system of $2n$ equations modelling $n$ interacting populations. When the kernel $g(\cdot)$ is taken to be the top-hat detection function, this is scenario 3 proposed in [42] in a bounded, one-dimensional spatial domain.

**Open Question:** In what context do solutions exist solving problems (2.11) and/or (2.12) when the detection kernel is the top-hat function? Different from model (2.10), the additional equation increases the complexity and the results found in [23] are not directly applicable.

In the models presented so far, we have discussed the inclusion of cognitive processes in movement population models that consider only movement (no birth or death of the population) between interacting populations. Sometimes, this is justifiable if one assumes that the movement process occur at a timescale that is much faster than that of a birth/death process. Readers should take caution, however, since such an assumption may invalidate the use of a quasi-steady state approximation. We discuss this point in more detail in Section 3.1.

**Existing Models with Population Dynamics**

Next, a classical consumer-resource model is considered with an additional term biasing the movement of the consumer through a number of differing mechanisms. A slightly more general formulation is considered in [57] which is currently under review. To this end, denote by $u(x,t)$, $v(x,t)$ the consumer and resource, respectively. The case we introduce here is the most straightforward: it is assumed that the consumers have knowledge of where the resources are. With perception, we then take the perception function $a(x,t) = \gamma \pi_{g,R}(x,t)$ for $\gamma > 0$. The consumer-resource model with knowledge and perception of resources is then described by

$$
\begin{align*}
\frac{\partial u}{\partial t} &= D_1 \Delta u - \gamma_1 \nabla \cdot (u \nabla \pi_{g,R}) + \frac{cv}{\alpha + v} - du, \\
\frac{\partial v}{\partial t} &= D_2 \Delta v + rv(1 - v/K) - \frac{\beta uv}{\alpha + v}.
\end{align*}
$$

(2.13)

This perspective is comparable to that found in model (2.10): instead of knowledge of the (current) density of other populations, the consumers have knowledge of the current resource density. Here, $r > 0$ denotes the maximum reproduction rate for the resource, while $K > 0$ is the carrying capacity for the resource. The consumer is assumed to grow according to a Holling type II functional response with the growth rate $\beta > 0$ and the half saturation constant $\alpha > 0$, and decay linearly at rate $d > 0$. The quantity $c > 0$ is a conversion efficiency by the consumer from the resource. Notice that if one takes $\gamma = 0$, the system is reduced to a classical consumer-resource model.

**New Models and Extensions**

In what follows, we discuss some new models that have not yet been investigated in the presented format. Some may require further development, however we still include them to motivate future researches with the advancements currently being made. The first, moderately simple generalization is applied to model (2.12). Some authors argue that a memory process should carry a similar derivation to the movement process itself [20], which in this case suggests that the cognitive map should also feature some rate of diffusion. This leads to the idea of memory *smearing*, which allows for random errors in memory recall. In this way, the diffusion smears the memory component $k^i$ so that the thing being remember is roughly accurate, but not remembered precisely. This simple modification results in the new model

$$
\begin{align*}
\frac{\partial u^i}{\partial t} &= d_i \Delta u^i - \gamma_i \nabla \cdot \left( u^i \nabla k_{g,R}^i \right), \\
\frac{\partial k^i}{\partial t} &= \varepsilon_i \Delta k^i + u^i \sum_{j=1}^n \rho_{ij} u^j - (\mu + \beta u^i)k^i,
\end{align*}
$$

(2.14)
where $0 < \varepsilon_i \ll d_i$. The parameters $\varepsilon_i$ is meant to include this mechanism of memory smearing, or how memories may be altered with respect to distance [4].

**Open Question:** In what context do solutions exist to problem (2.14) when the detection kernel is the top-hat function? Since the cognitive map features a diffusion term, which is known to regularize solutions, this may be easier to address than the no diffusion case.

Alternatively, one may modify the term which dictates areas where conflicts have occurred. In model (2.12), it is assumed that conflicts occur at exactly one point. However, the inclusion of perception may also be relevant in this term, since agents may be able to experience a conflict at a distance. This could result in a conflict (through a “stand-off”) that should also be remembered. Hence, the growth term for $k_1$ should feature the form $\rho \overline{\mu}_j \gamma_q$ in the case of two interacting species. For $n$ interacting species, this takes the form

$$
\begin{aligned}
\frac{\partial u}{\partial t} &= d_i \Delta u - \gamma_i \nabla \cdot \left( u \nabla \overline{\theta}_{g,R} \right), \\
\frac{\partial k}{\partial t} &= \overline{\mu}_j \sum_{j=1}^{n} \rho_{ij} u_{ij} - (\mu + \beta u) k_i.
\end{aligned}
$$

(2.15)

Naturally, these models assume the same perceptual kernel and perceptual radius for each population. This could be a reasonable assumption for studying more uniform populations, such as animals within the same species (e.g. wolf packs), but this may not be an accurate description if the interacting populations are significantly different, e.g. different species all together.

**Open Question:** How does non-local information gathering effect space use outcomes? What are the differences or similarities between local and non-local information gathering? Is non-local information gathering better or worse than local information gathering as found in (2.12)?

Next, we discuss some extensions of the consumer-resource model prototype introduced in (2.13). In general, model (2.13) can be written as

$$
\begin{aligned}
\frac{\partial u}{\partial t} &= D_1 \Delta u - \gamma \nabla \cdot (u \nabla a(x,t)) + \frac{\epsilon_{uv} u v}{\alpha + \gamma} - du, \\
\frac{\partial v}{\partial t} &= D_2 \Delta v + rv(1 - v/K) - \frac{\beta u v}{\alpha + \gamma}.
\end{aligned}
$$

(2.16)

for some attractive potential $a(x,t)$. We can then formulate two additional models based on different cognitive mapping mechanisms which are similar to those introduced in models (2.10), (2.11) and (2.12). In these cases, a third equation modelling the evolution of the cognitive map must then be included.

The first extension, following (2.13), considers a more realistic scenario where the attractive potential $a(x,t)$ is now dynamic in time (i.e. an additional equation will be introduced), and will be described as a cognitive map which we denote by $q(x,t)$. The map $q(x,t)$ is assumed to grow constantly at rate $b > 0$ with respect to resource density, while it decays linearly at a rate $\mu > 0$ due to finite memory capacity. The equation describing the evolution of the cognitive map is then described by

$$
\frac{\partial q}{\partial t} = bv - \mu q,
$$

and in (2.16) we take $a(x,t) = \overline{q}_{g,R}$. Model (2.16) then becomes

$$
\begin{aligned}
\frac{\partial u}{\partial t} &= D_1 \Delta u - \gamma \nabla \cdot (u \nabla \overline{q}_{g,R}) + \frac{\epsilon_{uv} u v}{\alpha + \gamma} - du, \\
\frac{\partial v}{\partial t} &= D_2 \Delta v + rv(1 - v/K) - \frac{\beta u v}{\alpha + \gamma}, \\
\frac{\partial q}{\partial t} &= bv - \mu q.
\end{aligned}
$$

(2.17)

This perspective is comparable to that found in model (2.11): instead of observing marks left on the landscape, consumers are assumed to detect the local resource density and are able to maintain a record of where they have previously found resources.

In the second extension, we consider another version of dynamic memory where the evolution of the cognitive map includes additional mechanisms. The cognitive map $q(x,t)$ is now assumed to grow proportional to the resource density and the density of the consumers at rate $b > 0$. This may be more reasonable than the previous model, as one of the implicit assumptions in these memory-based movement models is that foragers are able to share knowledge between individuals. Hence, a location with high resource density is more likely to be remembered if a larger number of foragers perceive it as such. Similar to the previous
model, it is assumed that the map decays linearly due to finite memory capacity at rate \( \mu \geq 0 \), however it is also assumed that the map can decay further at rate \( \xi \geq 0 \) should the consumer return to an area and find a low resource density. The evolution of the cognitive map \( q(x,t) \) is then given by

\[
\frac{\partial q}{\partial t} = buv - (\mu + \xi u)q.
\]

Taking again \( a(x,t) = \overline{g}_{g,R} \), model (2.16) becomes

\[
\begin{align*}
\frac{\partial u}{\partial t} &= D_1 \Delta u - \gamma \nabla \cdot (u \nabla \overline{g}_{g,R}) + \frac{\varepsilon uv}{\alpha + v} - du, \\
\frac{\partial v}{\partial t} &= D_2 \Delta v + rv(1 - v/K) - \frac{\beta\overline{g}}{\alpha + v}, \\
\frac{\partial q}{\partial t} &= buv - (\mu + \xi u)q.
\end{align*}
\]

(2.18)

This perspective is comparable to that found in model (2.15), and features similar mechanisms found in model (2.17). In this way, it is similar to the memory of direct animal interactions, but the direction of bias is opposite: consumers remember areas they are attracted to, not areas that they seek to avoid.

We conclude this section with a brief but important note concerning attracting and repelling quantities in relation to dynamically changing cognitive maps. Since each of the equations introduced to describe the cognitive map should satisfy a positivity lemma (i.e. they are all a linear differential equation with respect to the cognitive map variable), a single equation is insufficient to include both attracting and repelling quantities. Instead, they can only describe relative amounts of attraction or repulsion with respect to the variable being remembered, but not simultaneously. We explore this point further in Section 2.4.

### 2.3 Explicit Memory

In the models introduced thus far, we have seen the inclusion of perception and some forms of memory, yet these forms are all implicit in that they do not feature an explicit reference to past experiences. A more recent consideration, strongly motivated by the influence of memory on animal movement, is the inclusion of time delays in the advection term. In this way, foragers make explicit reference to their past experiences. This complicates the mathematical analysis significantly when compared to the previous models, which we discuss in Section 3. However, the increase in complexity is not unexpected and these models often yield rather interesting results, both mathematically and ecologically. Nevertheless, in comparison to the models introduced in Sections 2.2.1-2.2.2, the overarching theme remains the same: agents have a bias in their movement, and this bias is included within the advection term. The key difference is that the bias is now driven by information in the past. This is perhaps the most explicit inclusion of memory within deterministic population models, as there is a continual, explicit reference to previous experiences or information. Most of the existing models consider the discrete delay case, which features an explicit reference to information exactly \( \tau > 0 \) time units before the current time, but some also consider the more complicated case of a distributed delay, which includes a reference to all previous times under some weighting function. While the latter case is more logical, the former is often valued for its simplicity while still including an explicit reference to the past.

#### 2.3.1 Discrete Time Delays

**Scalar Equations**

We continue with our theme of motivation from the prototypical diffusion-advection model (2.1). Perhaps the most obvious way to include memory of past occurrences in this framework is to consider the case where the advective potential \( a(x,t) \) is given by the solution itself, \( \tau \) units in the past, i.e. \( a(x,t) = u(x,t - \tau) \), which shall henceforth be denoted by \( u_{\tau} := u(x,t - \tau) \). This is the form derived and investigated in [54], along with the additional inclusion of a birth/death process. The equation describing the evolution of the population \( u(x,t) \) is given by

\[
\frac{\partial u}{\partial t} = d_1 \Delta u - \gamma \nabla \cdot (u \nabla u_{\tau}) + f(u),
\]

(2.19)

where \( d_1 > 0, \gamma \in \mathbb{R} \), and the growth term \( f(u) \) is roughly of logistic type. We again emphasize the relation between the standard diffusion-advection equation (2.1) and the form found in equation (2.19): the bias in
movement is given explicitly by \( u_{\tau}(x, t) \), where \( \tau > 0 \) can be thought of as the \textit{averaged} spatial memory period. Intuitively, it does not make logical sense to assume that organisms make a constant reference to information obtained \textit{exactly} \( \tau \) time units ago. Indeed, higher forms of memory are expected to be more complicated than this in reality. However, it is a useful starting point to consider the effects of an explicit reference to the past, and in this context it makes the most sense to consider the discrete time delay to be an averaged amount of time that a forager will refer back to. In this way, the model favors simplicity. Similar to model (2.9), the advection rate \( \gamma \) may have different sign depending on the situation: \( \gamma < 0 \) represents a movement away from areas of high population density \( \tau \) time units ago, which is a natural phenomenon; on the other hand, \( \gamma > 0 \) represents movement towards high population densities \( \tau \) time units ago, which may be the case for social animals that aggregate for group defense (see the discussion of [54]). Notice also that while a form of memory is included, a perceptual range is not. Indeed, the bias in movement is based purely on \textit{local spatial} information while being non-local in time.

This model has since been extended in three ways, with focus given to adjustments in the growth term: first, a non-local spatial effect is considered in the growth term; second, a non-local temporal effect is considered in the growth term; third, a combination of both of these effects is considered in the growth term.

The first case [58] takes the form

\[
\partial u \over \partial t = D_1 \partial^2 u \over \partial x^2 + D_2 \partial \over \partial x \left(u \frac{\partial u_{\tau}}{\partial x}\right) + f(u, \overline{u}), \tag{2.20}
\]

where \( \overline{u}(t) = |\Omega|^{-1} \int_{\Omega} u(y, t)dy \) is the average population density over the entire domain, and so the growth term also depends on a spatial average of the population over the entire domain. Notice that, in comparison to the averaging with respect to a perceptual kernel, this average may change with respect to time but remains independent of space. The motivation behind such a formulation is due to a recognition that birth/growth/death rates almost certainly depend on population densities at other spatial locations, not purely on the single point where the organism is located. This particular form of non-local (in space) interaction is motivated by [18], which considers this to be the most straightforward way to include a non-local (spatial) interaction effect, though one may consider other forms as well.

The second case, considered in [52], has a comparable form:

\[
\partial u \over \partial t = D_1 \Delta u + D_2 \nabla \cdot (u \nabla u_{\tau}) + f(u, u_{\sigma}), \tag{2.21}
\]

where \( u_{\sigma} = u(x, t - \sigma) \), as previously defined. Biologically speaking, this accounts for a delay in the renewal of resources or the time necessary for animals to reach maturity. Readers should note that while the reaction terms may look similar between models (2.20) and (2.21), their interpretation is distinct and may take significantly different forms.

In the discussion of [52], it is noted that the effects of diffusion and time delays are not independent of each other. That is, the individuals located at \( x \) at a precise time may move to a new location at the present time. As a reasonable revision, the third case considers a combination of the previous two scenarios and includes non-local in space and non-local in time reaction terms simultaneously, as done in [2]. Their evolution equation takes the form

\[
\partial u \over \partial t = \Delta u + d \nabla \cdot (u \nabla u_{\tau}) + \lambda u F(u, \overline{u}_{\sigma}), \tag{2.22}
\]

where \( d \in \mathbb{R} \) and \( \overline{u}_{\sigma}(x, t) = \int_{\Omega} K(x, y)u(y, t - \sigma)dy \) for some reasonably smooth spatial kernel \( K(x, y) \) (see \textbf{H2} in [2]). The spatial scaling is chosen such that \( d \) represents the ratio of the memory-based advection coefficient to the regular diffusion coefficient, and \( \lambda > 0 \) is a scaled constant. The kernel \( K(x, y) \) accounts for the non-local intraspecific competition of the species for either resources or space. Readers should note that this kernel is distinct from the perceptual kernels introduced previously, and so we denote it by \( K \) instead of \( g \). The specific choice in function \( F \) depends greatly on the purpose of application, and so interested readers are directed to [2, 6, 8] for further details. Lastly, notice that (2.22) is somewhat a generalization of the simpler form introduced for model (2.20), where one recovers (2.20) by taking \( K(x, y) = |\Omega|^{-1} \) in (2.22).

\textbf{Systems of Equations}
A recent paper [59] considers a consumer-resource model, similar to model (2.16), where the consumer is assumed to have both random movement as well as memory-based movement through the inclusion of a discrete time delay. The resources are assumed to be plants or “no brainer” animals, so that the prey have no memory-based movement. So far, this is the only model to include an explicit memory mechanism within a system of partial differential equations. The evolution equations describing the consumer and resource in one dimension are given as

\[
\begin{aligned}
\frac{\partial u}{\partial t} &= d_{22} \frac{\partial^2 u}{\partial x^2} - d_{21} \frac{\partial}{\partial x} \left( u \frac{\partial v}{\partial x} \right) + f(u, v), \\
\frac{\partial v}{\partial t} &= d_{11} \frac{\partial^2 v}{\partial x^2} + g(u, v),
\end{aligned}
\]

for \( \tau > 0 \) and \( v_{\tau}(x, t) := v(x, t-\tau) \). The quantity \( v(x, t) \) is the density of the resource, which diffuses randomly at rate \( d_{11} > 0 \), and has growth/decay as \( g(u, v) \), depending on both the resource and the consumer. The quantity \( u(x, t) \) is the density of the consumer, which diffuses randomly at rate \( d_{22} > 0 \), and also grows/decays as \( f(u, v) \), depending on both the resource and consumer. To save confusion, readers should note that the role of \( u \) and \( v \) found here are opposite to that found in the original reference to keep the current work as consistent as possible. The consumer \( u(x, t) \) then moves up the gradient of the resource \( v(x, t) \), \( \tau \) units of time ago, at rate \( d_{21} \geq 0 \). Note that \( d_{21} \) is taken to be non-negative since it is reasonable to assume that the consumer is attracted to the resource. This model as presented is comparable to the resource-consumer model found in (2.16), except the cognitive map is taken to be a memory of the purely local resource density at some fixed time \( \tau \) units before the current time \( t \).

Finally, a competition-diffusion model in [53] features memory based self-diffusion and cross-diffusion. Following a similar motivation from previous discrete delay models, the model takes the following form:

\[
\begin{aligned}
\frac{\partial u}{\partial \tau} &= D_1 \Delta u + D_{11} \nabla \cdot (u \nabla u_{\tau}) + D_{12} \nabla \cdot (u \nabla v_{\tau}) + f(u, v), \\
\frac{\partial v}{\partial \tau} &= D_2 \Delta v + D_{21} \nabla \cdot (v \nabla u_{\tau}) + D_{22} \nabla \cdot (v \nabla v_{\tau}) + g(u, v),
\end{aligned}
\]

for \( \tau > 0 \). Similar to previous works discussed, the authors investigate the impact of memory-based self- and cross-diffusion by carrying out a stability and bifurcation analysis. In its general form, this model includes each form of attraction/repulsion to/from their same group or their competitor, depending on the sign of \( D_{ij} \). Due to the complexity of the analysis involved, some simplifications and specific cases are considered for clarity: a Lotka-Volterra competition is investigated, i.e.

\[
\begin{aligned}
f(u, v) &= u(1 - u - \alpha v), & g(u, v) &= \gamma v(1 - \beta u - v),
\end{aligned}
\]

for \( \alpha, \beta, \gamma > 0 \). The local stability of steady states is then explored in relation to the kinetics system (i.e. no diffusion or advection) with focus on the cases a) \( D_{12} = D_{21} = D_{22} = 0, D_{11} \neq 0 \), or b) \( D_{11} = D_{22} = D_{21} = 0, D_{12} \neq 0 \). Case a) has self-aggregation for species \( u \), \( \tau \) units in the past, while case b) has attraction or repulsion of species \( u \) to/from species \( v \), \( \tau \) units in the past. When they consider the weak competition case, i.e. \( \alpha \beta < 1 \), some interesting insights are obtained. First, if \( u \) is a timid competitor who moves away from the previous locations of its competitor \( (D_{12} > 0) \), then the constant coexistence steady state will be destabilized as the rate \( D_{12} \) increases. On the other hand, if \( u \) is an aggressive competitor who moves towards previous locations of its competitor, then a Hopf bifurcation occurs as the memory period \( \tau \) increases. This indicates that memory-based cross diffusion alters the more classical dynamics of competition-diffusion models in ways that are not found in the existing literature.

Extensions of Discrete Delay Models

As a logical extension of each of the models above, one may incorporate both memory and perception through a modification of the advective potential \( u_{\tau}(x, t) \). Instead, one may consider \( \pi g_{g, R}(x, t) \) for some detection kernel \( g \) and perceptual radius \( R \).

**Open Question:** How does the inclusion of perception into discrete delay models change the space use outcomes? More precisely, how does non-local perception alter the stability of the constant steady states?

### 2.3.2 Distributed Time Delays

In the scenario outlined above, the inclusion of time delays is purely discrete in the sense that the agents remember only exactly \( \tau \) units ago. As previously suggested, it is more realistic to instead consider a
distribution over all previous times, though it becomes more significantly more technical than each of the previous models [55]. Before introducing the model itself, it is instructive to first explore the following spatiotemporal convolution kernel:

$$v_u(x,t) = \mathcal{G} \ast u(x,t) := \int_{-\infty}^{t} \int_{\Omega} G(d_3, x, y, t - s) \mathcal{G}(t - s) u(y, s) dy ds. \quad (2.25)$$

We take a moment here to describe in detail the reasoning behind this form for $v_u(x,t)$. First, it is assumed that there is some biological reason for the inclusion of a time delay, which in our case is motivated by an explicit memory driven movement mechanism and a reference to previous experiences. More precisely, we assume that the cognitive map (in this case, the population density $u(x,t)$) at a time $t$ has contribution from itself at all previous times $s < t$, but not all previous times are equally as important. The choice in temporal kernel $\mathcal{G}(\cdot)$ then describes the weighting given to previous times, similar to how the perceptual kernel $g(\cdot)$ describes the weighting given to distant locations. One may even choose $\mathcal{G} = \delta(t - \tau)$ for some $\tau > 0$, for example, if one assumes that the population density exactly $\tau$ time units ago is most important. In order to determine the contribution from all previous times, we then multiply the density $u(x,\cdot)$ at time $s$ by the weight function at this time, which is $\mathcal{G}(t - s)$ since it is $-t + \tau$ time units ago. Then, the integration over space is meant to capture perceptual influences. Different from previous examples, the perception kernel depends on space as well as time, and $G(d_3, x, y, t)$ is chosen to be the fundamental solution to the heat equation with boundary conditions identical to those prescribed in the proposed movement model. Such a choice is due to mathematical convenience, but at the cost of biological realism.

In [55], the waning of memory due to the passage of time is considered through a Gamma distribution, with two specific cases referred to as a weak or strong kernel, respectively:

$$\mathcal{G}_w(t; \tau) = \tau^{-1} e^{-t/\tau}, \quad \mathcal{G}_s(t; \tau) = t \tau^{-2} e^{-t/\tau}, \quad \tau > 0. \quad (2.26)$$

In application, these two temporal kernels have different interpretations. In the context of memory and cognition, the weak kernel represents knowledge loss only due to waning memory. On the other hand, the strong kernel describes both knowledge gain due to learning and loss due to waning memory. As $t$ increases, the weak kernel $\mathcal{G}_w$ is monotonically decreasing, while the strong kernel $\mathcal{G}_s$ is increasing first and then decreasing. The cognitive map is then taken to be $a(x,t) = \gamma v_u(x,t)$, and evolution of motion with a birth/death process is described by

$$\frac{\partial u}{\partial t} = d_1 \Delta u - \gamma \nabla \cdot (u \nabla v_u) + f(u). \quad (2.27)$$

Here, the advective potential $v_u(x,t)$ as defined in (2.25) with either the weak or strong kernel defined in (2.26). Although the mathematical complexity is increased, the primary motivation remains the same: agents in population $u$ have a bias in their movement, and this bias is explicitly driven by an attraction (or repulsion) to (from) all previous locations at rate $\gamma$, with favor given to experiences closer in space and more recent in time. Model (2.27) can be thought of as the prototypical animal movement model which explicitly includes a distributed memory, in the same way that model (2.19) can be thought of as the prototypical animal movement model which explicitly includes memory through a discrete delay.

This prototypical model can be generalized to include different quantities as the cognitive map, as well as to include non-local effects in the growth term. As model (2.21) was a generalization of (2.19) in the discrete delay case, [59] generalizes (2.27) to include a distributed delay in the maturation process. To this end, define

$$v_i(x,t) := \mathcal{G}_i \ast u(x,t) := \int_{-\infty}^{t} \int_{\Omega} G(d_3, x, y, t - s) \mathcal{G}_i(t - s; \tau_i) u(y, s) dy ds, \quad (2.28)$$

for $i = 1, 2$. In this construction, there are two different delay values $\tau_i$, where $\tau_1$ is related to the delay for inclusion of memory and learning, and $\tau_2$ is related to the delay due to the maturation process. This is a similar interpretation to that found in model (2.21) for the discrete delay case. $G(d_3, x, y)$ is a spatial kernel which is again taken to be the Green’s function for the heat equation subject to homogeneous Neumann boundary conditions. Similar to model (2.27), the diffusion coefficient for this spatial kernel $G(d_3, x, y)$ is
taken to be identical to $d_1$, the diffusion rate for the random movement of the population. The kernels $G_i(t; \tau)$ are then taken to be either the weak or strong kernel defined previously. This leads to four possible combinations for each kernel type. The evolution equation is then given as

$$\frac{\partial u}{\partial t} = d_1 \Delta u + d_2 \nabla \cdot (u \nabla v_1) + f(u, v_2),$$

where $d_1 > 0$ and $d_2 \in \mathbb{R}$ corresponds to either attraction or repulsion depending on the sign.

Finally, we may note that the double convolution kernel introduced in (2.25) can be more generally thought of as

$$v_a(x,t) = G * * a(x,t) := \int_{-\infty}^{t} \int_{\Omega} G(x,y,t-s)G(t-s)a(y,s)dyds,$$

where $a(x,t)$ is any potential related to environmental covariates. For example, instead of taking $a(x,t) = u(x,t)$ as in the prototypical model (2.27), one could consider an additional equation explicitly modelling a dynamic map in relation to landmarks or conflict zones, for example. (2.30) then describes modifications to the cognitive map $a(x,t)$ with respect to both space and time. A general evolution equation which includes modifications in both space and time is then given as

$$\frac{\partial u}{\partial t} = d_1 \Delta u + d_2 \nabla \cdot (u \nabla v_a) + f(u).$$

To conclude, readers should note that the inclusion of distributed temporal delays is a rather recent development as applied to knowledge-based movement models. There are some interesting equivalencies between problem (2.27) and a two component system which is very similar to a Keller-Segel chemotaxis model (see [55, Section 2]; the lemma statement is found in Appendix B.3.2 for completeness). To motivate future efforts, we have also generalized the distributed delay to include any cognitive map, hence the arbitrary choice of $a(x,t)$ appearing in (2.30). It would be an interesting direction of study to consider cases where $a(x,t)$ is a more general cognitive map, similar to those introduced in Section 2.2, and to compare the results with and without the inclusion of a distributed temporal delay. This could provide insights into the effect of learning on animal movement models, motivating the following open question.

**Open Question:** How might a more general cognitive map be influenced by distributed or discrete delays? That is, how might the combined effects of a dynamic map as found in Section 2.2 with discrete or distributed delays as in the models introduced above influence the space use outcomes? What temporal kernel, if any, is optimal for species survival? If so, when are they optimal?

### 2.4 Short and Long Term Memory

In the derivation of the space use coefficients for model (2.1) (see Appendix A.2), it is suggested that there should be multiple *layers* or *channels* which together construct the cognitive map influencing movement. Each of these layers may incorporate a number of important environmental covariates known to influence animal movement, such as food acquisition, territory defence, or mate finding [13]. In each of the models discussed thus far, only one layer is incorporated into the cognitive map. One possible motivation for the inclusion of at least two layers that work together to inform the animal’s movement is through distinct short and long term memory components (sometimes referred to as a “bi-component” mechanism [5, 48, 65]).

From a biological standpoint, a wide range of species are believed to rely on short and long-term memory for seasonal or long-distance migration [28]. Recently, it has been shown in a stochastic setting that the inclusion of both short and long term memory layers are necessary in order to produce periodic movement in a periodic environment [32]. However, such a result may be less insightful in a PDE setting as a periodic environment guarantees the existence of a periodic solution depending on the sign of the principal eigenvalue to a linearized problem [22, Ch. 2 & 3].

In order to consider the effects of short and long term memory, one makes the following set of assumptions:

- short term memory memory has larger decay and uptake rates, while long term memory has smaller decay and uptake rates. In this way, long term memory takes more time to form but decays less rapidly; on the other hand, short term memory responds to changes quickly and fades easily. If we denote by $m_s$ and $m_l$
the short and long term memory, respectively, the evolution of the short and long term memory layers could be described by

\[
\begin{align*}
\frac{\partial m_s}{\partial t} &= \alpha_s a_s(x, t) - \beta_s m_s, \\
\frac{\partial m_l}{\partial t} &= \alpha_l a_l(x, t) - \beta_l m_l,
\end{align*}
\]

(2.32)

with motivation taken from models (2.12) and (2.11). The parameters should be chosen so that \( \alpha_l < \alpha_s \) and \( \beta_l < \beta_s \) to capture the effect of short versus long term memory and the respective decay/uptake rates. The functions \( a_s, a_l \) should, in general, depend on time and space, and will be associated with whichever environmental covariate is being tracked. For example, one could take \( a_s(x, t) = a_l(x, t) = m(x, t) \), where \( m(x, t) \) is a resource density. The cognitive map is then

\[ a(x, t) = c_1 m_s(x, t) + c_2 m_l(x, t), \]

and the evolution equation describing movement with perception and short and long-term memory is given by

\[
\frac{\partial u}{\partial t} = d\Delta u - \nabla \cdot (u \nabla g_{a,R}),
\]

(2.33)

for an appropriate detection kernel \( g(\cdot) \) and perceptual radius \( R \). The coefficients \( c_i, i = 1, 2 \), could be either positive or negative depending on application. In [32], for example, both coefficients are taken positive so that foragers are attracted to both short and long-term memories. On the other hand, [5] suggest that long-term memory could be an attractive force while short-term memory is repulsive, i.e. \( c_1 < 0 < c_2 \). For example, [5] incorporates an attraction to high resource areas through long term memory, while being repelled from areas it has recently been through short-term memory in order to allow resources to replenish. This formulation may be useful in exploring the effects of how the time since last visiting a location influences movement decisions in a number of contexts, such as in relation to resource density [5, 50, 44] or territory surveillance and prey management [51]. Readers should note that in these references, the models are stochastic or statistical. To our knowledge, (2.32)-(2.33) is the first incorporation of short and long-term memory in a diffusion-advection equation setting. This opens an interesting avenue of study, as this combination of effects allows one to include a prioritization of information through an ordering of advection rates. Logically, it makes sense to prioritize avoiding predators over obtained sustenance, for example. Variable rates of advection depending on other external factors (e.g. hunger or other satisfactions measures, see Section 2.5) provides a more complex mechanism which includes the effect of prioritization.

**Open Question:** What role does short vs. long term memory play in differing scenarios, e.g. competition outcomes, foraging outcomes or migration patterns? Are both short and long term memory components necessary and/or sufficient to predict certain space use phenomena, if ever?

**Open Question:** How does prioritization of information affect outcomes? One such conjecture can be stated roughly as follows: Avoiding predators before obtaining food results in better outcomes than if the priorities were reversed. How do interactions of other cognitive map types influence space use outcomes?

### 2.5 Learning

The phenomenon of *learning* is more difficult to quantify in the current setting. This is exacerbated by the fact that there are many different forms of learning, ranging from simple habituation (a change in behaviour through repeated exposure to stimulus) all the way to observational learning (learning through mere observation). If we take the psychologists definition of learning, “the cause-effect process leading to information acquisition that occurs as a result of an individual’s experience”, as some modellers have [30], one may argue that any model featuring a dynamic cognitive map constitutes learning. For example, if a cognitive map \( a(x, t) \) is described by an ordinary differential equation

\[
\frac{\partial a}{\partial t} (x, t) = \alpha(x, t) - \beta(x, t)a(x, t),
\]

then the growth term \( \alpha(x, t) \) is the learning mechanism already implicitly included in the model. Therefore, one can argue that all models presented in Section 2.2.2 that feature an additional equation describing the
The evolution of a cognitive map already includes learning implicitly. On the other hand, models featuring time delays (Section 2.3) may or may not include a learning process. Models that feature discrete time delays, for example, do not have a learning mechanism. On the other hand, distributed delays with a strong kernel include an implicit learning process. This definition of learning may be too broad for further study, however, since it does not allow one to distinguish a genuine “learning process” from other phenomena. This can be problematic in at least two different ways. First, some models feature a cognitive map that lies “outside” the mind of the foragers, as in model (2.11) describing the evolution of marks on the landscape. One can conceptually distinguish between a learning process and an external process, but they are described in an equivalent way and so cannot be easily distinguished from an analytical perspective. Second, this broad definition of learning may create difficulties when making comparisons to real world data. For example, if one seeks to compare a model to data in order to identify whether a learning process has occurred and make connections to the resultant movement pattern observed, it becomes impossible to distinguish whether a learning process has in fact occurred, or if the movement pattern is a consequence of another simpler mechanism.

Alternatively, some consider learning to be markedly different than memory when defined as “modifications to a forager’s behaviour through experience/knowledge acquisition” [63, 49], and therefore provides an avenue to determine whether or not a learning process has occurred. From this point of view, none of the models introduced thus far feature learning since their movement mechanism remains the same for all time. That is, the movement mechanism is an attraction (or repulsion) from the gradient of the cognitive map, however it may be constructed. This means that a population can never learn to change their behaviour in relation to environmental covariates and the way the components considered influence their movement decisions. This motivates one to consider the effect of variable rates of attraction or diffusion, e.g. given an advective potential $a(x,t)$ with rate of attraction $\gamma$, the rate $\gamma$ should be allowed to change in sign and/or magnitude as new information is obtained. Indeed, it has been empirically shown that certain factors can increase locomotion activity in some species [9]. One such motivating example is the concept of starvation driven diffusion [9], which incorporates a mechanism for the rate of diffusion to decrease when an organism is satisfied with its current environment, and will increase when the organism is unsatisfied with the current environment. To this end, a measure of “satisfaction” must be introduced, and this quantity can be phenomenologically viewed as a learning mechanism. To this end, a satisfaction measure was defined in [9] by

$$s = s(x,t) := \frac{\text{food supply}}{\text{food demand}},$$

and so it is assumed that the animals have knowledge of the current food supply and demand. If the resource distribution is given as a function $m(x,t)$, one then has

$$s(x,t) = \frac{m(x,t)}{u(x,t)},$$

where $u$ is the population density of foragers, and so it is assumed that food demand is related directly to the population density. Notice that $s(x,t)$ is similar to the static cognitive map suggested in model (2.8), but we now use this quantity to measure satisfaction as opposed to using it as the cognitive map itself. If $s > 1$, the food supply is larger than the food demand, and so motility should decrease. If $s < 1$, the food supply is smaller than the food demand and motility should increase. Thus, the changes to the rate of diffusion should be a composite function $\omega(s)$, where $\omega(\cdot)$ is small for arguments greater than one and large for arguments less than one. A simple example is the discontinuous function

$$\omega(s) = \begin{cases} d^+, & 0 \leq s < 1, \\ d^-, & 1 < s < \infty, \end{cases}$$

for $0 < d^- < d^+ < \infty$. In general, it is suggested that a satisfaction measure should satisfy

$$\omega(s) \nearrow d^+ \text{ as } s \searrow 0^+, \quad \omega(s) \searrow d^- \text{ as } s \nearrow \infty,$$

where $d^+$, $d^-$ are the maximum and minimum rates of diffusion.
This alternative view of learning can be incorporated into the advection term as well: instead of modifications to the rate of diffusion, one could introduce modifications to the advection speed, or even the sign of the advection speed. This may introduce mathematical difficulties, however, as one cannot include the advection speed outside of the gradient term. Instead, the rate must appear inside the gradient term as is found in the derivation of space use coefficients, see Appendix A.2. More precisely, suppose the cognitive map is given by \(a(x,t)\) and the advection rate is give by \(\gamma = \gamma(x,t,u)\), which is no longer constant and may depend on the population density \(u\) and other environmental factors. The correct form for this advection with perception would be \(-\nabla \cdot (u \nabla (\gamma(x,t,u)\sigma_{g,R}))\), as opposed to \(-\gamma(x,t,u)\nabla \cdot (u \nabla \sigma_{g,R})\). Furthermore, while \(\omega(\cdot)\) must remain positive when modifying rates of diffusion (in order to maintain parabolicity of the equation), \(\gamma(x,t,u)\) can change sign when modifying advection speeds. In such a case, a sign change would indicate a change from attraction to repulsion (or vice versa), which may be a stronger indication of learning as it indicates a change in kind rather than a change in amount. Such a consideration may also allow one to overcome an issue discussed in Section 2.2.2: a single equation could, in principle, be sufficient to describe a cognitive map if the rate of advection is allowed to change sign. This point is raised briefly in the discussion of [54], where it is suggested that the sign of the advection speed may change from aggregation to segregation depending on environmental conditions.

To demonstrate this point, we construct a simple model combining the effects of starvation driven advection and a den site. First, we assume that there is some constant rate of attraction to a central den site located at \(x_0 \in \Omega\). We then assume that foragers are attracted to the local resource density, but the rate at which they move up the gradient of the resource profile will now depend on the satisfaction measure \(s(x,t)\) defined in line (2.35). We then define a modified advection rate \(\tilde{\omega}(s)\) to be

\[
\tilde{\omega}(s) = \begin{cases} 
\gamma^+, & 0 \leq s < 1, \\
0, & 1 < s < \infty,
\end{cases}
\]

where \(0 < \gamma < \gamma^+\) so that the movement towards resources becomes priority when satisfaction is low (i.e. when hunger is high).

Motivated by the formulations above, one may incorporate the idea of a satisfaction measure with those introduced in [46] so that foragers may update their advection rates based on (possibly incomplete) knowledge of the local resource density and an expected resource density. Recall the form of the cognitive map \(m(x,t) / \bar{m}(t)\) introduced in model (2.7), where \(\bar{m}(t)\) denotes the average resource density over the entire habitat at time \(t\). Then, the advection rate can be taken as a composite function with the new satisfaction measure \(s(x,t) = m(x,t) / \bar{m}(t)\), where \(s > 1\) in above average resource areas and \(s < 1\) in below average resource areas. Depending on the expected behaviour of the forager, \(\omega(s)\) can be formulated so that the advection speed either increases or decreases (such as the discontinuous example above), or even changes sign, depending on this satisfaction ratio. This begins to address a key difference between spatial memory and attribute memory: the formulations above provide at least one direction to distinguish between the quality of movement decisions and behaviour, rather than the mere locations that are desirable or undesirable.

**Open Question:** For the new model (2.37), can the starvation-driven switch between two cognitive mechanisms, foraging and home attraction, generate nontrivial spatiotemporal dynamics and patterns? Biologically, would such realistic switches benefit or impair the species? What other satisfaction measures can be introduced to act as a behaviour modification mechanism?

### 3 Biological Insights Through Mathematical Analysis

In this section, we discuss some of the biological insights we have gained, and continue to gain, through the analysis of the models proposed so far. First we bring attention to some useful “Rules of Thumb” that any
modeller should consider after a model has been formulated. We also briefly discuss potential “Measures of Success” in animal movement models, as well as their importance to the exploration of these models and the insights gained through their study. In some cases, we are able to discuss existing insights gained through such measures. In other cases, we propose modifications or suggestions for further study. We conclude with a discussion of pattern formation, a quintessential feature of animal movement models, and how we may investigate this possible outcome in the relevant models introduced so far.

3.1 Rules of Thumb

In this section we introduce some rules of thumb that should be considered upon the development of any animal movement model. We divide these rules into “prerequisites” and “considerations”. Prerequisites are properties that the model should necessarily possess so that the conclusions made are logical and as applicable to the real world as possible. Considerations, on the other hand, are helpful suggestions that may be relevant for some models, but may not be strictly required by all models and can be more or less valid depending on the specific circumstances considered.

3.1.1 Prerequisites

We first discuss some necessary features any reasonable model should possess: existence and positivity. Without these two conditions verified, the model may be argued to be unreasonable and therefore an unrealistic representation of the real world, especially when applied to ecological modelling.

Existence: Existence simply refers to the existence of a solution to a given problem in some suitable context. It is well known that there are problems for which solutions exist, problems for which no solutions exist, and problems for which we do not know whether a solution exists or not. Furthermore, when the solution depends on a temporal variable, it may exist for all time \( t \in (0, \infty) \), or it may only exist on some finite interval \( (0, T) \) with solution blowup as \( t \to T^- \) [7, Ch. 1.6].

While it is not inappropriate to investigate ecological models without knowing the existence of solutions apriori (consider the famous Navier-Stokes equations), it is often possible to show the existence of solutions under relatively weak assumptions. Depending on the context, one may obtain classical solutions [67, Ch. 8] (sufficiently differentiable in both space and time), but other contexts may require a more general form of solution, such as weak or strong solutions [67, Ch. 2, 3 & Ch. 9] or mild solutions [40, 1]. These are solutions that may not be differentiable in the classical sense, but solutions can be well-defined through notions of weak differentiability or semi-group theory. The technical details of such formulations is beyond the scope of this review, however we have provided some key references for now standard results. It is also worth noting that many of the models proposed here do not necessarily fall within the standard theory. This leaves many open problems regarding the existence of solutions. See Appendix B.1 for a more explicit discussion of open questions concerning the existence of solutions.

Positivity: We conclude with the notion of positivity. This simply says that, given an initial data which is positive (somewhere), the solution will also be positive for all future times. In particular, it is highly unreasonable for a solution to attain negative values when it is meant to model the population density or probability density function of a group of organisms! This is sometimes a possibility in delay differential equations. For a wide range of problems with no time delays, a standard positivity result often follows immediately from the maximum principle. However, this is a highly nontrivial question for virtually all models discussed so far: since the model formulations are non-standard, we do not have much of the tools used in a classical analysis of partial differential equations. The flip side of positivity then says that if the time dependent solution ever becomes identically zero at a time \( t_0 \), it must remain zero for all future times \( t \geq t_0 \). Should this be violated, the model describes a phenomenon that can never happen in reality (e.g. a spontaneous birth process) and should be modified appropriately.

3.1.2 Considerations

Next, we briefly highlight some of the features anyone should consider when proposing a new model. These properties are by no means absolutely necessary for a model to be valid, but they offer insights into the question under what conditions or assumptions is the model most valid?
Uniqueness: The property of uniqueness ensures that there is exactly one non-trivial solution which solves the problem. For time dependent problems, this implies that there exists exactly one non-trivial solution corresponding to a given initial condition. In terms of ecological (and other) modelling efforts, it is a desirable property to hold in order to draw concrete conclusions from the model. Indeed, should a model have two or more solutions corresponding to the same set of initial data, it is no longer possible to determine which outcome could be realized in the real world. Furthermore, should a proposed model suffer from a lack of uniqueness, this may indicate that an important biological mechanism has been neglected or ignored, which may motivate alternative model constructions.

Continuity w.r.t. initial data: This property roughly says that small changes in the initial data result in comparably small changes in the solution. This is important for modelling purposes particularly when efforts are made towards fitting real world data: small errors in data measurements may produce unreasonable results if this continuity does not hold.

Initial Conditions: We briefly discuss some reflections on the nature of the initial conditions chosen. First, it is worth noting that the initial data chosen can be arbitrary in theory; the importance, perhaps, is on determining the impact that differing initial data may have on the space use outcomes. In many classical PDE models, initial conditions may have no impact on space use outcomes (e.g. Fisher-KPP equation); on the other hand, some models have significantly differing outcomes depending on the initial data chosen (e.g. a strong Allee effect). It is not clear how the choice in initial data effects space use outcomes in a general sense, especially in relation to the models introduced here. However, there is strong numerical evidence to suggest that patterns only form if the initial condition is sufficiently close to a segregation pattern, see e.g. [41]. In such a case, pattern formation can no longer be considered “spontaneous” in the classical sense. Instead, such outcomes may be thought of as “pattern stimulating”. There is, however, one exception to the arbitrary choice in initial data, and this is in relation to experimentation. Should a model be constructed such that it may be fit to experimental data, the initial data must match that of the experiment itself.

Boundary Conditions: Next, we consider the choice in boundary conditions. Modellers spend a significant amount of time describing what happens within the domain (e.g. the habitat), but we must also specify what occurs when an agent reaches the boundary of the domain (e.g. the edge of the habitat in a bounded domain). Often this is presented from a mathematical perspective with three main types (Dirichlet, Neumann and Robin type) introduced. However, we will approach this first from the biological perspective, and then make some brief comments on the connection to prospects of mathematical analysis.

To motivate the reader, we consider some of the models introduced thus far. Many feature no birth/death processes (see Sections 2.1-2.2.2, for example), and so it is reasonable to assume that the total population remains fixed for all time. In other words, the models proposed describe only the animal movement. In such a case, the most appropriate boundary condition is referred to as the zero-flux or reflecting boundary condition. A zero-flux boundary condition ensures the total population remains fixed for all time. To determine the appropriate zero-flux boundary condition, one simply integrates the equation over the domain and applies the divergence theorem to show that \( \frac{d}{dt} \int_0^L u(x,t)dx = 0 \), which naturally implies that the total population remains constant for all time, and is identical to the initial population density. As a simple but instructive example, consider (2.1) in \( \Omega = (0, L) \). Integrating over \( (0, L) \) and applying the divergence theorem yields

\[
0 = \frac{d}{dt} \int_0^L u(x,t)dx = \int_0^L (\partial_x u_x - a_x(x,t)u) \, dx = \frac{\partial u}{\partial x} \big|_0^L - \frac{\partial a}{\partial x} u \big|_0^L.
\]

Hence, the total population is conserved if we prescribe at the endpoints

\[
d\frac{\partial u}{\partial x}(0,t) - \frac{\partial a}{\partial x}(0,t)u(0,t) = d\frac{\partial u}{\partial x}(L,t) - \frac{\partial a}{\partial x}(L,t)u(L,t) = 0.
\]
We can then make a connection between this zero-flux boundary condition and another common boundary condition, referred to as a homogeneous Neumann boundary condition, which prescribes the outer normal derivative of the solution at the boundary:

\[ \frac{\partial u}{\partial n}(x,t) = 0, \quad \text{on } \partial \Omega. \]

Here, \( \frac{\partial}{\partial n} \) denotes the outer unit normal vector to the boundary of the domain \( \partial \Omega \). Sometimes, a homogeneous Neumann boundary condition is also referred to as a zero-flux boundary condition. This raises an interesting point that may also contribute to some confusion. In some cases, a homogeneous Neumann boundary condition is sufficient to ensure a conserved population. In these cases, a homogeneous Neumann condition is equivalent to a zero-flux condition. This would be the case in any model where the attractive potential \( a(x,t) \) depends on the solution \( u(x,t) \) itself. In this way, \( \frac{\partial a}{\partial n} = \frac{\partial u}{\partial n} = 0 \) along \( \partial \Omega \), and so (3.1) is automatically satisfied. This is the case for model (2.9), for example. On the other hand, a homogeneous Neumann boundary may not conserve the population if \( a(x,t) \) is independent of \( u(x,t) \). This would be the case in model (2.4) where \( a(x,t) \) is given apriori, for example, so that the zero-flux and the homogeneous Neumann condition are no longer equivalent.

In order to conserve the population in this case, we need to satisfy what is sometimes referred to as a Robin or mixed boundary condition. This condition is classically written as

\[ \alpha(x) \frac{\partial u}{\partial n} + \beta(x)u = 0, \quad \text{on } \partial \Omega, \quad (3.2) \]

but can be generalized to include cases where the coefficients are time dependent [38, Ch. 2]. If \( \alpha, \beta > 0 \), this boundary condition can be interpreted as a partial loss of the population at the boundary, for example. Choosing \( \alpha(x,t) = d, \beta(x,t) = -\frac{\partial a}{\partial n} \), the Robin condition (3.2) is equivalent to the zero-flux condition (3.1).

The final boundary condition we discuss is the homogeneous Dirichlet boundary condition

\[ u(x,t) = 0, \quad \text{on } \partial \Omega. \]

This is sometimes referred to as a hostile boundary condition. As this name suggests, this condition assumes that the boundary is completely lethal and any agent that reaches the edge is removed immediately and never returns (e.g. the animal dies). Such a condition is most appropriate for species on an island, surrounded by a cliff edge, or in a particular necessary ecological niche, for example. In general, a homogeneous Dirichlet condition will not conserve the population since a hostile boundary is a mechanism by which there is a loss of the population without an explicit birth or death process included in the model equation.

A variety of differing boundary conditions should be considered for biological purposes and application, including moving boundary conditions, e.g. [14], where a non-Stefan free boundary condition was proposed via a trade-off between shorter and longer spatial scales. In this sense, there are no correct or incorrect choices should sufficient justification be provided for the choice made. From a mathematical perspective, however, different boundary conditions may offer different routes for analysis. In the case of Dirichlet boundary conditions, for example, the well-known Poincaré inequality in \( H^1_0(\Omega) \) provides a compactness result that is important for proving the existence of solutions in a number of contexts. A zero-flux boundary condition has a similar result, since a conservation of the total population means that one can work in the space of \( H^1(\Omega) \) functions such that \( \int_\Omega u \, dx = 1 \), which has comparable compactness properties, see [11, Ch. 5] for an elementary treatment. Other choices, e.g. a homogeneous Neumann condition that does not result in a conserved population, may increase the difficulty of analysis due to a loss of compactness. Another popular boundary condition, the so-called periodic boundary condition, may also be chosen for convenience of analysis. This boundary condition is used in [23], for example, where a regularity result for the heat equation semigroup on a torus is used. Technically the periodic boundary condition can be extended to an infinite domain such that the detection function with a perceptual range needs no modification near the boundaries.

**Domains:** Related to the boundary conditions discussed above is the associated domain in which the model is studied. In most models introduced here, a one dimensional spatial domain is chosen for ease of analysis (some exceptions apply, as in some of the time-delay models introduced here). This can be a bounded interval or an unbounded interval. In the case of a bounded interval, it may not be entirely clear how to
define a quantity with some perceptual radius included near the boundary: indeed, a kernel with perceptual radius \( R > 0 \) must have a precise definition within \( R \) of the boundary. One solution is to “cut off” the kernel, essentially extending its domain outside of the habitat by zero.

**Timescales:** To conclude, we briefly discuss some issues one may encounter related to differing timescales of components included in a model. A significant portion of movement models described here do not include a birth/death process. In practice, this may be reasonable if one assumes that animal movement occurs at a timescale much faster than the birth/death processes [42]. However, this may result in difficulties in other respects. For example, a common technique in the analysis of these knowledge based movement models is to take advantage of the so-called quasi-steady state approximation in the equation for the cognitive map. That is, if a map \( a(x,t) \) is described by a dynamic equation, one sets \( \frac{\partial a}{\partial t} = 0 \) and can then solve for \( a(x,t) \) explicitly. This reduces the number of equations appearing in the model, but it may not be appropriate to make such a simplification if the growth of a cognitive map occurs at a timescale that is the same as that of the actual movement. This may be the case for model (2.11), where the “cognitive map” is given by marks on the landscape, which necessarily occur at the same timescale as the animal movement. Consequently, the biological insights may not be accurate. On the other hand, since perception occurs very quickly and influences movement for a longer period of time, it may be reasonable to use a quasi-steady state approximation for models which describe the cognitive map itself, such as model (2.12). Another possible remedy to address this issue is to include a birth/death process in the model as well so that the timescale over which these models are valid becomes less of an issue from the beginning.

We wish to emphasize that none of these rules are always wrong or always right. Instead, we hope to provide some additional insights into some of the assumptions commonly made, when they are most valid, and when they may need to be reconsidered.

### 3.2 Measures of Success in Movement Models

Having a unified framework to measure the success of certain cognitive mechanisms will be invaluable to the assessment of which mechanisms lead to more fruitful outcomes for the population. In this sense, a unified framework may be interpreted more accurately as unified within groups of comparable models. For example, it should be possible to compare cognitive mechanisms within models that describe a population foraging for available resources; on the other hand, it may not make sense to compare the success of foraging for resources versus the success of avoiding predators/competitors. Ultimately, a measure of success is highly dependent on context, and should be treated as such. In the following, we seek to highlight some existing methods that can be used to compare and contrast models that feature knowledge based movement mechanisms, as well as some alternative definitions and generalizations for future study.

#### 3.2.1 Measures with Explicit Resources

Currently, one of the only quantitative measures of success in regards to animal movement models is within the context of foraging for resources. This quantity is referred to as the foraging success, and has been used as a measure to determine optimal diffusion and advection rates for model (2.4), see [13]. The foraging success is defined as

\[
\text{Foraging Success} := \int_{t'}^{t_{\text{max}}} \int_{\Omega} u(x,t) m(x,t) dx dt,
\]

where the interval \((t', t_{\text{max}})\) is chosen so that the transient dynamics of the solution \( u(x,t) \) have settled down. Recalling that the function \( m(x,t) \) is the resource density, one may also note that if the resource density is static (i.e., \( m(x,t) = m(x) \)), the integration is over space only. The purpose of this measure is to quantify the effectiveness of consumer-resource tracking, in this case based on non-local information. As noted in [13], however, model (2.4) does not consider mutual interference or resource depletion, and so this measure may only be valid under the assumption of sparsely populated regions and that the resources degrade more quickly than they are depleted by foragers. This may then introduce further issues, as one of the common assumptions for a knowledge-based animal movement model framework is that there is a high enough population density so that the PDE model derived is an appropriate mean-field approximation of
the population density \[42\]. Finally, it should be noted that the foraging success used above is a \textit{predictor} of success, not a \textit{result} of success. In other words, if the foraging success is large, the foragers are expected to do well.

This motivates a need for alternative forms of foraging success, given a population density \(u(x,t)\) and a resource density \(m(x,t)\), as a \textit{result} of successful foraging. In other words, if the foragers are succeeding, the measure will reflect this in relation to the available resources. In this way, we propose an \textit{efficiency} measure, rather than an \textit{effectiveness} measure. This should have some general, reasonable qualitative features in order to accurately reflect “success”:

i) When the population density is low (high), the foraging success is low (high), but if the resource density is also low (high), a low (high) population density may still be considered a success (not a success) when compared to a low (high) population density with high (low) resource density.

ii) Foraging success should incorporate some averaging process over time, but should not be unbounded when \(t_{\text{max}} \to \infty\).

Point i) is an assumption based roughly on common sense: higher population densities should be viewed as more successful than lower population densities, as this suggests a thriving foraging community. However, this assumption alone is insufficient to describe success \textit{in relation to the resource density}. Indeed, if the objective is to measure the \textit{efficiency} of consumer-resource tracking, the success must be in relation to the resource density itself. This motivates the caveats mentioned, which allows one to more closely compare and contrast two successful or unsuccessful populations in relation to available resources. Point ii) is a condition one should implement in order to keep differing foraging strategies (perceptual radii, kernels, advection speeds etc.) comparable over a long time period. Without this condition, it is feasible that the measure of foraging success gives too much weight towards minor variation over long time periods. In such a case, it is not clear what the correct time interval to choose is. One possible remedy can be formulated in the case of a periodic attractor, i.e. a solution \(u^*(x,t)\) that solves the equation and satisfies \(u^*(x,0) = u^*(x,T)\) for some period \(T > 0\). Often, this period is related (or is exactly) the period of the resource function \(m(x,t)\). In this case, it makes most sense to take the integral over one full period one the solution \(u(x,t)\) is sufficiently close to this attractor \(u^*(x,t)\). To this end, we propose a modification to the foraging success defined above as follows:

\[
\text{Modified Foraging Success} := \frac{1}{T} \int_0^T \frac{1}{|\Omega|} \int_{\Omega} \frac{u^*(x,t)}{m(x,t)} \, dx \, dt, \tag{3.4}
\]

for a given periodic resource density \(m(x,t) > 0\) satisfying \(m(x,t) = m(x,t + T)\) for all \(t > 0\) with periodic attractor \(u^*(x,t)\) with the same period. This provides an average “population density per unit resource density”. The averaging over time addresses issues raised above, while an averaging over space allows one to compare differing habitat sizes. For further details on periodic-parabolic initial value problems, we refer readers to [22, III.21]. Numerically this can be achieved by running the solution until the maxima and minima over a distant time interval are no longer changing; one may then integrate between successive maxima or minima to obtain one full cycle.

**Open Question:** For knowledge-based movement models, what are the results of modified foraging success? Are they consistent or opposite to the existing results of foraging success?

**Open Question:** How can we generalize the foraging success measure further? What about other measures for scenarios that do not include explicit resources (see Section 3.2.2 below for some preliminary suggestions)? Should the foraging success be constant over an average period of time, or dynamically changing with time?

### 3.2.2 Measures with Population Dynamics

In the previous discussion, it is assumed that there is a true resource density with no population dynamics. In other cases, such as the models appearing at the end of Section 2.2 and all found in Section 2.3, population dynamics are also considered. Foraging success may still be considered if explicit resources are included. However, there are alternate, more quantitative ways to determine the more “successful” strategy, where success is now measured with respect to growth rates and/or survival as opposed to effective or efficient resource tracking.
The first alternative, motivated by net growth rates in cell growth described in works such as [26], considers the rate of change of the total population over a prescribed time interval:

$$\text{Net Growth} := \int_{t'}^{t_{\text{max}}} (\text{Change in total population density}). \quad (3.5)$$

Thus, if $u(x,t)$ measures the population density, the change in total population density is $\frac{d}{dt} \int_{\Omega} u(x,t)dx$, and so

$$\text{Net growth} := \int_{t'}^{t_{\text{max}}} \frac{d}{dt} \int_{\Omega} u(x,t)dxdt = \int_{t'}^{t_{\text{max}}} \int_{\Omega} \frac{\partial u}{\partial t}(x,t)dxdt.$$

From the discussion of boundary conditions found in Section 3.1, the population should be conserved in the absence of growth/death dynamics, and so the Net Growth becomes

$$\text{Net Growth} := \int_{t'}^{t_{\text{max}}} \int_{\Omega} f(x,t,u)dxdt, \quad (3.6)$$

where $f$ describes population growth/death. A numerical exploration of this quantity seems to be the most fruitful direction, as the formulation above leaves the dependence of Net Growth on potential knowledge-based parameters implicit.

**Open Question:** What are the consequences and predictions of knowledge-based movement models in the presence of reaction terms made by the measures of success introduced above? Are they qualitatively consistent in their predictions, and if so, when? Are there alternate definitions of success, perhaps similar to those proposed above, that may be useful in their predictive power?

A second alternative that can be used is a comparison of competitive outcomes, assuming that both survival and extinction are possibilities within the modelling framework considered. In this sense, success is obtained when the population persists as $t \to \infty$. A foraging strategy is unsuccessful if the population goes extinct. This is considered in [9], for example, where it was shown that starvation-driven diffusion may reverse outcomes predicted by constant diffusion alone. Hence, starvation-driven diffusion may be a more successful foraging strategy than random diffusion. Readers should carefully note that this does not mean that starvation-driven diffusion is the more correct mechanism, but rather that it is an alternate explanation should we find results in nature inconsistent with predictions made by simpler models. This could be studied analytically through changes in a principal eigenvalue which determines local or global stability. To be concrete, suppose there exists a quantity $\mu_1$, depending on various foraging strategies introduced in the model considered, such that $\mu_1 < 0$ implies persistence and $\mu_1 > 0$ implies deterministic extinction. Then, changes in $\mu_1$ with respect to available parameters can indicate a successful or unsuccessful foraging strategy.

As a simple example, consider the diffusive logistic equation

$$\frac{\partial u}{\partial t} = D\Delta u + ru(1-u)$$

in a one dimensional domain $(0,L)$ subject to homogeneous Neumann boundary data. It is well known that the stability of the constant steady state $u^* = 1$ is given by the sign of the principal eigenvalue, which in this case is $\mu_1 = -r < 0$, and so $u \to 1$ as $t \to \infty$ for any $r > 0$ from the theory of monotone flows. Taking the derivative of $\mu_1$ with respect to the parameter $r$ yields $\mu'(r) = -1 < 0$, and so increasing $r$ is a quantity which contributes to the success of the species in terms of survival. In contrast, the stability does not depend on the domain size $L$, or the diffusion rate $D$, which can be seen since the derivative of $\mu_1$ with respect to either of these variables is zero. On the other hand, if one considers the same problem subject to homogeneous Dirichlet boundary conditions, it is well known that the principal eigenvalue $\mu_1 = -\frac{r}{D} + \left(\frac{L}{D}\right)^2$ depends on $D$, $r$, as well as the domain size $L$. Taking the derivative of $\mu_1$ with respect to these variables yields

$$\frac{d\mu_1}{dD} = \frac{r}{D^2} > 0, \quad \frac{d\mu_1}{dr} = -\frac{1}{D} < 0, \quad \frac{d\mu_1}{dL} = -\frac{2\pi^2}{L^3} < 0,$$
and so we see that increasing $r$ and $L$ are beneficial strategies, while increasing $D$ is not. An analysis of this sort could be applied to a wide range of models, with the difficulty dependent on the difficulty of the stability properties of the problem at hand, providing phenomenological insights into the impacts certain factors have on population outcomes. In the example above, the relation was found to be one directional. More interesting examples may find non-monotone behaviour, providing an optimal window in which certain strategies are most successful.

**Open Question:** How might we formulate useful eigenvalue problems related to memory and non-local perception? This will necessarily include a careful treatment of the non-local effect of the perceptual kernel, with attention directed towards the top-hat detection function. Is it possible to determine the dependence on quantities such as: perceptual radius, advection speed, rates of diffusion, memory uptake and decay rates, etc.?

### 3.2.3 Measures with No Population Dynamics or Resources

In many of the models introduced in Section 2.2, there are no population dynamics or explicit resources, and so it does not make sense to measure quantities such as foraging success or net growth rates of the entire population. Indeed, in many cases a zero-flux boundary condition is chosen so that the total population remains fixed! In these cases, quantitative measures may not be so obvious.

On the other hand, proving the existence of non-constant steady states may also provide a more qualitative measure of success. Indeed, the emergence of pattern formation without explicit environmental heterogeneity is often considered one of the hallmarks of animal movement models. One may then compare differing memory mechanisms (learning rates/functional forms, memory decay rates etc.) in order to determine the necessity or sufficiency of certain mechanisms to produce territorial patterns. This is discussed in more detail in Section 3.4.

### 3.3 Comparisons through Measures of Success

We begin with model (2.4). In [13], the objective is to use the foraging success (defined in (3.3)) as a measure to compare different diffusion and advection rates, as well as differing forms of perception kernels and perceptual radii. Interestingly, the foraging success is found to be non-monotone with respect to perceptual radius $R$ in certain environments, see [13, Figure 3]. On the other hand, the foraging success is monotonically increasing with respect to the advection rate $\alpha$ under the special case when $R \to 0^+$, see [13, Figure 2]. This provides further insights into questions such as: What is the optimal detection scale for non-local information gathering? For what kinds of movement and detection does non-local information provide a benefit? For what kind of landscapes is non-local information most useful? One important aspect to note here is that these questions are possible to answer due to an available measure to infer comparatively better or worse scenarios within the context of this single model.

Model (2.4) is most closely related to the new models (2.7) and (2.8) in that the resource density is a given function. There are no existing results concerning these models, and so this opens an interesting avenue for investigation: which knowledge strategies are optimal in terms of the foraging success of each model? It may be the case that one strategy is not always optimal, and so it becomes interesting to consider under which conditions each strategy is optimal.

Motivated by these types of investigations, one may then expand this exploration to any model which can be investigated through any of the alternative measures of success introduced in Section 3.2. For models that feature an explicit resource density, it would be interesting to compare the results between foraging success and a modified foraging success as defined in (3.3) and (3.4), respectively. For models that feature population dynamics, a similar exploration could be done through the Net Growth measure defined in (3.6). Each of these measures is best studied through numerical simulation, as these quantities are almost surely intractable through analysis alone.

Finally, one direction of analysis that may be fruitful is the study of eigenvalue problems and the impact that varying quantities have on the sign of the principal eigenvalue. This would ultimately require the development of new techniques in order to deal with non-local problems (in space). For example, is it possible to represent a principal eigenvalue for any of the problems here through a variational characterization? If so, it may be possible to determine the impact of parameters as they appear in the characterization itself,
whether beneficial or detrimental to survival. On the other hand, if there are no population dynamics, the sign of the eigenvalue may determine the appearance of pattern formation in the steady states (see Section B.2 for further details). In the case of periodic solutions, a time-dependent eigenvalue problem may be relevant, in which case a variational characterization is unavailable. However, one may still investigate possible impacts based on the solution to the eigenvalue problem itself.

**Open Question:** How does the foraging success change when the landscape is dynamic? That is, when \( m(x, t) \) is not given apriori, but solved through an additional equation. Is the direction of the foraging success in relation to available parameters the same or different when compared to prescribed resources, and if different, when are they different?

### 3.4 Emergence of Patterns

One of the hallmarks in the field of movement ecology is the generation of patterns in space, time or both. Classically, this is discussed through linearization techniques which gives rise to so-called Turing instability, where a constant steady state can be destabilized in the presence of diffusion. There are now a significant number of resources exploring in great detail the possible outcomes, including not only when patterns will form, but also which types of patterns one might expect to find. Furthermore, these patterns can emerge independent of environmental heterogeneity. While the original motivation was to apply these models to cell growth phenomena, some of these same ideas and motivations transfer to the study of movement ecology.

Indeed, one of the primary goals often discussed in movement ecology is the potential emergence of patterns and their relation to the underlying mechanisms and parameters appearing in the movement models. As it currently stands, we have little understanding of when patterns may emerge, let alone which types of patterns we might expect to appear. For the implicit memory models discussed in Section 2.2, the same linearization techniques used in classical setting yield some preliminary insights: foragers must have sufficient uptake of memory, and cannot forget experiences too quickly in order for patterns to emerge. Additionally, the advection rate must be sufficiently large compared to the diffusion rate in order for patterns to form. However, we still do not have a robust understanding of necessary and sufficient conditions for patterns to emerge. A similar trend holds for the explicit memory models discussed in Section 2.3: the advection rate must be sufficiently large in relation to the diffusion rate for constant steady states to be destabilized. Of course, the techniques used in these two settings are quite different and should be separately treated analytically.

From a mathematical perspective, this yields a rich area of analysis which can be roughly understood across all models that investigate such a question: when do patterns form, and what kinds of patterns are possible? It is not clear what kinds of patterns are possible for implicit memory models; existing works consider only the possibility of patterns emerging in relation to parameters appearing in the model. We conjecture that some models may in fact have infinitely many piece-wise constant steady states (see [41, Sec. 3.4]). In the case of explicit memory models, we have a better understanding of what types of patterns may form, at least from a simulation perspective: spatially constant temporally periodic steady states, temporally constant spatially varying steady states (stripe patterns), and spatiotemporally varying steady states (checkerboard patterns) have all been observed. However, it is still not clear exactly when these patterns are expected to form, and whether these are all of the possibilities (we believe not).

Ultimately, linearization techniques alone are not enough to obtain the deepest insights that are desired. While they provide insights into when patterns may form, these techniques do not tell us which patterns will emerge, and more importantly do not tell us the mechanisms behind specific patterns that emerge.

**Open Question:** How do these new modelling components introduced in this review affect the stability of constant steady states? Can we identify and classify the possible patterns? What are the mechanism behind the generation of certain patterns? Can we connect existing simulations with analytical insights? Can these be connected to patterns found in nature, where data is available? How should we interpret these patterns, both biologically and mathematically? All patterns are ultimately mysterious: are they typical? Are these strange cases or normal for these models?
4 Concluding Remarks

We conclude our discussion with some overarching themes and broad impacts of the works discussed in this review. First, we have taken care to introduce multiple key cognitive mechanisms included in these pioneering models. Most broadly, this includes perception, memory, and learning, which are incorporated through an advective term which biases animal movement in addition to passive diffusion. Perception, included through a spatial convolution, incorporates differing perceptual capabilities through a perceptual radius $R$ and perceptual kernel $g(\cdot)$. Memory, included as an implicit static quantity, an implicit dynamic quantity, or an explicit quantity through time delays, incorporates the process of encoding, storing, and retrieving information within the equations describing movement. Often, this is included through a so-called cognitive map which can take a wide variety of forms, also described here in detail. Learning, either implicitly through memory uptake functions, or explicitly through variable diffusion/advection rates via satisfaction measures, allows one to consider, compare and contrast the learning process through these differing perspectives. In each of these categories, we have described in detail some of the prototypical models appearing in each setting, as well as the connections and comparisons between each. From these formulations, we have provided some of the important insights gained through studying these models. This includes some of the technical details concerning the development and analysis of these new model formulations, as well as the current direction of study within particular classes of models. To help provide direction for future study, we propose a wide variety of new models and related extensions to existing models throughout, in addition to providing insights into the current measures of success used in the literature. This includes critical reflections on the current tools used, and how we might improve and modify these measures for future use.

While we do not assert that we have provided a complete description of all existing cognitive mechanisms included within a diffusion-advection equation framework, we have made a significant effort to include many of the common tools and techniques used in this area of study. In cases where the big picture is perhaps treated as more important than the fine-grain detail, we have provided all relevant references that we are aware of to accommodate further reading. To motivate researchers further, we include a wide range of open questions and interesting directions of study throughout the manuscript. These questions highlight how primitive some of the existing results are when compared to more mature areas of study, and emphasizes how much room there is for growth.

Based on the points raised above and the precise details found throughout this review, it is clear that these models and their extensions can have a very broad impact, both for applied mathematicians and biologists alike. There are many insights provided here, but there are many more connections to be made. This includes a rich, diverse, and challenging branch of mathematical models which will require study from many different perspectives. New insights can be found through a more detailed exploration using existing mathematical techniques, while further insights will require new tools and techniques, leaving much room for up-and-coming researchers to become pioneers in this growing field of study. Furthermore, these mathematical explorations will be aided greatly by the contribution of knowledgeable biologists who can help make connections between analytical insights and biological ones, help make models as biologically reasonable while favoring simplicity, and aid in the further development of new models and extensions beyond those discussed here. We hope this review will encourage new researchers to contribute to this exciting new direction of applied mathematics.

Appendices

A Derivations

A.1 Derivation of Fokker-Planck Equation

In this section, we derive a continuous-time, continuous-space model via the master equation

$$u(x, t + \tau) = \int_{\Omega} f(x, y, t; \tau) u(y, t) dy$$

(A.1)

in an unbounded domain $\Omega = \mathbb{R}^2$. Higher dimensional derivations can be considered in a similar fashion. The master equation above keeps track of the density function for the location of animals over time and
space via a conservation law. The function $f$ is a probability density function which describes the movement from a point $y$ to $x$ over the time interval $[t, t + \tau]$ (see Supplementary Material A.2 for a concrete example).

First, consider the case where an individual is released with its location given by the probability density function $u_0(x)$ at time $t = 0$. To make this more concrete, one could consider the specific case where an individual is released at a point $x_0 \in \Omega$. In this case, $u_0(x) = \delta(x-x_0)$. If we do not know the exact location of release, we use a general probability density function (PDF) $u_0(x)$ satisfying

$$1 = \int_{\Omega} u_0(y) dy.$$ 

We now use a Taylor series approach while taking the limit as $\tau \to 0^+$, see [3] for example. Starting from the master equation (A.1), reminding readers that $x = (x_1, x_2)$ is a vector, we define the new vector $z = x - y$. Then we write $f(x, y, t; \tau) =: f_z(z, y, t; \tau)$. Through this change of variables we obtain from (A.1)

$$u(x, t + \tau) = \int_{\Omega} f_z(z, x-z, t; \tau) u(x-z, t) dz.$$ (A.2)

Expanding the right hand side in space about 0, we find

$$u(x, t + \tau) = \int_{\Omega} f_z(z, x, t; \tau) u(x, t) dz$$
$$- \int_{\Omega} \left(z_1 \frac{\partial}{\partial x_1} (u(x, t) f_z(z, x, t; \tau)) + z_2 \frac{\partial}{\partial x_2} (u(x, t) f_z(z, x, t; \tau)) \right) dz$$
$$+ \int_{\Omega} \left(\frac{z_1^2}{2} \frac{\partial^2}{\partial x_1^2} (u(x, t) f_z(z, x, t; \tau)) + \frac{z_2^2}{2} \frac{\partial^2}{\partial x_2^2} (u(x, t) f_z(z, x, t; \tau)) \right) dz$$
$$+ \int_{\Omega} \left(z_1 z_2 \frac{\partial^2}{\partial x_1 \partial x_2} (u(x, t) f_z(z, x, t; \tau)) \right) dz + O(z^3),$$ (A.3)

where we have assumed that the mixed partial derivatives of $f_z u$ agree. Using the fact that $\int_{\Omega} f_z(z, \tau) u(x, t) dz = u(x, t)$, we may move this term to the left hand side of (A.3) and divide by $\tau > 0$ to obtain

$$\frac{u(x, t + \tau) - u(x, t)}{\tau} = -\frac{1}{\tau} \int_{\Omega} \left(z_1 \frac{\partial}{\partial x_1} (u(x, t) f_z(z, x, t; \tau)) + a_2 \frac{\partial}{\partial x_2} (u(x, t) f_z(z, x, t; \tau)) \right) dz$$
$$+ \frac{1}{\tau} \int_{\Omega} \left(\frac{z_1^2}{2} \frac{\partial^2}{\partial x_1^2} (u(x, t) f_z(z, x, t; \tau)) + \frac{z_2^2}{2} \frac{\partial^2}{\partial x_2^2} (u(x, t) f_z(z, x, t; \tau)) \right) dz$$
$$+ \frac{1}{\tau} \int_{\Omega} \left(z_1 z_2 \frac{\partial^2}{\partial x_1 \partial x_2} (u(x, t) f_z(z, x, t; \tau)) \right) dz$$
$$+ O(\tau^{-1} z^3).$$ (A.4)

Taking the limit as $\tau \to 0^+$, we find that

$$\frac{\partial u}{\partial t} = -\nabla \cdot (c(x, t) u) + \sum_{i,j=1}^2 \frac{\partial^2}{\partial x_i x_j} (d_{ij}(x, t) u),$$ (A.5)

where

$$c(x, t) := \lim_{\tau \to 0^+} \frac{1}{\tau} \int_{\Omega} z f_z(z, x, t; \tau) dz,$$ (A.6)

$$d_{ij}(x, t) := \lim_{\tau \to 0^+} \frac{1}{2\tau} \int_{\Omega} z_i z_j f_z(z, x, t; \tau) dz.$$ (A.7)

Note that we tacitly assume that all higher order terms vanish as $\tau \to 0^+$. Interested readers are directed to [3] and [33, Ch. 2.3] for comparable derivations and further discussion.
A.2 Derivation of Space Use Coefficients from Utilization Distribution Function

We derive a general form of the space use coefficients $c, d$ from the utilization distribution function

$$f(x, y, t; \tau) = \frac{K(x - y; \tau)w(A(x, t))}{\int_{\Omega} K(z - y; \tau)w(A(z, t))dz}, \quad (A.8)$$

with further considerations found in [33]. The kernel $K(x - y; \tau)$ is the spatial dispersal found in the absence of external factors (e.g., in a spatially uniform environment), where we assume $K$ is symmetric for simplicity.

The weighting function $w(A(x, t))$ describes how other factors, such as memories and environmental covariates, influence the movement behaviour towards the point $x$ and time $t$. This is perhaps the most important term in this derivation, as this is where we are able to explicitly include cognitive factors into the bias in movement behaviour from first principles.

Under the same change of variables as in Supplementary Material A.1, the utilization distribution function is given by

$$f_{z}(z, x, t; \tau) = \frac{K(z; \tau)w(A(z + x, t))}{\int_{\Omega} K(\tilde{z}; \tau)w(A(\tilde{z} + x, t))d\tilde{z}}, \quad (A.9)$$

and so if one performs another Taylor expansion of $w$ in space about 0 appearing in the numerator of (A.9), we see that

$$K(z; \tau)w(A(z + x, t)) \sim w(A(x, t))K(z; \tau) + z\nabla w(A(x, t)) + O(z^2). \quad (A.10)$$

Hence, under the assumption that the kernel $K$ is symmetric, it is readily seen that

$$\int_{\Omega} z_1 z_2 f_{z}(z, x; \tau)dz \sim w(A(x, t))\int_{\Omega} z_1 z_2 K(z; \tau)dz + O(z^3)$$

$$= 0 + O(z^3). \quad (A.11)$$

Hence, referring to (A.7) one finds that in fact $d_{ij} \equiv 0$ for $i \neq j$. When $i = j$, we further expand the denominator of (A.9) to obtain

$$\int_{\Omega} K(\tilde{z}; \tau)w(A(\tilde{z} + x, t))d\tilde{z} \sim w(A(x, t))$$

$$+ \frac{\Delta w(A(x, t))}{2!}M_2(\tau) + \ldots, \quad (A.12)$$

where

$$M_p(\tau) = \int_{\Omega} |a|^p K(a; \tau)da \quad (A.13)$$

is the $p$th moment of the dispersal kernel.

Notice that the first order term in (A.12) vanishes in the same way as $d_{ij}$ for $i \neq j$ since $K$ is symmetric, i.e. $M_1(\tau) = 0$. Similarly, the mixed terms also vanish. Putting expansions (A.10) and (A.12) into (A.7), we find in the limit that

$$d_{ii}(x, t) = \lim_{\tau \to 0^+} \frac{1}{2\tau} w(A(x, t))\frac{M_2(\tau)}{2} + \ldots$$

$$= \lim_{\tau \to 0^+} \frac{M_2(\tau)}{4\tau} + O(\tau)$$

$$= d \ (= \text{const}). \quad (A.14)$$
Using a similar procedure, one may insert expansions (A.10) and (A.12) into (A.6) to find

\[ c(x,t) = \lim_{\tau \to 0^+} \frac{1}{\tau} \frac{\nabla w(A(x,t)) M_2(\tau)}{w(A(x,t))} + \ldots \]

\[ = \lim_{\tau \to 0^+} \left( \frac{M_2(\tau)}{2} \frac{\nabla w(A(x,t))}{w(A(x,t))} \right) + O(\tau) \]

\[ = 2d \nabla \log(w(A(x,t))). \]  

(A.15)

Finally, we discuss the choice of weighting function \( w \) as it applies to knowledge-based animal movement. Indeed, the final form of the equation describing animal movement is, roughly, a diffusion equation with bias in movement given by advection up the gradient of the log of the weighting function \( w \). We now consider the end result under the assumption of exponential weighting of covariates, which is the most common choice of selection function appearing in the literature [24]. As described previously, this includes environmental factors as well as cognitive factors influencing movement. In such a case,

\[ w(x,t) \propto \exp \left( \sum_i \beta_i a_i(x,t) \right), \]

where \( \beta_i > 0 \) indicates attraction towards the auxiliary variable and \( \beta_i < 0 \) indicates repulsion. Thus, from the derivation of \( c(x,t) \) appearing in line (A.15), we see that

\[ c(x,t) = 2d \nabla \log(w) \]

\[ = 2d \sum_i \beta_i \nabla a_i(x,t). \]  

(A.16)

Thus, the final description of motion in two spatial dimensions is given by

\[ \frac{\partial u(x,t)}{\partial t} = d \Delta u(x,t) - 2d \nabla \cdot \left( u(x,t) \nabla \left( \sum_i \beta_i \nabla a_i(x,t) \right) \right). \]  

(A.17)

## B Comments on Mathematical Analysis

### B.1 Well-posedness of Existing Models

#### B.1.1 Implicit Memory Models

We begin with a discussion of some fundamental results (or lack thereof) in the implicit memory models introduced in Section 2.2. Given the relatively sparse literature available on these systems, it is easiest to describe the cases where existence is known, leaving the rest as a rather large group of open problems.

In general, model (2.1) has a large body of literature concerning the well-posedness of the problem under appropriate regularity assumptions on the function \( a(x,t) \). If \( a(x,t) \) is twice continuously differentiable, for example, a unique solution exists for all time. Thus, any model which provides \( a(x,t) \) explicitly enjoys well-posedness, essentially following from standard theory of linear partial differential equations, see e.g. [67, Ch. 8], [16], [11, Part II]. This means that models (2.4), (2.6), (2.7) have a unique, global solution, so long as the resource \( m(x,t) \) is sufficiently smooth. On the other hand, model (2.8) is nonlinear and existence is not so obvious.

The difficulty is increased substantially with models introduced in Section 2.2.2. Indeed, such systems are nonlinear, and moreover are nonlinear at a higher order (i.e. the coupling occurs inside the gradient). Proving the well-posedness of such models is highly nontrivial in general, with at least one exception. If the perceptual kernel \( g(\cdot) \) is sufficiently smooth, the regularity requirements for the nonlinear term can be transferred to the kernel itself, and so existence follows from standard \( L^p \)-theory of parabolic equations [67, Ch. 9]. Recently, the well-posedness of model (2.10) has been shown under the assumption that the perceptual kernel is twice continuously differentiable [19]. This was shown using semi-group theory [1, 40]. We note that a similar existence result should be possible using \( L^p \)-theory of parabolic equations when the
perceptual kernel is twice differentiable, and its second derivative belongs to $L^\infty(Q_T)$. Indeed, estimates found in [67, Ch. 9] hold somewhat trivially, since we may transfer the derivatives from the solution $u$ itself to the spatial kernel instead. Hence, the existence of a strong solution belonging to $W^{2,1}_{2,1}(Q_T)$ where $Q_T = \Omega \times (0, T)$ for a smooth, bounded domain $\Omega$ and $T > 0$ fixed follows from a mere $L^2(Q_T)$ estimate on the solution $u$. We leave further details of this solution method for future work.

To conclude, the well-posedness of any model featuring a Gaussian kernel enjoys well-posedness due to the regularity of the kernel. The exponential kernel is more difficult due to the lack of differentiability at zero. The most difficult case, and perhaps most biologically interesting, is the top-hat detection function, which is discontinuous. This fact is precisely what makes these models so difficult to study analytically, and new techniques or tools must be used. On the other hand, modellers may be satisfied with approximating the top-hat detection function through a regularization technique, e.g. mollification, in which case well-posedness is less of an issue and will follow from a careful application of existing theory.

### B.1.2 Explicit Memory Models

In contrast to the issues discussed above, explicit memory models (i.e. models with delay) found in Section 2.3 generally have more results concerning their well-posedness. In this sense, their complexity is offset by the existing literature for delay equations and do not often require a study of non-local (in space) effects. We highlight some key existing results below.

Model (2.19) has a unique, global solution under mild conditions, see [54, Proposition 2.1]. This follows from a standard bootstrapping method, assuming regular initial data and applying classical theory of parabolic equations.

[58] does not consider the existence of solutions to model (2.20), and is thus an open problem. This may follow from standard arguments, since the non-locality appears in the reaction term rather than inside the gradient.

Model (2.21) has a unique, global solution under mild conditions, see [52, Theorem 2.1]. This also follows from a standard bootstrapping argument with minor modifications to deal with two time delays.

[2] considers the existence of inhomogeneous steady states to model 2.22 under some stronger regularity assumptions on the growth term and non-local kernel appearing within it. This is achieved through a Lyapunov-Schmidt reduction, see [2, Theorem 2.1]. Existence of solutions to the time dependent problem is not considered, and so this is an open question.

[59] does not consider the existence of solutions to problem (2.23), and so this is an open question.

[55] does not explicitly consider the existence of solutions to model (2.27), however [55, Lemma 2.2] (the statement of which is found in the Appendix B.3.2) provides an interesting equivalence result between the delay differential equations and a certain form of Keller-Segel chemotaxis model. Hence, existence and non-existence of solutions may follow from the vast literature concerning Keller-Segel type systems.

[59] does not consider the existence of solutions to model (2.29), however some references concerning periodic solutions and travelling wave solutions are provided. The study of the well-posedness of this model is an open question.

To conclude, models which feature a spatiotemporal convolution, defined most generally in line (2.25), do not have many results concerning the well-posedness of solutions. Similar to the caveats made in Section B.1.1, existence is less of an issue when the kernels chosen are appropriately smooth. The most interesting cases, however, do not often feature such regularity (as in the case of the top-hat detection function, or when the kernel is taken to be the fundamental solution to the heat equation), and so further development of analytical tools is necessary.

### B.2 Pattern Formation through Linear Stability Analysis

#### B.2.1 Implicit Memory Models

We discuss some of the results found in [41] and the general trends based on dispersal relations (see [41, Figure 1]), at least as they are found in the case of two interacting species. These results are most closely related to model (2.12), however readers are reminded that the model appearing in [41] treats the memory component $k_i$ as a probability distribution as opposed to a magnitude alone (see the brief discussion in Section 2.2 before model (2.12) is introduced).
First, it is found that in the limit as the perceptual radius $R$ decreases, the set of wavenumbers at which patterns may form increases in size. In particular, in the limit as $R \to 0^+$, patterns may form at arbitrarily high wavenumbers, and so the problem becomes ill-posed. This indicates that a decreasing perceptual radius has a destabilizing effect on the constant steady states.

Another set of important parameters considered is the rate at which memories decay (denoted by $\mu$) or the rate at which foragers update their cognitive map should they revisit a site and deem it safe (denoted by $\beta$). It was found that if either $\mu$ is too large, or $\beta$ is too small, patterns cannot occur at any wavenumber. Hence, in order for territorial patterns to form, foragers cannot forget information too quickly, and they must have some mechanism to update their cognitive map based on new information. In the case of conflict zones, this means that foragers must have some mechanism through which they feel safe should they revisit a site and experience no conflict. Similarly, it was found that the set of wavenumbers at which patterns may form increases as the advection rate increases. This suggests that foragers must move quickly enough towards safe areas or away from conflict zones in order for patterns to occur.

On the other hand, it was found that the rate of diffusion and rate at which conflicts occur does not change the set of wavenumbers at which patterns may form, however the rate at which these patterns grow is smaller when the conflict rate $\rho$ is larger or the diffusion rate is smaller.

This exploration provides some preliminary insights into the relative effects these parameters have on the possibility of pattern formation, and some further investigation of these effects is explored in [42], where some simplifying assumptions allow the authors to unify models (2.10), (2.12) and (2.11) into one common form. These simplifications are convenient for their analysis, however it may be interesting to investigate the effects that differing parameter regimes may have without these unifying assumptions. In particular, one may consider the effects of different rates of diffusion, rates of advection, as well as the relative effects of the memory decay rates, without an appeal to a quasi steady-state approximation, without which these results may not hold. Furthermore, readers should note that in the case of three or more interacting populations, less is known, and the numerical results found in [42] suggest that the dynamics may be chaotic.

Finally, readers should note that the above holds when we know the existence of a positive steady state. This is usually easiest when constant steady states are a possibility, however this is highly dependent on the boundary conditions chosen. If a homogeneous Dirichlet boundary condition is used, for example, a constant steady state is no longer a possibility (other than the trivial solution, of course), and so an additional step must be executed: proving the existence of non-trivial steady states. This is the direction taken in [2], for example.

**Open Question:** How do different parameters change the insights obtained above? That is, unequal diffusion rates, advection rates, perceptual radii etc. without appealing to a simplification process?

**Open Question:** How do results change with more than two interacting populations? Some partial answers exist suggesting the dynamics are incredibly rich, and this is a highly non-trivial question in general.

### B.2.2 Explicit Memory Models

In the case of explicit memory models, i.e. models which feature an explicit time delay as found in Section 2.3, a linear stability analysis allows one to investigate the possibility of pattern formation similar to the method discussed above. However, due to the complex nature of delay differential equations, the resultant analysis of local stability of constant steady states is significantly more involved. A good introductory reference for partial delay differential equations is found in [66], however, readers should carefully note that all delay parameters appear in the lower order terms. For this reason, new tools and techniques need to be developed in the case of knowledge based movement models since the delay parameters appear at a higher order (i.e. inside the gradient term), increasing the difficulty substantially. Despite this, we may still discuss some of the key insights found in the use of temporal delays, which is primarily done through changes in stability of possible constant steady states (model (2.22) is the one exception due to the hostile boundary condition). Readers should note that these are general trends, and to cover every possible outcome in detail here is more challenging than in the cases without delays since the possible outcomes are much richer and can vary significantly across models. For this reason, we focus on the impacts of memory and cognition as opposed to focusing on other mechanisms included in the models.

We begin with the prototypical delay model (2.19). It was shown in [54] that the stability of the constant steady state depends on the ratio of diffusion rate and advection speed, but is independent of the discrete the
time delay $\tau$. Roughly, the advection speed away (or towards) high density areas must be sufficiently large in relation to the diffusion in order to destabilize the constant steady state. This suggests that the average time which the foragers reference back to does not influence the emergence of patterns, but there must be a mechanism by which the foragers move towards these preferred areas more quickly than the random diffusive movements. Intuitively, this makes sense: if the random motion is too large, this overtakes any possibility of aggregation/segregation and the population density will remain uniformly distributed throughout the environment.

Model (2.20) generalizes this through an inclusion of non-local effect in the growth of the population. Interestingly, [58] shows that the changes in stability remain roughly the same, however these changes are no longer independent of the delay parameter $\tau$. Similar to the previous results, the rate of advection must be sufficiently large in magnitude in order for destabilization of the constant steady state to occur. This is shown through the appearance of a Turing-Hopf or double Hopf bifurcation for some values of $\tau$ with respect to the advection speed. This is most easily viewed through [58, Figure 5] where a stability region is provided with respect to advection and delay parameters. Furthermore, there are many different forms of steady states that are possible (periodic solutions, non-constant steady states), as opposed to only a constant steady state. This demonstrates that a non-local effect in the growth term promotes a wider variety of potential outcomes in animal space use.

Model (2.21) is most similar to the first prototypical model in that the advection speed must be sufficiently large in comparison to the diffusion rate in order to destabilize the constant steady state, and this occurs independent of both delay parameters related to memory and maturation times. However, when the advection speed is small in relation to the diffusion rate, the memory delay parameter plays a key role in determining the stability of the constant steady state.

Model (2.22) is an exception to other delay model results, as the hostile boundary condition implies that a constant (non-trivial) steady state is not possible. Therefore, some results must first be shown to ensure that a non-trivial steady state exists. Once this was established, a similar trend appears to hold: destabilization of this steady state is induced by a sufficiently large advection speed. Numerically, it is observed that the hostile boundary condition results in “stripe” patterns, as opposed to “checkerboard” patterns as found in the case of zero-flux boundary conditions.

Different from the models discussed above, model (2.23) features both consumers as well as resources, and so it is fundamentally different than a single species model to begin with. In particular, this model features memory as well as dynamic resources, which may be more realistic than static or given resource densities apriori. Despite this, some of the general trends still hold: when the advection speed is small, there are no changes in stability of the constant steady state; when the advection rate is moderate, the delay parameter may have a destabilizing effect; when the advection speed is large, there is a critical delay value $\tau^*$ for which the constant steady state is stable when $\tau < \tau^*$, and unstable when $\tau > \tau^*$. Therefore, interaction between consumers and resources introduces a key difference from single species models in that the delay parameter can drastically change the long term dynamics. This is similar to the case of included maturation delays discussed earlier.

The final two models introduced feature distributed delays, as opposed to discrete delays. Model (2.27) features a single species model with distributed delay, as well as differing kernels determining how previous information influences movement. First, it is shown that the choices of kernels found in [58] correspond to an equivalence of systems, particularly a well know Keller-Segel system. This means that the analytical tools used to study the dynamics of Keller-Segel models can be used on this distributed delay model. Despite the departure from the discrete delay models, the same trend holds once more: the advection speed must be sufficiently large in order to destabilize the constant steady state.

Finally, model (2.29) includes a distributed delay in both memory as well as in the growth term. In this case, the destabilizing effect of advection speeds no longer holds, at least in the weak kernel case: when the advection speed is positive (segregation effect), patterns cannot occur; when the advection speed is negative (aggregation), bifurcations can occur at any advection speed, depending on the delay parameter $\tau$ appearing in the distributed memory delay kernel. In particular, this suggests that large advection speeds in magnitude is no longer sufficient to induce destabilization. Similarly, when the memory delay is held fixed, bifurcations occur in relation to the advection speed and maturation delay parameter $\sigma$. 

35
B.3 Unification and Equivalence of Existing Models

In some cases, it is possible to show the equivalence of certain models in specific scenarios. In other cases, some models can be subsumed in a unified form. We discuss these two aspects here.

B.3.1 Unification of Models

In order to combine these models into a cohesive form, one may generalize the spatial kernel introduced in (2.2) to include both spatial and temporal influences. To this end, suppose we are given a quantity $a(x,t)$ (e.g., a cognitive map). We can then define

$$
\bar{a}(x,t) := \int_{-\infty}^{t} \int_{\Omega} K(x,y,t,s)a(y,s)dyds,
$$

(B.1)

where $K$ is some reasonably defined space-time kernel describing modifications to the quantity $a(x,t)$ with respect to both distance and time.

As a special case, we may consider space-time kernels that are separable in their variables, i.e., $K(x,y,t,s) = g(x,y)G(t,s)$. We then define

$$
\bar{a}(x,t) := \int_{-\infty}^{t} \int_{\Omega} g(x-y)G(t-s)a(y,s)dyds.
$$

(B.2)

Given that this form separates the space and time variables through the kernels $g(x)$ and $G(t)$, one may consider two “averaging” processes so that combining both yield the form above. That is, define the linear operators $S$ and $T$ by

$$
S[a](x,t) := \int_{\Omega} g(x-y)a(y,t)dy,
$$

$$
T[a](x,t) := \int_{-\infty}^{t} G(t-s)a(x,s)ds.
$$

Assuming these integrals are well defined, we may take the composition

$$
T[S[a]](x,t) = T \left[ \int_{\Omega} g(x-y)a(y,t)dy \right]
$$

$$
= \int_{-\infty}^{t} \left( \int_{\Omega} g(x-y)a(y,s)dy \right) G(t-s)ds
$$

$$
= \bar{a}(x,t),
$$

to recover (B.2).

One may then relate all forms presented so far through various choices in $K(\cdot)$, $g(\cdot)$, $G(\cdot)$, as well as the quantity $a(x,t)$ itself. For example, any model presented that does not feature any integration of information over previous times, we may choose $g(t) = \delta(t)$. In doing so, (B.2) collapses to the form in (2.2), and we recover the models found in Section 2.2. When a discrete delay is included, as presented in Section 2.3.1, we may choose $G(t) = \delta(t-\tau)$ and $g(x) = \delta(x)$ since there is no perception included in those examples. In the case of distributed delays, one must consider the more general form (B.1) and choose $K$ to be a product of the Green’s function for the heat equation and the weak or strong kernel defined in (2.26), which recovers model (2.27) and (2.29) found in Section 2.3.2. The table below highlights these connections explicitly for most models introduced in this manuscript. In it, the advective potential $a(x,t)$ as found in (B.1) is described, along with the relevant kernels $g(x)$ and $G(t)$.

It should be noted that these generalizations do not necessarily lend themselves to a concrete mathematical analysis, however in some cases it may provide a stepping stone to results concerning the existence and uniqueness of solutions to the time-dependent problems.
| Model | Reference(s) | Adective Potential | Detection Kernel | Temporal Kernel |
|-------|--------------|--------------------|-----------------|-----------------|
| (2.4) | [13]         | Resource $m(x,t)$  | top-hat, exponential, Gaussian | $\delta(t)$ (local temporal information only) |
| (2.6) | [35, 25]     | Euclidean distance $\|x-x_0\|$ from den site located at $x_0$ | $\delta(x)$ (local spatial information only) | $\delta(t)$ (local temporal information only) |
| (2.7) | None         | Ratio of resource density over average resource density, $m(x,t)/\bar{m}$ | $\delta(x)$, top-hat, exponential, Gaussian | $\delta(t)$ (local temporal information only) |
| (2.8) | None         | Ratio of resource density over population density, $m(x,t)/u(x,t)$ | $\delta(x)$, top-hat, exponential, Gaussian | $\delta(t)$ (local temporal information only) |
| (2.10)| [42, 23]     | Density of all other species $u_i$, $i = 1, \ldots, n$ | top-hat         | $\delta(t)$ (local temporal information only) |
| (2.11)| [31, 42]     | Density of scent marks left on landscape by other populations, $\rho(x,t)$ | top-hat         | $\delta(t)$ (local temporal information only) |
| (2.12)| [41, 42]     | Cognitive map given by memory of conflict zones, $\kappa(x,t)$ | top-hat         | $\delta(t)$ (local temporal information only) |
| (2.13)| [57]         | Density of prey, $v(x,t)$; extensions include density of marks left on landscape (2.17) or memory of areas where resources were previously found (2.18) | top-hat         | $\delta(t)$ (local temporal information only) |
| (2.19), (2.20), (2.21), (2.22) | [54, 58, 52, 2] | Population density $u(x,t)$ | $\delta(x)$ (local spatial information only) | $\delta(t - \tau)$ (temporal information exactly $\tau > 0$ time units in the past) |
| (2.23)| [59]         | Density of prey, $v(x,t)$ | $\delta(x)$ (local spatial information only) | $\delta(t - \tau)$ (temporal information exactly $\tau > 0$ time units in the past) |
| (2.24)| [53]         | Density of own population and competitors, $u(x,t)$ and $v(x,t)$ | $\delta(x)$ (local spatial information only) | $\delta(t - \tau)$ (temporal information exactly $\tau > 0$ time units in the past) |
| (2.27), (2.29) | [55, 59] | Density of own population $u(x,t)$ | Green’s function for the heat equation, $G(x,y,t)$ (depends on both space and time) | Weak and strong temporal kernel, see (2.26) |
| (2.33)| None         | Memory compartment with both short- and long-term memory mechanisms, $a(x,t)$ | top-hat, exponential, Gaussian, general form as described in Section 2.1 | $\delta(t)$ (local temporal information only) |
B.3.2 Equivalence of Models

In some cases, the models presented in this review can be reformulated into an equivalent model. We first write the full problem studied in [54]:

\[
\begin{align*}
\frac{\partial u}{\partial t} &= d_1 \Delta u + d_2 \nabla \cdot (u \nabla v) + f(u), & x \in \Omega, & t > 0, \\
\frac{\partial u}{\partial n} &= 0, & x \in \partial \Omega, & t > 0, \\
u(x, t) &= \eta(x, t), & x \in \Omega, & -\infty < t \leq 0,
\end{align*}
\]  

(B.3)

in a smooth, bounded domain \( \Omega \), \( \eta(x, t) \) is given initial data, and \( v(x, t) \) is defined as

\[
v(x, t) = G \ast u = \int_{-\infty}^{t} \int_{\Omega} G(d_3, x, y, t-s)G(t-s)u(y, s)dyds,
\]

where \( G \) is the Green’s function for the heat equation in \( \Omega \) subject to homogeneous Neumann boundary data, and \( G \) is either the weak or strong kernel defined in (2.26). We first state Lemma 2.1 found in [55]. The lemma is stated as follows.

Lemma B.1. Suppose that kernel \( G(t) \) is chosen to be the weak kernel defined in (2.26) and define

\[
v(x, t) = G_w \ast u(x, t) = \int_{-\infty}^{t} \int_{\Omega} G(d_3, x, y, t-s)G_w(t-s)u(y, s)dyds,
\]

where \( G \) is the Green’s function for the heat equation subject to homogeneous Neumann boundary data. Then,

1. If \( u(x, t) \) is the solution of (B.3), then \( (u(x, t), v(x, t)) \) is the solution of

\[
\begin{align*}
\frac{\partial u}{\partial t} &= d_1 \Delta u + d_2 \nabla \cdot (u \nabla v) + f(u), & x \in \Omega, & t > 0, \\
\frac{\partial u}{\partial n} &= d_3 \Delta u + \tau^{-1}(u - v), & x \in \Omega, & t > 0, \\
u(x, t) &= \eta(x, t), & x \in \Omega, & t > 0, \\
v(x, 0) &= \tau^{-1} \int_{-\infty}^{0} \int_{\Omega} G(x, y, -s)e^{s\tau^{-1}}\eta(y, s)dyds.
\end{align*}
\]  

(B.4)

2. If \( (u(x, t), v(x, t)) \) is a solution of

\[
\begin{align*}
\frac{\partial u}{\partial t} &= d_1 \Delta u + d_2 \nabla \cdot (u \nabla v) + f(u), & x \in \Omega, & t \in \mathbb{R}, \\
\frac{\partial u}{\partial n} &= d_3 \Delta u + \tau^{-1}(u - v), & x \in \Omega, & t \in \mathbb{R}, \\
u(x, t) &= \eta(x, t), & x \in \Omega, & t \in \mathbb{R},
\end{align*}
\]  

(B.5)

then \( u(x, t) \) satisfies equation (B.3) such that \( \eta(x, s) = u(x, s), -\infty < s < 0 \). In particular, if \( (u(x), v(x)) \) is a steady state of (B.5), then \( u(x) \) is a steady state of (B.3); if \( (u(x), v(x)) \) is a periodic solution of (B.5), then \( u(x, t) \) is a periodic solution of (B.3).

This is an interesting result for two reasons. First, it is interesting to see that model (B.3) is actually equivalent to a Keller-Segel chemotaxis model when the weak kernel is chosen. Second, as a result of this first fact, one can then use the huge body of literature studying chemotaxis models in order to gain insights into this new delay partial differential equation. In the case where one chooses the strong kernel, there is another equivalent system consisting of 3 equations and similar insights can be gathered. This is Lemma 2.2 in [55], which we do not provide here.

References

[1] Amann, H. Linear and Quasilinear Parabolic Problems Volume I: Abstract Linear Theory. Birkhäuser Basel, 2021.
[2] An, Q., Wang, C., and Wang, H. Analysis of a spatial memory model with nonlocal maturation delay and hostile boundary condition. *Discrete and Continuous Dynamical Systems-A* 40(10) (2020), 5845–5868.

[3] Bharucha-Reid, A. *Elements of the Theory of Markov Processes and Their Applications*. McGraw-Hill, New York, 1960.

[4] Bracis, C., Gurarie, E., and Bram Van Moorter, R. Memory effects on movement behavior in animal foraging. *PloS one* 10, 8 (2015).

[5] Bracis, C., Gurarie, E., Van Moorter, B., and Goodwin, R. A. Memory effects on movement behavior in animal foraging chloe. *PloS ONE* 10 (2015), 1–21.

[6] Britton, N. F. Spatial structures and periodic traveling waves in an integro-differential reaction diffusion population model. *SIAM J. Anal. Math.* 50 (1990), 1663–1688.

[7] Cantrell, R., and Cosner, C. *Spatial Ecology via Reaction-Diffusion Equations*. Wiley Series in Mathematical & Computational Biology. Wiley, 2004.

[8] Chen, S., and Yu, J. Stability and bifurcations in a nonlocal delayed reaction-diffusion population model. *J. Differential Equations* 260 (2016), 218–240.

[9] Cho, E., and Kim, Y.-J. Starvation driven diffusion as a survival strategy of biological organisms. *Bull. Math. Biol.* 75 (2013), 845–870.

[10] DeAngelis, D. L., and Mooij, W. M. Individual-based modeling of ecological and evolutionary processes. *Annual Review of Ecology, Evolution, and Systematics* 36 (2005), 147–168.

[11] Evans, L., and Society, A. M. *Partial Differential Equations*. Graduate studies in mathematics. American Mathematical Society, 1998.

[12] Fagan, W., Lewis, M., Auger-Methe, M., Avgar, T., Benhamou, S., Breed, G., LaDage, L., Schlager, U., , w. Tang, W., Papastamatiou, Y., Forester, J., and Mueller, T. Spatial memory and animal movement. *Ecology Letters* 16 (2013), 1316–1329.

[13] Fagan, W. F., Gurarie, E., Bewick, S., Howard, A., Cantrell, R. S., and Cosner, C. Perceptual ranges, information gathering, and foraging success in dynamic landscapes. *The American Naturalist* 189, 5 (2017), 474–489.

[14] Feng, C., Lewis, M. A., Wang, C., and Wang, H. A fisher-kpp model with a nonlocal weighted free boundary, 2021.

[15] Fenton, M. B. Bats navigate with cognitive maps. *Science* 369, 6500 (2020), 142.

[16] Friedman, A. *Partial Differential Equations of Parabolic Type*. Prentice-Hall, 1964.

[17] Fronhofer, E. A., Hovestadt, T., and Poethke, H.-J. From random walks to informed movement. *Oikos* 122, 6 (2013), 857–866.

[18] Furter, J., and Grinfeld, M. Local vs. non-local interactions in population dynamics. *Journal of Mathematical Biology* 27, 1 (1989), 65–80.

[19] Giunta, V., Hillen, T., Lewis, M. A., and Potts, J. R. Local and global existence for non-local multi-species advection-diffusion models, 2021.

[20] Gourley, S. A., and So, J. W.-H. Dynamics of a food-limited population model incorporating nonlocal delays on a finite domain. *J. Math. Biol.* 44 (2002), 49–78.

[21] Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., Goss-Custard, J., Grand, T., Heinz, S. K., Huse, G., Huth, A., Jepsen, J. U., Jorgensen, C., Mooij, W. M., Muller, B., Pe‘er, G., Piou, C., Railsback, S. F., Robbins, A. M., Robbins, M. M., Rossmanith, E., Rüger, N., Strand, E., Souissi, S., Stillman, R. A., Vabo, R., Visser, U., and DeAngelis, D. L. A standard protocol for describing individual-based and agent-based models. *Ecological Modelling* 198, 1-2 (2006), 115–126.

[22] Hess, P. *Periodic-Parabolic Boundary Value Problems and Positivity*. John-Wiley and Sons, 1991.

[23] Hillen, T., and Painter, K. J. A user’s guide to PDE models for chemotaxis. *Journal of Mathematical Biology* 58(1-2) (2009), 183–217.
[24] J, J. F., J, J. S., B, B. S., and AVGAR, T. 'how to' guide for interpreting parameters in habitat-selection analyses. J. Anim. Ecol. 90 (2021), 1027–1043.

[25] JOHNSTON, S., AND PAINTER, K. The impact of short- and long-range perception on population movements. Journal of Theoretical Biology 460 (2019), 227–242.

[26] KAROWE, D., AND MARTIN, M. The effects of quantity and quality of diet nitrogen on the growth, efficiency of food utilization, nitrogen budget, and metabolic rate of fifth-instar spodoptera eridania larvae (lepidoptera: Noctuidae). J. Insect Physiol. 35 (1989), 699–708.

[27] KELLER, E. F., AND SEGEL, L. A. Initiation of slime mold aggregation viewed as an instability. Journal of Theoretical Biology 26(3) (1970), 399–415.

[28] KITCHIN, R., AND BLADES, M. The Cognition of Geographic Space, vol. 4. I. B. Tauris, 2002.

[29] LELE, S. R., AND KEIM, J. L. Weighted distributions and estimation of resource selection probability functions. Ecology 87 (2006), 3021–3028.

[30] LEWIS, M., FAGAN, W., AUGER-METHE, M., AND ET AL. Learning and animal movement. Preprint: Authorea (2020).

[31] LEWIS, M. A., AND MURRAY, J. D. Modelling territoriality and wolf-deer interactions. Nature 366 (1993), 738–740.

[32] LIN, H. Y., FAGAN, W. F., AND JABIN, P. E. Memory-driven movement model for periodic migrations. Journal of Theoretical Biology 508 (2021).

[33] MARLEY, J. Animal selection and movement analysis: Finding missed connections. Master’s thesis, University of Alberta, 2020.

[34] MOORCROF, P. R., MOORCROFT, P., AND LEWIS., M. A. Mechanistic home range analysis. Princeton University Press, 2006.

[35] MOORCROFT, P. R., LEWIS, M. A., AND CRABTREE, R. L. Home range analysis using a mechanistic home range model. Ecology 80 (1999), 1656–1665.

[36] OTHMER, H. G., DUNBAR, S. R., AND ALT, W. Models of dispersal in biological systems. J. Math. Biol. 26 (1988), 263–298.

[37] PAINTER, K. Mathematical models for chemotaxis and their applications in self-organisation phenomena. J. Theo. Biol. (2018).

[38] PAO, C. Nonlinear Parabolic and Elliptic Equations. Fems Symposium. Springer US, 1992.

[39] PATLAK, C. S. Random walk with persistence and external bias. Bulletin of Mathematical Biophysics 15, 3 (1953), 311–338.

[40] PAZY, A. Semigroups of linear operators and applications to PDEs. Springer-Verlag, 1983.

[41] POTTS, J. R., AND LEWIS, M. A. How memory of direct animal interactions can lead to territorial pattern formation. Journal of the Royal Society Interface 13 (2016).

[42] POTTS, J. R., AND LEWIS, M. A. Spatial memory and taxis-driven pattern formation in model ecosystems. Bull. Math. Biol. 81 (2019), 2725–2747.

[43] POTTS, J. R., AND SCHLÄGEL, U. Parametrising diffusion-taxis equations from animal movement trajectories using step selection analysis. Methods in Ecology and Evolution (2020).

[44] RANC, N., CAGNACCI, F., AND MOORCROFT, P. R. Memory drives the formation of animal home ranges: evidence from a reintroduction. bioRxiv preprint (2020), –.

[45] RANC, N., MOORCROFT, P. R., OSSI, F., AND CAGNACCI, F. Experimental evidence of memory-based foraging decisions in a large wild mammal. bioRxiv preprint (2020), –.

[46] RANTA, E., LUNDBERG, P., AND KAITALA, V. Resource matching with limited knowledge. OIKOS 86 (1999), 383–385.

[47] RENNER, I., AND WARTON, D. Equivalence of maxent and poisson point process models for species distribution modeling in ecology. Biometrics 69 (2013).
[48] RIOTTE-LAMBERT, L., BENHAMOU, S., AND CHAMAILLÉ-JAMMES, S. How memory-based movement leads to nonterritorial spatial segregation. *The American Naturalist* 185, 4 (2015), E103–E116.

[49] SALMANIW, Y., THOMPSON, P., WANG, H., WANG, X., MARLEY, J., AND LEWIS, M. Memory and learning in animal movement models. *In Preparation* (2022).

[50] SCHLÄGEL, U. E., AND LEWIS, M. A. Detecting effects of spatial memory and dynamic information on animal movement decisions. *Methods in Ecology and Evolution* 5, 11 (2014), 1236–1246.

[51] SCHLAGEL, U. E., MERRILL, E. H., AND LEWIS, M. A. Territory surveillance and prey management: Wolves keep track of space and time. *Ecol Evol* 7, 20 (2017), 8388–8405.

[52] SHI, J., WANG, C., AND WANG, H. Diffusive spatial movement with memory and maturation delays. *Nonlinearity* 32 (2019), 3188–3208.

[53] SHI, J., WANG, C., AND WANG, H. Spatial movement with diffusion and memory-based self-diffusion and cross-diffusion. *J. Diff. Eqs.* 305 (2021), 242–269.

[54] SHI, J., WANG, C., WANG, H., AND YAN, X. Diffusive spatial movement with memory. *Journal of Dynamics and Differential Equations* 32 (2019), 979–1002.

[55] SHI, Q., SHI, J., AND WANG, H. Spatial Movement with Distributed Memory. *Journal of Mathematical Biology* 82 (2021).

[56] SMOUSE, P. E., FOCARDI, S., MOORCROFT, P. R., KIE, J. G., FORESTER, J. D., AND MORALES, J. M. Stochastic modelling of animal movement. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365, 1550 (2010), 2201–2211.

[57] SONG, Y., WANG, H., AND WANG, J. Cognitive consumer-resource dynamics with nonlocal perception. *In Review* (2022).

[58] SONG, Y., WU, S., AND WANG, H. Spatiotemporal dynamics in the single population model with memory-based diffusion and nonlocal effect. *Journal of Differential Equations* 267 (2019), 6316–6351.

[59] SONG, Y., WU, S., AND WANG, H. Memory-based movement with spatiotemporal distributed delays in diffusion and reaction. *Applied Math. and Comp.* 404 (2021), 1–20.

[60] SPIEGEL, O., AND CROFOOT, M. C. The feedback between where we go and what we know — information shapes movement, but movement also impacts information acquisition. *Current Opinion in Behavioral Sciences* 12 (2016), 90–96.

[61] TANG, W., AND BENNET, D. A. Agent-based modeling of animal movement: a review. *Geography Compass* 4, 7 (2010), 682–700.

[62] TIETELBAUM, C., CONVERSE, S. J., FAGAN, W., BÖHNING-GAESE, K., O’HARA, R., LACY, A., AND MUELLER, T. Experience drives innovation of new migration patterns of whooping cranes in response to global change. *Nature Communications* 7 (2016).

[63] THRUN, S., AND PRATT, L. *Learning to Learn*. Springer US, 1998.

[64] TOLEDO, S., SHOHAMI, D., SCHIFFNER, I., LOURIE, E., ORCHAN, Y., BARTAN, Y., AND NATHAN, R. Cognitive map–based navigation in wild bats revealed by a new high-throughput tracking system. *Science* 369 (07 2020), 188–193.

[65] VAN MOORTER, B., VISSCHER, D., BENHAMOU, S., BÖRGER, L., BOYCE, M. S., AND GAILLARD, J.-M. Memory keeps you at home: a mechanistic model for home range emergence. *Oikos* 118, 5 (2009), 641–652.

[66] WU, J. *Theory and Applications of Partial Functional Differential Equations*. Applied Mathematical Sciences. Springer New York, 1996.

[67] WU, Z., YIN, J., AND WANG, C. *Elliptic & Parabolic Equations*. Elliptic & Parabolic Equations. World Scientific, 2006.

[68] ZOLLNER, P. Comparing the landscape level perceptual abilities of forest sciurids in fragmented agricultural landscapes. *Landscape Ecology* 15 (2000), 523–533.