

Mean-field analysis of multipopulation dynamics with label switching

*Original*

Mean-field analysis of multipopulation dynamics with label switching / Morandotti, M.; Solombrino, F.. - In: SIAM JOURNAL ON MATHEMATICAL ANALYSIS. - ISSN 0036-1410. - STAMPA. - 52:2(2020), pp. 1427-1462. [10.1137/19M1273426]

*Availability:*

This version is available at: 11583/2850824 since: 2020-11-02T17:51:30Z

*Publisher:*

Society for Industrial and Applied Mathematics Publications

*Published*

DOI:10.1137/19M1273426

*Terms of use:*

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*

GENERICO -- per es. Nature : semplice rinvio dal preprint/submitted, o postprint/AAM [ex default]

The original publication is available at <https://epubs.siam.org/doi/10.1137/19M1273426> / <http://dx.doi.org/10.1137/19M1273426>.

(Article begins on next page)

## MEAN-FIELD ANALYSIS OF MULTIPOPULATION DYNAMICS WITH LABEL SWITCHING\*

MARCO MORANDOTTI<sup>†</sup> AND FRANCESCO SOLOMBRINO<sup>‡</sup>

**Abstract.** The mean-field analysis of a multipopulation agent-based model is performed. The model couples a particle dynamics driven by a nonlocal velocity with a Markov-type jump process on the probability that each agent has of belonging to a given population. A general functional analytic framework for the well-posedness of the problem is established, and some concrete applications are presented, both in the cases of a discrete and continuous set of labels. In the particular case of a leader-follower dynamics, the existence and approximation results recently obtained in [G. Albi et al., *Math. Models Methods Appl. Sci.*, 29 (2019), pp. 633–679] are recovered and generalized as a byproduct of the abstract approach proposed.

**Key words.** kinetic equations, mean-field limits, continuity equations, jump processes, superposition principle

**AMS subject classifications.** 35Q91, 60J75, 37C10, 47J35, 58D25

**DOI.** 10.1137/19M1273426

**1. Introduction.** The concept of mean-field interaction, originally used in statistical physics by Kac [32] and then McKean [35] to describe the collisions between particles in a gas, has later proved to be a powerful tool to analyze the asymptotic behavior of systems of interacting agents. Recent applications range from biological, social, and economic phenomena [1, 11, 20] to automatic learning [13, 31] and optimization heuristics [26, 33]. The underlying idea is that the collective behavior of large systems of particles (agent-based models, usually consisting of a set of ODEs) can be efficiently treated by replacing the influence of all the other individuals in the dynamics on a given agent by a single averaged effect. From a mathematical point of view, this amounts to passing from a particle description to a kinetic description, consisting of a limit PDE whose unknown is the particle density distribution in the state space. The well-posedness of such models has therefore to be proven in spaces of measures (see, for instance, the results in [19]).

In some of the applications, the interacting agents are assumed to belong to a number of different species, or populations [3, 4, 22, 24, 28]. This is a useful modeling assumption, e.g., in the theory of mean-field games, or in control theory, where it can be used to distinguish informed agents steering pedestrians to leave unknown environments [17], or to highlight the influence of few key investors in the stock market (see the discussion in [12, introduction]). In a multipopulation setting, source (or sink) terms can be added to the model, to account for the “birth” and “death” rate of a single population (see, for instance, [40, sections 4–5]). Furthermore, even

---

\*Received by the editors July 9, 2019; accepted for publication (in revised form) February 3, 2020; published electronically March 30, 2020.

<https://doi.org/10.1137/19M1273426>

**Funding:** The work of the first author was supported by the Starting grant per giovani ricercatori of Politecnico di Torino. The work of the second author was supported by the project “Variational methods for stationary and evolution problems with singularities and interfaces” PRIN 2017 financed by the Italian Ministry of Education, University, and Research.

<sup>†</sup>Dipartimento di Scienze Matematiche, Politecnico di Torino, Torino, 10129, Italy (marco.morandotti@polito.it).

<sup>‡</sup>Dipartimento di Matematica e Applicazioni, Università di Napoli Federico II, Napoli, 80126, Italy (francesco.solombrino@unina.it).

in the case where the number of agents remains constant along the evolution, the possibility of having some individuals switching from one population to another has been considered [41].

As a matter of fact, exchange rates among populations are a common feature in several applications. For instance, in models of chemical reaction networks a particle may change its type as a result of the interaction with the others, at a stochastic rate which may also depend on its position [34, 36, 38]. Another relevant example comes from social dynamics, where transition rates appear in some kinetic models of opinion formation in the presence of strong leaders, as the one proposed in [27], inspired by the earlier contribution [42]. In this model, the opinion of the agents, described by their position in the state space, evolves because of the exchange of opinion with the others. This is encoded by a nonlocal transport term taking into account the presence, among the overall population, of a restricted number of “leaders” promoting their opinion with a strong influence on the “followers.” It is natural to postulate, as the authors do (see [27, section 3.b] for a detailed discussion), that opinion leadership is not constant over time: someone who is an opinion leader today may lose this role tomorrow, or a follower may become a leader in the future.

Summarizing, the multipopulation dynamics we are interested in attaches to an agent sitting at a position  $x$  a probability measure  $\lambda \in \mathcal{P}(U)$ , where  $U$  is a compact space of labels accounting for the population to which the agent belongs. In fact, we find it natural not to assume any deterministic knowledge of the label of a single agent, since transitions are usually modeled as the outcome of a stochastic process. Hence,  $\lambda$  expresses the probability that the agent has, at a certain moment in time, of belonging to a subset of  $U$ , and may itself evolve as a consequence of the interaction among the agents. The rate of change of the position of an agent, as well as the transition rates among different populations, may be influenced by the global state of the system. For instance, in [27] if the mass of leaders sharing a similar opinion (i.e., sitting in the same small region of space) exceeds a certain threshold, the transition rates from leaders to followers becomes higher. Similarly, the velocity of the system, inducing the spatial evolution of the agents (their change of opinion, in the model proposed in [27]) is the result of pairwise interactions which depend on the mutual positions and are weighted differently depending on the actions of members of different populations (leaders and followers in [27]).

The minimal model for evolution that we propose<sup>1</sup> consists of the coupling of a nonlocal transport dynamics with a Markov-type jump process. We, namely, assume that

- a particle at position  $x_i$  with probability distribution of labels  $\lambda_i$  experiences a velocity field

$$\dot{x}_i = v_{\Psi^N}(x_i, \lambda_i)$$

influenced by the global state of the system  $\Psi^N$ , the empirical measure  $\frac{1}{N} \sum_{i=1}^N \delta_{(x_i, \lambda_i)}$ ;

- the probability distribution of labels  $\lambda_i$  evolves according to

$$\dot{\lambda}_i = \mathcal{T}_{\Psi^N}(x_i, \lambda_i),$$

where the operator  $\mathcal{T}_{\Psi^N}$ , accounting for the transitions among the labels, may depend on the position of the particle and on the state of the system.

<sup>1</sup>We neglect in the present contribution the possible presence of diffusion terms.

In the case of two populations  $U = \{F, L\} = \{\text{followers, leaders}\}$ , a model of this type has been recently analyzed in [2], as a (partial) discrete counterpart of the PDE model in [27]. Some simplifying assumptions had, however, to be added in order to perform a mean-field analysis. In particular, the velocity field

$$(1.1a) \quad v_{\Psi^N}(x_i) = \frac{1}{N} \sum_{j=1}^N K^F(x_i - x_j) \lambda_j(\{F\}) + \frac{1}{N} \sum_{j=1}^N K^L(x_i - x_j) \lambda_j(\{L\})$$

did not depend on the probability of the labels  $\lambda_i$  and the transition operators

$$(1.1b) \quad \begin{aligned} (\mathcal{T}_{\Psi^N}(\lambda_i))(\{F\}) &= -\alpha_F(\Psi^N) \lambda_i(\{F\}) + \alpha_L(\Psi^N)(1 - \lambda_i(\{F\})), \\ (\mathcal{T}_{\Psi^N}(\lambda_i))(\{L\}) &= \alpha_F(\Psi^N) \lambda_i(\{F\}) - \alpha_L(\Psi^N)(1 - \lambda_i(\{F\})) \end{aligned}$$

did not depend on the position  $x_i$ . The goal of the present paper is to get rid of these simplifying assumptions, providing an *appropriate functional setting* as well as a *general set of assumptions* on the velocity field  $v_{\Psi}$  and on the operator  $\mathcal{T}_{\Psi}$  (see section 3) which allow one to perform a rigorous mean-field analysis of agent-based models of the kind discussed above.

In order to develop our analysis, we will borrow some functional-analytic and some measure-theoretic tools that have been recently specified for the different, although related, context of *spatially inhomogeneous evolutionary games* in [5]. The analogue of (1.1) is described in [5, section 3.2], where the dynamics considered is

$$(1.2) \quad \begin{cases} \dot{x}_i = a(x_i, \sigma_i) = \int_U e(x_i, u) \, d\sigma_i(u), \\ \dot{\sigma}_i = \left( \frac{1}{N} \sum_{j=1}^N \left( \int_U J(x_i, \cdot, x_j, u') \, d\sigma_j(u') \right. \right. \\ \left. \left. - \int_U \int_U J(x_i, w, x_j, u') \, d\sigma_j(u') \, d\sigma_i(w) \right) \right) \sigma_i, \end{cases}$$

for  $x \in \mathbb{R}^d$  tracking the position of a player and  $\sigma \in \mathcal{P}(U)$  denoting their mixed strategy. In (1.2), the velocity  $\dot{x}_i$  is determined by the  $x$ -dependent vector field  $e(x, u)$  through an averaging process with respect to the strategies available to the player  $x_i$ ; the evolution of the mixed strategy  $\sigma_i$  is dictated by the *replicator equation* (see [30]): the term in the parentheses in the right-hand side of the second equation above determines the performance of the strategy played by player  $x_i$  with respect to all of the available strategies, averaging out all of the possible distributions of the opponents  $x_j$  with their mixed strategies  $\sigma_j$ . In particular,  $J(x_i, u, x_j, u')$  is the payoff of the game between player  $x_i$  with strategy  $u$  against player  $x_j$  with strategy  $u'$ . For the heuristic interpretation and the modeling issues related to (1.2) we refer the interested reader to [15, 18].

The paper [5] introduced suitable notions (namely, Lagrangian and Eulerian) of solutions to the mean-field limit of (1.2) showing convergence of the particle model to the limit description by means of stability estimates in the Wasserstein metric arising from the well-posedness theory of ODEs in Banach spaces [16, 21, 23].

While sharing some common features, the dynamics described by (1.1) and (1.2) show a relevant difference in the structures of the velocity fields in the right-hand sides. Indeed, in our setting, it is natural to postulate that the velocity of the agents is also

depending on the behavior of the ones around them, at least in a small neighborhood. Such a feature is not present in the first equation in (1.2), whereas it is encoded in our model by considering a velocity field  $v_\Psi$  depending on the global state of the system. Beside this main distinction, the right-hand sides of the equations for  $\lambda_i$  are more general than those for  $\sigma_i$  since we do not need any quadratic dependence on  $\lambda_i$  but only that some structural assumptions be satisfied. These are in particular fulfilled by the operators in (1.2), as we discuss in Proposition 5.8.

The novel contribution of this work is in providing well-posedness and mean-field analysis for our particle system (1.1) by suitably adapting the tools from [5] to our case. In doing this, we will pursue a more abstract point of view than the one considered in [5], by focusing on the structural assumptions on the velocity field  $v_\Psi$  and on the operator  $\mathcal{T}_\Psi$  that guarantee the well-posedness of the mean-field analysis. However, we will show in sections 4 and 5 that velocities and transition rates modeled by interaction kernels as those considered in (1.1) and (1.2) (in this last case, with the difference that we pointed out before) also fit in the setting that we propose in section 3. Thus, we can retrieve an even more general version of the results of [2] as a by-product of our analysis (see Theorem 4.8, where an explicit dependence on the space variable  $x$  of the transition rates is also allowed).

In the last part of this introduction, we briefly sketch the mathematical framework in which we formulate our results for the benefit of the interested reader. The general assumptions on  $v_\Psi$  and  $\mathcal{T}_\Psi$  are presented at the beginning of section 3; see (v1)–(v3) and (T0)–(T3) below. The dependence of  $v_\Psi$  on the global status of the system calls for a locally Lipschitz dependence both on the status variable  $(x, \lambda)$  (in a suitable topology) and on  $\Psi$  with respect to the Wasserstein distance between probability measures, which is customary for the mean-field analysis of transport equations with nonlocal vector fields; see, e.g., [6, 25, 29, 39, 43], also [14] in connection with optimal control problems. Furthermore, the sublinearity condition (v3) guarantees long-time existence of the solution. Concerning the operator  $\mathcal{T}_\Psi$ , it must comply with analogous local Lipschitz continuity conditions and sublinearity conditions; see (T1)–(T2) below; in addition, we require that

- constants belong to the kernel of  $\mathcal{T}_\Psi$ ;
- there exists  $\delta > 0$  such that  $\mathcal{T}_\Psi(x, \lambda) + \delta\lambda \geq 0$ .

The two conditions above (see (T0) and (T3) below for a precise statement) ensure that the velocities  $\dot{\lambda}_i$ 's lie on the tangent plane to the probabilities on  $\{F, L\}$  and that the positivity of the measures are preserved, respectively. Altogether, conditions (v1)–(v3) and (T0)–(T3) allow us to show that the particle model, which we can rewrite in the form

$$(1.3) \quad \dot{y}_i = \begin{pmatrix} \dot{x}_i \\ \dot{\lambda}_i \end{pmatrix} = \begin{pmatrix} v_{\Psi^N}(x_i, \lambda_i) \\ \mathcal{T}_{\Psi^N}(x_i, \lambda_i) \end{pmatrix} =: b_{\Psi^N}(y_i),$$

is well-posed (see Proposition 3.3).

We subsequently show that in the limit as  $N \rightarrow \infty$  the empirical measures  $\Psi^N$  converge to a continuous path of probability measures  $\Psi_t$  that solves the continuity-type equation

$$(1.4) \quad \partial_t \Psi_t + \operatorname{div}(b_{\Psi_t} \Psi_t) = 0$$

in the space of probability measures on the product space  $Y = \mathbb{R}^d \times \mathcal{P}(U)$ .

A key point of the proof is that the solutions to (1.4), which are defined via duality with test functions (Eulerian solutions; see Definition 3.4), can be equivalently

characterized as Lagrangian solutions satisfying the fixed point equation

$$(1.5) \quad \Psi_t = \mathbf{Y}_\Psi(t, 0, \cdot) \# \bar{\Psi} \quad \text{for every } t \geq 0,$$

where  $\mathbf{Y}(t, 0, \cdot)$  is the transition map (see Definition 3.7) associated with the ODE (1.3) (with  $\Psi$  in place of  $\Psi^N$ ),  $\bar{\Psi}$  is a given initial distribution of agents and probability on the labels, and the symbol  $\#$  denotes the push-forward measure (see Definition 2.1).

In section 4 we also retrieve and extend the main result of [2] as a special case of our analysis. We associate with a solution to (1.4) the followers and leaders distributions

$$(1.6) \quad \mu_\Psi^F(B) := \int_{B \times \mathcal{P}(\{F, L\})} \lambda(\{F\}) \, d\Psi(x, \lambda), \quad \mu_\Psi^L(B) := \int_{B \times \mathcal{P}(\{F, L\})} \lambda(\{L\}) \, d\Psi(x, \lambda),$$

for each Borel set  $B \subset \mathbb{R}^d$ . If the vector field  $v_\Psi$  has the special structure (1.1a) and under suitable structural assumptions on the transition rates  $\alpha_F$  and  $\alpha_L$  (see (4.7) below), using the definition of Eulerian solution, we are able to show that  $\mu_\Psi^F$  and  $\mu_\Psi^L$  are indeed the unique solutions to the system

$$(1.7) \quad \begin{cases} \partial_t \mu_t^F = -\operatorname{div}((K^F \star \mu_t^F + K^L \star \mu_t^L) \mu_t^F) - \alpha_F(x, \mu_t^F, \mu_t^L) \mu_t^F + \alpha_L(x, \mu_t^F, \mu_t^L) \mu_t^L, \\ \partial_t \mu_t^L = -\operatorname{div}((K^F \star \mu_t^F + K^L \star \mu_t^L) \mu_t^L) + \alpha_F(x, \mu_t^F, \mu_t^L) \mu_t^F - \alpha_L(x, \mu_t^F, \mu_t^L) \mu_t^L \end{cases}$$

considered in [2]. Notice that, as discussed in Remark 4.6, rewriting (3.28) as a system of equations in  $\mu^F$  and  $\mu^L$  is not possible for a velocity field explicitly depending on  $\lambda$  such as

$$\begin{aligned} & v_{\Psi^N}(x_i, \lambda_i) \\ &= \lambda_i(\{F\}) \left( \frac{1}{N} \sum_{j=1}^N K^{FF}(x_i - x_j) \lambda_j(\{F\}) + \frac{1}{N} \sum_{j=1}^N K^{LF}(x_i - x_j) \lambda_j(\{L\}) \right) \\ &+ \lambda_i(\{L\}) \left( \frac{1}{N} \sum_{j=1}^N K^{FL}(x_i - x_j) \lambda_j(\{F\}) + \frac{1}{N} \sum_{j=1}^N K^{LL}(x_i - x_j) \lambda_j(\{L\}) \right) \end{aligned}$$

considered for instance in [24]. In the presence of exchange rates, the correct mean-field description of the above particle system is given by (3.28) in the product space  $\mathbb{R}^d \times \mathcal{P}_1(\{F, L\})$ .

In subsection 5.1 we consider the case  $U = \{1, \dots, H\}$  and prove that the results of the leader-follower dynamics can be generalized to the discrete setting of an arbitrary finite number  $H > 2$  of populations. In particular, an existence result for a system of the kind (see (5.13) below)

$$\partial_t \mu_t^h = -\operatorname{div} \left( \left( \sum_{k \in U} K^k \star \mu_t^k \right) \mu_t^h \right) - \alpha_{hh}(x, \mu_t^1, \dots, \mu_t^H) \mu_t^h + \sum_{k \neq h} \alpha_{kh}(x, \mu_t^1, \dots, \mu_t^H) \mu_t^k$$

for  $h, k = 1, \dots, H$ , is proved. In the system above, our assumptions (T0) and (T3) amount to requiring that the transition rates  $\alpha_{hk}$  build, for fixed  $x$  and  $\mu_t^1, \dots, \mu_t^H$ , a so-called Q-matrix, whose use is customary in the theory of Markov chains (see Remark 5.7). We also remark that the situation we consider is, however, more general,

as we may allow for a possibly nonlinear dependence of  $\alpha_{hk}$  on the space variable and of the state of the system.

The paper is organized as follows. In section 2, we recall the basic notions of measure theory that will be needed in the following, and prove a corollary of a theorem by Brezis [16, section I.3, Theorem 1.4, Corollary 1.1] on the well-posedness of ODEs in Banach spaces. In section 3 we introduce our abstract model and we prove our main result, Theorem 3.5, on the mean-field limit of the dynamics. Section 4 is devoted to the special case of  $U = \{F, L\}$ , modeling the leader-follower dynamics. We apply the abstract results to this case and recover the results of [2], extending them to the case of  $x$ -dependent transition rates. Finally, subsection 5.1 we extend the results of section 4 to any finite numbers of labels, whereas in subsection 5.2 we discuss some explicit examples of velocity fields  $v_\Psi$  and transition operators  $\mathcal{T}_\Psi$  which are encompassed by our setting in the case of a continuum of labels, and we compare them with those considered in [5].

## 2. Preliminaries.

**2.1. Basic notation.** If  $(X, d_X)$  is a metric space, we denote by  $\mathcal{M}(X)$  the space of signed Borel measures in  $X$  with finite total variation; by  $\mathcal{M}_+(X)$  and  $\mathcal{P}(X)$  the convex subsets of nonnegative measures and probability measures, respectively. The notation  $\mathcal{P}_c(X)$  will be used for measures having compact support in  $X$ . For  $\mu \in \mathcal{M}(X)$ ,  $|\mu|$  denotes the total variation of  $\mu$ . If we denote by  $C_0(X)$  the space of continuous functions vanishing at the boundary of  $X$ , and by  $C_b(X)$  the space of bounded continuous functions, the weak\* and narrow convergence in  $\mathcal{M}(X)$  are defined by the convergence of the duality products

$$\int_X \phi \, d\mu_h \rightarrow \int_X \phi \, d\mu, \quad h \rightarrow \infty,$$

for each  $\phi \in C_0(X)$  and  $\phi \in C_b(X)$ , respectively.

Whenever  $X = \mathbb{R}^d$ ,  $d \geq 1$ , it remains understood that it is endowed with the Euclidean norm (and induced distance), which shall be simply denoted by  $|\cdot|$ .

For a Lipschitz function  $f: X \rightarrow \mathbb{R}$  we denote by

$$\text{Lip}(f) := \sup_{\substack{x, y \in X \\ x \neq y}} \frac{|f(x) - f(y)|}{d_X(x, y)}$$

the Lipschitz constant. The notations  $\text{Lip}(X)$  and  $\text{Lip}_b(X)$  will be used for the spaces of Lipschitz and bounded Lipschitz functions on  $X$ , respectively. Both are normed spaces with the norm  $\|f\| := \|f\|_\infty + \text{Lip}(f)$ , where  $\|\cdot\|_\infty$  denoted the supremum norm.

In a complete and separable metric space  $(X, d_X)$ , we shall use the Kantorovich–Rubinstein distance  $\mathcal{W}_1$  in the class  $\mathcal{P}(X)$ , defined as

$$\mathcal{W}_1(\mu, \nu) := \sup \left\{ \int_X \varphi \, d\mu - \int_X \varphi \, d\nu : \varphi \in \text{Lip}_b(X), \text{Lip}(\varphi) \leq 1 \right\}$$

or, equivalently (thanks to the Kantorovich duality), as

$$\mathcal{W}_1(\mu, \nu) := \inf \left\{ \int_{X \times X} d_X(x, y) \, d\Pi(x, y) : \Pi(A \times X) = \mu(A), \Pi(X \times B) = \nu(B) \right\},$$

involving couplings  $\Pi$  of  $\mu$  and  $\nu$ . Notice that  $\mathcal{W}_1(\mu, \nu)$  is finite if  $\mu$  and  $\nu$  belong to the space

$$\mathcal{P}_1(X) := \left\{ \mu \in \mathcal{P}(X) : \int_X d_X(x, \bar{x}) d\mu(x) < +\infty \text{ for some } \bar{x} \in X \right\}$$

and that  $(\mathcal{P}_1(X), \mathcal{W}_1)$  is complete if  $(X, d_X)$  is complete. For a positive measure  $\mu \in \mathcal{M}_+(E)$ , for  $E$  being a Banach space, we define the first moment  $m_1(\mu)$  as

$$m_1(\mu) := \int_E \|x\|_E d\mu.$$

Notice that, for a probability measure  $\mu$ , finiteness of the integral above is equivalent to  $\mu \in \mathcal{P}_1(E)$ , whenever  $E$  is endowed with the distance induced by the norm  $\|\cdot\|$ .

We now recall the definition of push-forward measure.

**DEFINITION 2.1.** *Let  $\mu \in \mathcal{M}_+(X)$  and  $f: X \rightarrow Z$  a  $\mu$ -measurable function be given. The push-forward measure  $f_{\#}\mu \in \mathcal{M}_+(Z)$  is defined by  $f_{\#}\mu(B) = \mu(f^{-1}(B))$  for any Borel set  $B \subset Z$ . The push-forward measures has the same total mass as  $\mu$ , namely,  $\mu(X) = f_{\#}\mu(Z)$ . It also holds the change of variables formula*

$$\int_Z g df_{\#}\mu = \int_X g \circ f d\mu$$

whenever either one of the integrals is well defined.

For  $E$  being a Banach space, the notation  $C_b^1(E)$  will be used to denote the subspace of  $C_b(E)$  of functions having bounded continuous Fréchet differential at each point. The notation  $D\phi(\cdot)$  will be used to denote the Fréchet differential. In the case of a function  $\phi: [0, T] \times E \rightarrow \mathbb{R}$ , the symbol  $\partial_t$  will be used to denote partial differentiation with respect to  $t$ , while  $D$  will only stand for differentiation with respect to the variables in  $E$ .

**2.2. Functional setting.** The space of labels  $U$  will be assumed to be a compact metric space. The state variable of the system is  $y := (x, \lambda) \in \mathbb{R}^d \times \mathcal{P}(U) =: Y$ . The component  $x \in \mathbb{R}^d$  describes the location of an agent in space, whereas the component  $\lambda \in \mathcal{P}(U)$  describes the distribution of labels of the agent. A probability distribution  $\Psi \in \mathcal{P}(Y)$  denotes a distribution of agents with labels.

To define the functional setting for the dynamics, we need to consider the *free space*

$$(2.1) \quad \mathcal{F}(U) := \overline{\text{span}(\mathcal{P}(U))}^{\|\cdot\|_{\text{BL}}} \subset (\text{Lip}(U))',$$

which is also called in the literature the Arens–Eells space; see [8, 10] and [44, Chapter 3]. The closure in (2.1) is taken with respect to the *bounded Lipschitz* norm  $\|\cdot\|_{\text{BL}}$ , which is defined, for  $\mu \in (\text{Lip}(U))'$ , by

$$\|\mu\|_{\text{BL}} := \sup \{ \langle \mu, \varphi \rangle : \varphi \in \text{Lip}(U), \|\varphi\|_{\text{Lip}} \leq 1 \}.$$

With the free space  $\mathcal{F}(U)$  at hand, we define  $\bar{Y} := \mathbb{R}^d \times \mathcal{F}(U)$  and the norm  $\|\cdot\|_{\bar{Y}}$  by

$$(2.2) \quad \|y\|_{\bar{Y}} = \|(x, \lambda)\|_{\bar{Y}} := |x| + \|\lambda\|_{\text{BL}}.$$

For a given  $R > 0$ , we denote by  $B_R$  the closed ball of radius  $R$  in  $\mathbb{R}^d$  and by  $B_R^Y$  the ball of radius  $R$  in  $Y$ , namely,  $B_R^Y = \{y \in Y : \|y\|_{\bar{Y}} \leq R\}$ , and observe that it



is a compact set, since  $Y$  is locally compact by our assumptions on  $U$ . The Banach space structure of  $\bar{Y} \supset Y$  allows us to define the first moment  $m_1(\Psi)$  for a probability measure  $\Psi \in \mathcal{P}(Y)$  as

$$m_1(\Psi) := \int_Y \|y\|_{\bar{Y}} d\Psi$$

so that the space  $\mathcal{P}_1(Y)$  can be equivalently characterized as

$$\mathcal{P}_1(Y) = \{\Psi \in \mathcal{P}(Y) : m_1(\Psi) < +\infty\}.$$

We will sometimes use the notation  $\mathcal{P}(K)$  to denote probability measures with support contained on a given compact subset  $K \subset Y$ . Notice that trivially we have  $\mathcal{P}(K) \subset \mathcal{P}_1(Y)$ .

**2.3. Well-posedness of ODEs in Banach spaces.** We recall here a theorem by Brezis [16, section I.3, Theorem 1.4, Corollary 1.1] on the well-posedness of ODEs in Banach spaces.

**THEOREM 2.2.** *Let  $(E, \|\cdot\|_E)$  be a Banach space,  $C$  a closed convex subset of  $E$ , and let  $A(t, \cdot) : C \rightarrow E$ ,  $t \in [0, T]$ , be a family of operators satisfying the following properties:*

- (i) *there exists a constant  $L \geq 0$  such that for every  $c_1, c_2 \in C$  and  $t \in [0, T]$*

$$(2.3) \quad \|A(t, c_1) - A(t, c_2)\|_E \leq L\|c_1 - c_2\|_E;$$

- (ii) *for every  $c \in C$  the map  $t \mapsto A(t, c)$  is continuous in  $[0, T]$ ;*  
 (iii) *for every  $R > 0$  there exists  $\theta > 0$  such that*

$$(2.4) \quad c \in C, \|c\|_E \leq R \quad \Rightarrow \quad c + \theta A(t, c) \in C.$$

*Then for every  $\bar{c} \in C$  there exists a unique curve  $c : [0, T] \rightarrow C$  of class  $C^1$  satisfying  $c_t \in C$  for all  $t \in [0, T]$  and*

$$(2.5) \quad \frac{d}{dt}c_t = A(t, c_t) \quad \text{in } [0, T], \quad c_0 = \bar{c}.$$

*Moreover, if  $c^1, c^2$  are the solutions starting from the initial data  $\bar{c}^1, \bar{c}^2 \in C$ , respectively, we have*

$$(2.6) \quad \|c_t^1 - c_t^2\|_E \leq e^{Lt}\|\bar{c}^1 - \bar{c}^2\|_E \quad \text{for every } t \in [0, T].$$

For our purposes, we need the following generalization where assumption (i) in Theorem 2.2 is weakened to a local Lipschitz continuity condition, provided we additionally assume at most linear growth of the operator  $A$ .

**COROLLARY 2.3.** *Let hypotheses (ii) and (iii) of Theorem 2.2 hold for a family of operators  $A(t, \cdot) : C \rightarrow E$ ,  $t \in [0, T]$ . Assume, in addition, that*

- (i') *for every  $R > 0$  there exists a constant  $L_R \geq 0$  such that for every  $c_1, c_2 \in C \cap B_R$  and  $t \in [0, T]$ ,*

$$(2.7) \quad \|A(t, c_1) - A(t, c_2)\|_E \leq L_R\|c_1 - c_2\|_E;$$

- (i'') *there exists  $M > 0$  such that for every  $c \in C$ , there holds*

$$(2.8) \quad \|A(t, c)\|_E \leq M(1 + \|c\|_E).$$

Then for every  $\bar{c} \in C$  there exists a unique curve  $c: [0, T] \rightarrow C$  of class  $C^1$  satisfying  $c_t \in C$  for all  $t \in [0, T]$  and

$$(2.9) \quad \frac{d}{dt}c_t = A(t, c_t) \quad \text{in } [0, T], \quad c_0 = \bar{c}.$$

Moreover, if  $c^1, c^2$  are the solutions starting from the initial data  $\bar{c}^1, \bar{c}^2 \in C \cap B_R$ , respectively, there exists a constant  $L = L(M, R, T) > 0$  such that

$$(2.10) \quad \|c_t^1 - c_t^2\|_E \leq e^{Lt} \|\bar{c}^1 - \bar{c}^2\|_E \quad \text{for every } t \in [0, T].$$

*Proof.* Let us fix the initial datum  $\bar{c} \in C$ , and let us choose  $\bar{R} := (\|\bar{c}\|_E + MT)e^{MT}$ . Consider a smooth function with compact support  $\chi: \mathbb{R}^+ \rightarrow [0, 1]$  such that  $\chi(r) = 1$  for every  $r \leq \bar{R}$  and set  $B(t, c) := \chi(\|c\|_E)A(t, c)$ . Then one can see that  $B$  satisfies hypotheses (i) and (ii) of Theorem 2.2. To see that hypothesis (iii) is also satisfied, it suffices to notice that, by convexity and since  $0 \leq \chi \leq 1$ ,  $c + \theta\chi(\|c\|_E)A(t, c) \in C$  whenever  $c + \theta A(t, c) \in C$ . Therefore there exists a unique solution  $t \mapsto c(t)$  of class  $C^1$  of

$$(2.11) \quad \frac{d}{dt}c_t = B(t, c_t) \quad \text{in } [0, T], \quad c_0 = \bar{c}.$$

Using again that  $0 \leq \chi \leq 1$  and (2.8), one can see that

$$(2.12) \quad \|c_t\|_E \leq \|\bar{c}\|_E + MT + M \int_0^t \|c_s\|_E ds,$$

hence, Gronwall’s lemma implies that  $\|c_t\|_E \leq \bar{R}$  for every  $t \in [0, T]$ . With this,  $c_t$  solves (2.9). A similar argument shows that any other solution  $t \mapsto \hat{c}_t$  to (2.9) must satisfy  $\|\hat{c}_t\|_E \leq \bar{R}$  for every  $t \in [0, T]$ . Thus, uniqueness of solutions for (2.9) follows from the uniqueness of solutions to (2.11). A similar argument also yields (2.10).  $\square$

**3. The abstract model.** The state of our system is described pairs  $y := (x, \lambda)$ . The element  $x \in \mathbb{R}^d$  denotes the position of an agent, whereas the element  $\lambda \in \mathcal{P}(U)$  denotes a (probability) distribution over the space  $U$ , which we assume to be a compact metric space, which can be interpreted as a space of strategies (as in [5], in which case an element of  $\mathcal{P}(U)$  is a mixed strategy) or as a space of labels (as in [2], where the case  $U = \{\text{leader, follower}\}$  was considered). A distribution of states will be described by an element  $\Lambda \in \mathcal{P}(Y)$ , where  $Y := \mathbb{R}^d \times \mathcal{P}(U)$ . We will be concerned with the evolution of  $\Lambda$ , given an initial  $\Lambda^0$ , determined by the laws of evolution of  $x$  and  $\lambda$ , which are going to be discussed below.

For  $y = (x, \lambda) \in Y$  and  $\Psi \in \mathcal{P}_1(Y)$ , we define a vector field  $b_\Psi: Y \rightarrow \bar{Y}$  through

$$(3.1) \quad b_\Psi(y) := \begin{pmatrix} v_\Psi(y) \\ \mathcal{T}_\Psi(y) \end{pmatrix}.$$

The first component of  $b_\Psi$  is a velocity field in  $\mathbb{R}^d$  determined by the global state of the system  $\Psi$ ; the second component is expressed in terms of the adjoint  $\mathcal{T}^*(x, \Psi)$  of an operator  $\mathcal{T}(x, \Psi)$  which sees the location of the agents and the global state of the system around them. In order to state the regularity assumptions that we make on  $b_\Psi$ , we will discuss separately the assumptions on  $v_\Psi$  and on  $\mathcal{T}_\Psi$ .

We assume that the velocity field  $v_\Psi : Y \rightarrow \mathbb{R}^d$  satisfies the following conditions:

- (v1) for every  $R > 0$ , for every  $\Psi \in \mathcal{P}(B_R^Y)$ ,  $v_\Psi \in \text{Lip}(B_R^Y; \mathbb{R}^d)$  uniformly with respect to  $\Psi$ , namely, there exists a constant  $L_{v,R} > 0$  such that

$$(3.2) \quad |v_\Psi(y^1) - v_\Psi(y^2)| \leq L_{v,R} \|y^1 - y^2\|_{\bar{Y}};$$

- (v2) for every  $R > 0$ , for every  $\Psi \in \mathcal{P}(B_R^Y)$ , there exists a constant  $L_{v,R} > 0$  such that for every  $y \in B_R^Y$  and for every  $\Psi^1, \Psi^2 \in \mathcal{P}(B_R^Y)$

$$(3.3) \quad |v_{\Psi^1}(y) - v_{\Psi^2}(y)| \leq L_{v,R} \mathcal{W}_1(\Psi^1, \Psi^2);$$

- (v3) there exists  $M_v > 0$  such that for every  $y \in Y$  and for every  $\Psi \in \mathcal{P}_1(Y)$  there holds

$$(3.4) \quad |v_\Psi(y)| \leq M_v (1 + \|y\|_{\bar{Y}} + m_1(\Psi)).$$

We now describe the assumptions on  $\mathcal{T}$ . For  $\Psi \in \mathcal{P}_1(Y)$ , let  $\mathcal{T}_\Psi : Y \rightarrow \mathcal{F}(U)$  be an operator such that

- (T0) for every  $(y, \Psi) \in Y \times \mathcal{P}_1(Y)$ , constants are in the kernel of  $\mathcal{T}_\Psi(y)$ , that is,

$$(3.5) \quad \langle \mathcal{T}_\Psi(y), 1 \rangle_{\mathcal{F}(U), \text{Lip}(U)} = 0;$$

- (T1) for every  $(y, \Psi) \in Y \times \mathcal{P}_1(Y)$ , there exists a constant  $M_{\mathcal{T}} > 0$  such that

$$(3.6) \quad \|\mathcal{T}_\Psi(y)\|_{\text{BL}} \leq M_{\mathcal{T}} (1 + \|y\|_{\bar{Y}} + m_1(\Psi));$$

- (T2) for every  $R > 0$  there exists  $L_{\mathcal{T},R} > 0$  such that, for every  $(y^1, \Psi^1), (y^2, \Psi^2) \in B_R^Y \times \mathcal{P}(B_R^Y)$ ,

$$(3.7) \quad \|\mathcal{T}_{\Psi^1}(y^1) - \mathcal{T}_{\Psi^2}(y^2)\|_{\text{BL}} \leq L_{\mathcal{T},R} (\|y^1 - y^2\|_{\bar{Y}} + \mathcal{W}_1(\Psi^1, \Psi^2));$$

- (T3) for every  $R > 0$  there exists  $\delta_R > 0$  such that for every  $(y, \Psi) \in B_R^Y \times \mathcal{P}_1(Y)$  we have

$$(3.8) \quad \mathcal{T}_\Psi(y) + \delta_R \lambda \geq 0.$$

**PROPOSITION 3.1.** *For  $y \in Y$  and  $\Psi \in \mathcal{P}_1(Y)$ , define  $b_\Psi(y)$  as in (3.1). Assume that  $v_\Psi : Y \rightarrow \mathbb{R}^d$  satisfies (v1)–(v3) and  $\mathcal{T}_\Psi : Y \rightarrow \mathcal{F}(U)$  satisfies (T0)–(T3). Then*

- (i) *for every  $R > 0$ , for every  $\Psi \in \mathcal{P}(B_R^Y)$ , and for every  $y^1, y^2 \in B_R^Y$ , there exists  $L_R > 0$  such that*

$$(3.9) \quad \|b_\Psi(y^1) - b_\Psi(y^2)\|_{\bar{Y}} \leq L_R \|y^1 - y^2\|_{\bar{Y}};$$

- (ii) *for every  $R > 0$ , for every  $\Psi^1, \Psi^2 \in \mathcal{P}(B_R^Y)$ , and for every  $y \in B_R^Y$ , there exists  $L_R > 0$  such that*

$$(3.10) \quad \|b_{\Psi^1}(y) - b_{\Psi^2}(y)\|_{\bar{Y}} \leq L_R \mathcal{W}_1(\Psi^1, \Psi^2);$$

- (iii) *for every  $R > 0$ , there exists  $\theta > 0$  such that for every  $y \in B_R^Y$  and for every  $\Psi \in \mathcal{P}(B_R^Y)$*

$$(3.11) \quad y + \theta b_\Psi(y) \in Y;$$

(iv) *there exists  $M > 0$  such that for every  $y \in Y$  and for every  $\Psi \in \mathcal{P}_1(Y)$  there holds*

$$(3.12) \quad \|b_\Psi(y)\|_{\overline{Y}} \leq M(1 + \|y\|_{\overline{Y}} + m_1(\Psi)).$$

*Proof.* Property (i) is a consequence of (3.2) and (3.7); property (ii) follows from (3.3) and (3.7). Observing that for all  $y = (x, \lambda) \in Y$  we have that  $\|\lambda\|_{BL} \leq 1$ , we obtain (3.12) from (3.4) and (3.6). Finally, to prove (iii), since  $\mathbb{R}^d$  is convex, we simply have to show that for every  $R > 0$ , there exists  $\theta > 0$  such that for every  $y = (x, \lambda) \in B_R^Y$  and for every  $\Psi \in \mathcal{P}(B_R^Y)$

$$(3.13) \quad \lambda + \theta \mathcal{T}_\Psi(y)\lambda \in \mathcal{P}(U).$$

From (3.8), we have that, for  $\theta = 1/\delta_R$ ,  $\langle \lambda + \theta \mathcal{T}_\Psi(y), f \rangle_{\mathcal{F}(U), \text{Lip}(U)} \geq 0$ , for all  $f \in \text{Lip}(U)$ ,  $f \geq 0$ . Since for all positive functionals  $\zeta \in (\text{Lip}(U))'$  it holds

$$|\langle \zeta, f \rangle_{(\text{Lip}(U))', \text{Lip}(U)}| \leq \|f\|_\infty \langle \zeta, 1 \rangle$$

for all  $f \in \text{Lip}(U)$ , and since  $\text{Lip}(U)$  is dense in  $C(U)$  we have that  $\lambda + \theta \mathcal{T}_\Psi(y)$  can be extended to a positive linear functional in  $(C(U))'$ . Therefore, by the Riesz representation theorem,  $\lambda + \theta \mathcal{T}_\Psi(y) \in \mathcal{M}_+(U)$ . Now, by (T0),  $\langle \lambda + \theta \mathcal{T}_\Psi(y), 1 \rangle_{\mathcal{F}(U), \text{Lip}(U)} = \langle \lambda, f \rangle = 1$  for all  $\lambda \in \mathcal{P}(U)$ , which implies (3.13) and we conclude the proof.  $\square$

We conclude this subsection by providing an explicit example of operators satisfying (T0)–(T3) which will be used in our examples of sections 4 and 5. The following lemma is easily proved.

LEMMA 3.2. *Let  $\mathcal{Q}(x, \Psi) \in \mathcal{L}(\text{Lip}(U); \text{Lip}(U))$  satisfy*

(Q0) *for every  $(x, \Psi) \in \mathbb{R}^d \times \mathcal{P}_1(Y)$ , constants are in the kernel of  $\mathcal{Q}(x, \Psi)$ , that is,*

$$(3.14) \quad \mathcal{Q}(x, \Psi)1 = 0;$$

(Q1) *for every  $(x, \Psi) \in \mathbb{R}^d \times \mathcal{P}_1(Y)$ , there exists a constant  $M_{\mathcal{Q}} > 0$  such that the operator norm satisfies*

$$(3.15) \quad \|\mathcal{Q}(x, \Psi)\|_{\mathcal{L}(\text{Lip}(U); \text{Lip}(U))} \leq M_{\mathcal{Q}}(1 + |x| + m_1(\Psi));$$

(Q2) *for every  $R > 0$  there exists  $L_{\mathcal{Q}, R} > 0$  such that, for every  $(x^1, \Psi^1), (x^2, \Psi^2) \in B_R \times \mathcal{P}(B_R^Y)$ ,*

$$(3.16) \quad \|\mathcal{Q}(x^1, \Psi^1) - \mathcal{Q}(x^2, \Psi^2)\|_{\mathcal{L}(\text{Lip}(U); \text{Lip}(U))} \leq L_{\mathcal{Q}, R}(|x^1 - x^2| + \mathcal{W}_1(\Psi^1, \Psi^2));$$

(Q3) *for every  $R > 0$  there exists  $\delta_R > 0$  such that for every  $(x, \Psi) \in B_R \times \mathcal{P}_1(Y)$  we have*

$$(3.17) \quad \mathcal{Q}(x, \Psi) + \delta_R I \geq 0.$$

*Then the adjoint operator  $\mathcal{Q}^*(x, \Psi)$  maps  $\mathcal{P}(U)$  into  $\mathcal{F}(U)$  and  $\mathcal{T}_\Psi(y) := \mathcal{Q}^*(x, \Psi)\lambda$  satisfies (T0)–(T3) above.*

*Proof.* By (3.17) we have  $\mathcal{Q}^*(x, \Psi)\lambda + \delta_R \lambda \geq 0$  as an element of  $(\text{Lip}(U))'$ . By (3.14)  $\langle \mathcal{Q}^*(x, \Psi)\lambda, 1 \rangle_{\text{Lip}(U)', \text{Lip}(U)} = \langle \lambda, \mathcal{Q}(x, \Psi)1 \rangle = 0$ . Observe that these are exactly (T0) and (T3) once we prove that  $\mathcal{T}_\Psi(y) \in \mathcal{F}(U)$  for all  $(y, \Psi) \in Y \times \mathcal{P}_1(Y)$ . To see this, we argue as in the proof of Proposition 3.1 to obtain that  $\mathcal{Q}^*(x, \Psi)\lambda + \delta_R \lambda \in \mathcal{P}(U)$

for all  $\lambda \in \mathcal{P}(U)$ . Hence,  $\mathcal{T}_\Psi(y) = (\mathcal{Q}^*(x, \Psi)\lambda + \delta_R\lambda) - \delta_R\lambda$  is an element of  $\mathcal{F}(U)$  and (T0) and (T3) hold true.

It remains to see that (T1) and (T2) hold. By the definition of the operator norm, we have

$$\|\mathcal{Q}^*(x, \Psi)\|_{\mathcal{L}(\mathcal{F}(U); \mathcal{F}(U))} := \sup \{ \|\mathcal{Q}^*(x, \Psi)\mu\|_{\text{BL}} : \mu \in \mathcal{F}(U), \|\mu\|_{\text{BL}} \leq 1 \}.$$

For  $(x^1, \Psi^1), (x^2, \Psi^2) \in B_R \times \mathcal{P}(B_R^Y)$ , we have

$$\begin{aligned} & \|\mathcal{Q}^*(x^1, \Psi^1) - \mathcal{Q}^*(x^2, \Psi^2)\|_{\mathcal{L}(\mathcal{F}(U); \mathcal{F}(U))} \\ &= \sup \{ \|(\mathcal{Q}^*(x^1, \Psi^1) - \mathcal{Q}^*(x^2, \Psi^2))\mu\|_{\text{BL}} : \mu \in \mathcal{F}(U), \|\mu\|_{\text{BL}} \leq 1 \} \\ &= \sup \left\{ \sup \{ \langle (\mathcal{Q}^*(x^1, \Psi^1) - \mathcal{Q}^*(x^2, \Psi^2))\mu, \varphi \rangle : \|\varphi\|_{\text{Lip}} \leq 1 \} \right. \\ &\quad \left. : \mu \in \mathcal{F}(U), \|\mu\|_{\text{BL}} \leq 1 \right\} \\ &= \sup \left\{ \sup \{ \langle \mu, (\mathcal{Q}(x^1, \Psi^1) - \mathcal{Q}(x^2, \Psi^2))\varphi \rangle : \|\varphi\|_{\text{Lip}} \leq 1 \} : \mu \in \mathcal{F}(U), \|\mu\|_{\text{BL}} \leq 1 \right\} \\ &\leq \sup \left\{ \sup \{ \|\mu\|_{\text{BL}} \|(\mathcal{Q}(x^1, \Psi^1) - \mathcal{Q}(x^2, \Psi^2))\varphi\|_{\text{Lip}} : \|\varphi\|_{\text{Lip}} \leq 1 \} \right. \\ &\quad \left. : \mu \in \mathcal{F}(U), \|\mu\|_{\text{BL}} \leq 1 \right\} \\ &\leq \sup \{ \|(\mathcal{Q}(x^1, \Psi^1) - \mathcal{Q}(x^2, \Psi^2))\varphi\|_{\text{Lip}} : \|\varphi\|_{\text{Lip}} \leq 1 \} \\ &\leq \|\mathcal{Q}(x^1, \Psi^1) - \mathcal{Q}(x^2, \Psi^2)\|_{\mathcal{L}(\text{Lip}(U); \text{Lip}(U))}, \end{aligned}$$

so that (S2) gives

$$(3.18) \quad \|\mathcal{Q}^*(x^1, \Psi^1) - \mathcal{Q}^*(x^2, \Psi^2)\|_{\mathcal{L}(\mathcal{F}(U); \mathcal{F}(U))} \leq L_{\mathcal{Q}, R}(|x^1 - x^2| + \mathcal{W}_1(\Psi^1, \Psi^2)).$$

A similar argument, taking (S1) into account, gives

$$\|\mathcal{Q}^*(x, \Psi)\|_{\mathcal{L}(\mathcal{F}(U); \mathcal{F}(U))} \leq M_{\mathcal{Q}}(1 + |x| + m_1(\Psi))$$

for every  $(x, \Psi) \in \mathbb{R}^d \times \mathcal{P}_1(Y)$ . Since  $\|\lambda\|_{\text{BL}} \leq 1$  for all  $\lambda \in \mathcal{P}(U)$ , we deduce (T1). Writing

$$\mathcal{T}_{\Psi^1}(y^1) - \mathcal{T}_{\Psi^2}(y^2) = \mathcal{Q}^*(x^1, \Psi^1)(\lambda^1 - \lambda^2) + (\mathcal{Q}^*(x^1, \Psi^1) - \mathcal{Q}^*(x^2, \Psi^2))\lambda^2$$

and using (3.18), we get (T2). □

**3.1. The discrete problem and statement of the main result.** We consider a particle system of  $N$  agents evolving according to

$$(3.19) \quad \begin{cases} \dot{x}_t^i = v_{\Lambda_t^N}(x_t^i, \lambda_t^i), \\ \dot{\lambda}_t^i = \mathcal{T}_{\Lambda_t^N}(x_t^i, \lambda_t^i) \end{cases} \quad \text{for } i = 1, \dots, N, t \in [0, T],$$

where  $x_t^i \in \mathbb{R}^d, \lambda_t^i \in \mathcal{P}(U)$  for each  $i \in \{1, \dots, N\}$ , and

$$(3.20) \quad \Lambda_t^N := \frac{1}{N} \sum_{i=1}^N \delta_{(x_t^i, \lambda_t^i)}$$

is the empirical measure associated with the system. Recalling the definition of  $b$  in (3.1), the evolution (3.19) can be written in compact form as

$$(3.21) \quad \dot{y}_t^i = b_{\Lambda_t^N}(y_t^i) \quad \text{for } i = 1, \dots, N, t \in [0, T].$$

We first discuss the well-posedness of system (3.19) for every choice of an initial datum  $\bar{y}^i = (\bar{x}^i, \bar{\lambda}^i)$ , for  $i = 1, \dots, N$ .

PROPOSITION 3.3. Assume that for every  $y \in Y$  and  $\Psi \in \mathcal{P}_1(Y)$  the velocity  $v_\Psi : Y \rightarrow \mathbb{R}^d$  satisfies (v1)–(v3) and the operator  $\mathcal{T}_\Psi : Y \rightarrow \mathcal{F}(U)$  satisfies (T0)–(T3). Then, for every choice of  $\bar{y}^i \in Y, i = 1, \dots, N$ , the system (3.21) has a unique solution in  $[0, T]$ .

Proof. We introduce the vector-valued variable  $\mathbf{y} := (y^1, \dots, y^N) \in Y^N \subset \bar{Y}^N$ , which we endow with the norm

$$(3.22) \quad \|\mathbf{y}\|_{\bar{Y}^N} := \frac{1}{N} \sum_{i=1}^N \|y^i\|_{\bar{Y}},$$

and the associated empirical measure  $\Lambda^N := \frac{1}{N} \sum_{i=1}^N \delta_{y^i}$ , which belongs to  $\mathcal{P}(B_R^Y)$  whenever  $\mathbf{y} \in (B_R^Y)^N$ . Consider the map  $\mathbf{b}^N : Y^N \rightarrow \bar{Y}^N$  whose components are defined through

$$(3.23) \quad b_i^N(\mathbf{y}) := b_{\Lambda^N}(y^i).$$

Then the Cauchy problem associated with (3.21) can be written as

$$(3.24) \quad \begin{cases} \dot{\mathbf{y}}_t = \mathbf{b}^N(\mathbf{y}_t), \\ \mathbf{y}_0 = \bar{\mathbf{y}}. \end{cases}$$

In order to apply Corollary 2.3 to the system above, we first notice that assumption (ii) in Theorem 2.2 is automatically satisfied since the system is autonomous. To check the other assumptions, we fix a ball  $B_R^{Y^N}$  and notice that  $B_R^{Y^N} \subset (B_R^Y)^N$ .

Applying (3.11) with  $\Psi = \Lambda^N$  to each component  $y^i$  of  $\mathbf{y}$ , we get that assumption (iii) of Theorem 2.2 is satisfied.

We now show that assumption (i') holds. Fix  $\mathbf{y}_1, \mathbf{y}_2 \in B_R^{Y^N} \subset (B_R^Y)^N$ , and let  $\Lambda_1^N, \Lambda_2^N$  be the associated empirical measures. Notice that

$$(3.25) \quad \mathcal{W}_1(\Lambda_1^N, \Lambda_2^N) \leq \frac{C_U}{N} \sum_{i=1}^N \|y_1^i - y_2^i\|_{\bar{Y}} = C_U \|\mathbf{y}_1 - \mathbf{y}_2\|_{\bar{Y}^N},$$

where  $C_U$  depends on the compact set  $U$ . With this, by the triangle inequality, (3.9), and (3.10), we estimate

$$(3.26) \quad \begin{aligned} & \|\mathbf{b}^N(\mathbf{y}_1) - \mathbf{b}^N(\mathbf{y}_2)\|_{\bar{Y}^N} \\ &= \frac{1}{N} \sum_{i=1}^N \|b_{\Lambda_1^N}(y_1^i) - b_{\Lambda_2^N}(y_2^i)\|_{\bar{Y}} \\ &\leq L_R \mathcal{W}_1(\Lambda_1^N, \Lambda_2^N) + \frac{L_R}{N} \sum_{i=1}^N \|y_1^i - y_2^i\|_{\bar{Y}} \leq L_R(1 + C_U) \|\mathbf{y}_1 - \mathbf{y}_2\|_{\bar{Y}^N}, \end{aligned}$$

which yields (2.7) up to renaming the constant. To see that also (2.8) and, therefore, assumption (i''), holds, we apply (3.12) and obtain, upon noticing that  $m_1(\Lambda^N) = \|\mathbf{y}\|_{\bar{Y}^N}$ ,

$$(3.27) \quad \begin{aligned} \|\mathbf{b}^N(\mathbf{y})\|_{\bar{Y}^N} &= \frac{1}{N} \sum_{i=1}^N \|b_{\Lambda^N}(y^i)\|_{\bar{Y}} \\ &\leq \frac{M}{N} \sum_{i=1}^N (1 + \|y^i\|_{\bar{Y}} + m_1(\Lambda^N)) = M(1 + 2\|\mathbf{y}\|_{\bar{Y}^N}). \end{aligned}$$

Therefore we can apply Corollary 2.3, which proves the statement. □

We are now in a position to state the main result of our paper, concerning the mean-field limit as  $N \rightarrow \infty$  of the solutions  $(y_t^1, \dots, y_t^N)$  to (3.21) or, equivalently, the limiting behavior of the associated empirical measures  $\Lambda_t^N$ . In order to do so, we first need to recall the concept of Eulerian solution to the continuity equation.

**DEFINITION 3.4 (Eulerian solution).** *Let  $\Lambda \in C^0([0, T]; (\mathcal{P}_1(Y), \mathcal{W}_1))$  and let  $\bar{\Lambda} \in \mathcal{P}_c(Y)$  be a given initial datum. We say that  $\Lambda$  is an Eulerian solution to the initial value problem for the equation*

$$(3.28) \quad \partial_t \Lambda_t + \operatorname{div}(b_{\Lambda_t} \Lambda_t) = 0$$

starting from  $\bar{\Lambda}$  if and only if  $\Lambda_0 = \bar{\Lambda}$  and, for every  $\phi \in C_b^1([0, T] \times \bar{Y})$ ,

$$(3.29) \quad \int_Y \phi(t, y) d\Lambda_t(y) - \int_Y \phi(0, y) d\Lambda_0(y) = \int_0^t \int_Y (\partial_t \phi(s, y) + D\phi(s, y) \cdot b_{\Lambda_s}(y)) d\Lambda_s(y) ds,$$

where  $D\phi(s, y)$  is the Fréchet differential of  $\phi$  in the  $y$  variable.

The main result of our paper is the following theorem, stating the existence of a unique Eulerian solution to (3.28) and its characterization as the mean-field limit of solutions to the discrete problem (3.21).

**THEOREM 3.5.** *Let  $r > 0$  and  $\bar{\Lambda} \in \mathcal{P}(B_r^Y)$  be a given initial datum. Then*

- (i) *there exists a unique Eulerian solution  $t \mapsto \Lambda_t$  to (3.28) starting from  $\bar{\Lambda}$ ;*
- (ii) *if  $\bar{\Lambda}^N = \frac{1}{N} \sum_{i=1}^N \delta_{\bar{y}^{N,i}}$  is a sequence of atomic measures in  $\mathcal{P}(B_r^Y)$  such that*

$$\lim_{N \rightarrow \infty} \mathcal{W}_1(\bar{\Lambda}^N, \bar{\Lambda}) = 0$$

*and, for fixed  $N$ ,  $\Lambda_t^N$  are the empirical measures associated with the unique solution to (3.21) with initial datum  $\bar{y}^{N,i}$ , we have*

$$\lim_{N \rightarrow \infty} \mathcal{W}_1(\Lambda_t^N, \Lambda_t) = 0 \quad \text{uniformly with respect to } t \in [0, T].$$

The proof of Theorem 3.5 will be based on a fixed point argument and on the notion of Lagrangian solution, which are going to be introduced in the next subsection.

**3.2. Lagrangian solutions.** We start by proving an auxiliary well-posedness result for an ODE in  $Y$  of the form

$$(3.30) \quad \dot{y}_t = b_{\Psi_t}(y_t), \quad y_0 = \bar{y},$$

where  $[0, T] \ni t \mapsto \Psi_t \in \mathcal{P}_1(Y)$  is a given continuous curve and  $\bar{y} \in Y$ .

**PROPOSITION 3.6.** *Assume that for every  $y \in Y$  and  $\Psi \in \mathcal{P}_1(Y)$  the velocity  $v_\Psi: Y \rightarrow \mathbb{R}^d$  satisfies (v1)–(v3) and the operator  $\mathcal{T}_\Psi: Y \rightarrow \mathcal{F}(U)$  satisfies (T0)–(T3). Let  $\Psi \in C^0([0, T]; (\mathcal{P}_1(Y), \mathcal{W}_1))$  and assume that there exists  $R > 0$  such that, in addition,  $\Psi_t \in \mathcal{P}(B_R^Y)$  for all  $t \in [0, T]$ . Then, for every choice of  $\bar{y} \in Y$ , the ODE (3.30) has a unique solution.*

*Proof.* We set  $b(t, y) := b_{\Psi_t}(y)$  according to (3.1). Since  $t \mapsto \Psi_t$  is continuous, using (3.10) we get that, for any fixed  $y \in Y$ ,  $b(\cdot, y)$  is continuous, which is condition (ii) in Corollary 2.3. Condition (iii) of Corollary 2.3 is a direct consequence of (3.11). Furthermore, (3.9) and (3.12) yield (2.7) and (2.8), respectively. The proof is concluded by Corollary 2.3.  $\square$

In view of the previous result, the following definition is justified.

DEFINITION 3.7 (transition map). *The transition map  $\mathbf{Y}_\Psi(t, s, \bar{y})$  associated with the ODE (3.30), replacing the initial condition by  $y_s = \bar{y}$ , is defined through*

$$\mathbf{Y}_\Psi(t, s, \bar{y}) = y_t,$$

where  $t \mapsto y_t$  is the unique solution to (3.30) .

We can now proceed to defining the notion of Lagrangian solution to (3.28).

DEFINITION 3.8 (Lagrangian solutions). *Let  $\Lambda \in C^0([0, T]; (\mathcal{P}_1(Y), \mathcal{W}_1))$  and let  $\bar{\Lambda} \in \mathcal{P}_c(Y)$  be a given initial datum. We say that  $\Lambda$  is a Lagrangian solution to the initial value problem for (3.28) starting from  $\bar{\Lambda}$  if and only if it satisfies the fixed point condition*

$$(3.31) \quad \Lambda_t = \mathbf{Y}_\Lambda(t, 0, \cdot) \# \bar{\Lambda} \quad \text{for every } 0 \leq t \leq T,$$

where  $\mathbf{Y}_\Lambda(t, s, y)$  are the transition maps associated with the ODE (3.30).

Remark 3.9. It follows from Definition 2.1 that Lagrangian solutions are also Eulerian solutions.

Remark 3.10. For a fixed  $N \in \mathbb{N}$ , let  $\Lambda_t^N$  be the empirical measures associated with the unique solution to (3.21) with initial datum  $\bar{y}^i, i = 1, \dots, N$ . If we now set  $\bar{\Lambda}^N := \frac{1}{N} \sum_{i=1}^N \delta_{\bar{y}^i}$ , by Definition 3.7 there holds

$$(3.32) \quad \Lambda_t^N = \mathbf{Y}_{\Lambda^N}(t, 0, \cdot) \# \bar{\Lambda}^N \quad \text{for every } 0 \leq t \leq T.$$

Hence,  $\Lambda^N$  is a Lagrangian and Eulerian solution to (3.28) starting from  $\bar{\Lambda}^N$ .

We now want to show that an infinite-dimensional converse of Proposition 3.6 holds, proving that indeed, in our case, every Eulerian solution is also a Lagrangian solution. This stems out of a general abstract principle known as the *superposition principle* in the version introduced in [5] (see also [7, Theorem 8.2.1] and [9, Theorem 7.1]). In the statement below, the *evaluation map*  $ev_t$  is defined, at a given  $t \in [0, T]$ , by

$$ev_t(\gamma) := \gamma(t) \quad \text{for all } \gamma \in C([0, T]; E);$$

we also use the notion of *cylindrical functions*, which are defined in the following way. We say that  $\phi \in C_b^1(E)$  is a cylindrical function if there exist a function  $\varphi \in C_b^1(\mathbb{R}^N)$  and  $z'_1, \dots, z'_N \in E'$  such that

$$\phi(y) = \varphi(\langle y, z'_1 \rangle, \langle y, z'_2 \rangle, \dots, \langle y, z'_N \rangle),$$

where  $\langle \cdot, \cdot \rangle$  denotes the duality map between  $E$  and  $E'$ .

THEOREM 3.11 (superposition principle). *Let  $(E, \|\cdot\|_E)$  be a separable Banach space, let  $b: (0, T) \times E \rightarrow E$  be a Borel vector field, and let  $\mu_t \in \mathcal{P}(E), t \in [0, T]$ , be a continuous curve with*

$$(3.33) \quad \int_0^T \int_E \|b_t\|_E \, d\mu_t \, dt < +\infty.$$

If

$$\frac{d}{dt} \mu_t + \operatorname{div}(b_t \mu_t) = 0$$



in duality with cylindrical functions  $\phi \in C_b^1(E)$ ,<sup>2</sup> then there exists  $\boldsymbol{\eta} \in \mathcal{P}(C([0, T]; E))$  concentrated on absolutely continuous solutions to the ODE  $\dot{y} = b_t(y)$  and with  $(\text{ev}_t)_\# \boldsymbol{\eta} = \mu_t$  for all  $t \in [0, T]$ .

*Proof.* See [5, Theorem 5.2] for the proof. □

Combining the abstract result Theorem 3.11 with the uniqueness granted by Proposition 3.6, we can prove the announced equivalence result. Notice that the proof has an intermediate step, since in order to apply Proposition 3.6 we must first ensure that an Eulerian solution  $\Lambda_t$  has (equi)compact support for all  $t$ . We are able to deduce this from Theorem 3.11 and the assumption that the initial datum  $\bar{\Lambda} \in \mathcal{P}_c(Y)$ .

**THEOREM 3.12.** *Let  $\Lambda \in C^0([0, T]; (\mathcal{P}_1(Y), \mathcal{W}_1))$  and let  $\bar{\Lambda} \in \mathcal{P}_c(Y)$  be a given initial datum. Assume that  $\Lambda$  is an Eulerian solution to the initial value problem for  $\partial_t \Lambda_t + \text{div}(b_{\Lambda_t} \Lambda_t) = 0$  (see (3.28)) starting from  $\bar{\Lambda}$ , in the sense of (3.29), then there exists  $R > 0$  such that  $\Lambda_t \in \mathcal{P}(B_R^Y)$  for all  $t \in [0, T]$  and*

$$\Lambda_t = \mathbf{Y}_\Lambda(t, 0, \cdot)_\# \bar{\Lambda} \quad \text{for every } 0 \leq t \leq T,$$

where  $\mathbf{Y}_\Lambda(t, s, y)$  are the transition maps associated with the ODE (3.30).

*Proof.* Since  $\Lambda \in C^0([0, T]; (\mathcal{P}_1(Y), \mathcal{W}_1))$ , the map

$$(3.34) \quad t \mapsto m_1(\Lambda_t) = \int_Y \|y\|_{\bar{Y}} d\Lambda_t(y)$$

is continuous and, hence, bounded, in  $[0, T]$ . Set  $b_t(y) := b_{\Lambda_t}(y)$  for  $y \in Y$ , and extend it to zero on  $\bar{Y} \setminus Y$ . Using (3.12), we have

$$(3.35) \quad \begin{aligned} \int_0^T \int_{\bar{Y}} \|b_t\|_{\bar{Y}} d\Lambda_t(y) dt &= \int_0^T \int_Y \|b_t\|_{\bar{Y}} d\Lambda_t(y) dt \\ &\leq \int_0^T \int_Y M(1 + \|y\|_{\bar{Y}} + m_1(\Lambda_t)) d\Lambda_t(y) dt \\ &\leq TM \left( 1 + 2 \max_{t \in [0, T]} m_1(\Lambda_t) \right) < +\infty. \end{aligned}$$

Hence, we can apply Theorem 3.11 with  $E = \bar{Y}$  and  $\mu_t = \Lambda_t$ , obtaining that  $\Lambda_t = (\text{ev}_t)_\# \boldsymbol{\eta}$  for a suitable  $\boldsymbol{\eta} \in \mathcal{P}(C([0, T]; \bar{Y}))$  concentrated on absolutely continuous solutions to ODE

$$(3.36) \quad \dot{y} = b_{\Lambda_t}(y) \quad \text{in } [0, T].$$

Now, using (3.12) again, we have

$$(3.37) \quad \|b_{\Lambda_t}(y)\|_{\bar{Y}} \leq M \left( 1 + \max_{t \in [0, T]} m_1(\Lambda_t) + \|y\|_{\bar{Y}} \right) \leq M_\Lambda (1 + \|y\|_{\bar{Y}}),$$

where we set  $M_\Lambda := M(1 + \max_{t \in [0, T]} m_1(\Lambda_t))$ . The equality  $\bar{\Lambda} = \Lambda_0 = (\text{ev}_0)_\# \boldsymbol{\eta}$ , which reads

$$\int_{\mathcal{P}(C([0, T]; \bar{Y}))} \phi(\gamma(0)) d\boldsymbol{\eta}(\gamma) = \int_Y \phi(y) d\bar{\Lambda}(y)$$

<sup>2</sup>By this we mean that the definition of weak solutions to the continuity equation has only to be satisfied for test functions that are cylindrical.

for each  $\phi \in C_b(Y)$ , implies that  $\boldsymbol{\eta}$  is concentrated on the set of solutions to (3.36) satisfying  $y(0) \in B_r^Y$ , where  $r$  is such that  $\text{supp } \bar{\Lambda} \subset B_r^Y$ . With (3.37) and the Grönwall inequality, each of these solutions must satisfy  $y(t) \in B_R^Y$ , where  $R$  is explicitly given by

$$R := R_{r,M,\Lambda,T} = (r + M_\Lambda T)e^{M_\Lambda T}.$$

From the equality  $\Lambda_t = (\text{ev}_t)_\# \boldsymbol{\eta}$  we then deduce that  $\Lambda_t \in \mathcal{P}_1(B_R^Y)$  for all  $t \in [0, T]$ . We can therefore apply Proposition 3.6 and exploit the uniqueness of the solution to the Cauchy problem (3.30) to deduce the representation

$$\gamma(t) = \mathbf{Y}_\Lambda(t, 0, \gamma(0))$$

with  $\gamma(0) \in B_r^Y$ , for each continuous path  $\gamma \in \text{supp } \boldsymbol{\eta}$ . With the equality  $\Lambda_t = (\text{ev}_t)_\# \boldsymbol{\eta}$ , this gives

$$\int_Y \phi(y) d\Lambda_t(y) = \int_{\mathcal{P}(C([0,T]; \bar{Y}))} \phi(\mathbf{Y}_\Lambda(t, 0, \gamma(0))) d\boldsymbol{\eta}(\gamma) = \int_Y \phi(\mathbf{Y}_\Lambda(t, 0, y)) d\bar{\Lambda}(y)$$

for each  $\phi \in C_b(Y)$ , which implies the conclusion. □

**3.3. Proof of Theorem 3.5.** As a preliminary step towards the proof, we need the following lemma, ensuring that the size of the support of a Lagrangian solution in the sense of (3.31) can be a priori estimated from the data of the problem.

**LEMMA 3.13.** *Assume that for every  $y \in Y$  and  $\Psi \in \mathcal{P}_1(Y)$  the velocity  $v_\Psi: Y \rightarrow \mathbb{R}^d$  satisfies (v1)–(v3) and the operator  $\mathcal{T}_\Psi: Y \rightarrow \mathcal{F}(U)$  satisfies (T0)–(T3). Let  $\Lambda \in C^0([0, T]; (\mathcal{P}_1(Y), \mathcal{W}_1))$  and let  $\bar{\Lambda} \in \mathcal{P}_c(Y)$  be a given initial datum. Fix  $r > 0$  such that  $\bar{\Lambda}$  has support in  $B_r^Y$ , and let  $M$  be the constant given by (3.12). Assume that  $\Lambda$  is a Lagrangian solution to the initial value problem for (3.28) starting from  $\bar{\Lambda}$  in the sense of (3.31). Then, for  $R = (r + MT)e^{2MT}$  we have*

$$\Lambda_t \in \mathcal{P}_1(B_R^Y) \quad \text{for all } t \in [0, T].$$

*Proof.* For  $r, R$  as in the statement, it suffices to show that we have

$$(3.38) \quad \max_{y \in B_r^Y} \|\mathbf{Y}_{\Lambda_t}(t, 0, y)\|_{\bar{Y}} \leq R$$

for all  $t \in [0, T]$ . Indeed, if this holds the statement immediately follows by (3.31) and elementary properties of the push-forward measure, taking into account that  $\bar{\Lambda}$  has support in  $B_r^Y$ . To prove the above claim, we first observe that by the definition of Lagrangian solutions and the fact that  $\bar{\Lambda} \in \mathcal{P}(B_r^Y)$  we immediately have

$$(3.39) \quad m_1(\Lambda_t) \leq \max_{y \in B_r^Y} \|\mathbf{Y}_\Lambda(t, 0, y)\|_{\bar{Y}}$$

for all  $t \in [0, T]$ . We now set  $f(s) = \max_{y \in B_r^Y} \|\mathbf{Y}_\Lambda(s, 0, y)\|_{\bar{Y}}$ . Then, one has by definition of the transition map, (3.12), and (3.39) that for every choice of  $y \in B_r^Y$

$$\|\mathbf{Y}_\Lambda(t, 0, y)\|_{\bar{Y}} \leq r + M \int_0^t (1 + \|\mathbf{Y}_\Lambda(s, 0, y)\|_{\bar{Y}} + m_1(\Lambda_s)) ds \leq r + M \int_0^t (1 + 2f(s)) ds$$

which implies by the Gronwall inequality  $f(t) \leq (r + Mt)e^{2Mt}$  for all  $t$ , confirming (3.38). □

*Proof of Theorem 3.5.* The proof goes through a finite-dimensional approximation and involves three steps.

*Step 1: Stability of Lagrangian solutions.* We fix  $r > 0$ , two initial data  $\bar{\Lambda}^1, \bar{\Lambda}^2 \in \mathcal{P}(B_r^Y)$ , and assume that two Lagrangian solutions  $\Lambda_t^1, \Lambda_t^2$ , starting from  $\bar{\Lambda}^1$  and  $\bar{\Lambda}^2$ , respectively, exist. We fix  $R = (r + MT)e^{2MT}$  and the corresponding constant  $L_R$  provided by (3.9)–(3.10). We claim that

$$(3.40) \quad \mathcal{W}_1(\Lambda_t^1, \Lambda_t^2) \leq e^{L_R t + e^{L_R t} - 1} \mathcal{W}_1(\bar{\Lambda}^1, \bar{\Lambda}^2) \quad \text{for all } t \in [0, T].$$

To prove this claim, we fix  $y_1$  and  $y_2 \in B_r^Y$  and observe that

$$(3.41) \quad \|\mathbf{Y}_{\Lambda^i}(t, 0, y_i)\|_{\mathcal{Y}} \leq R$$

for all  $t \in [0, T]$  and  $i = 1, 2$ . This can be proved along similar lines as in the proof of (3.39). With (3.41) and (3.9)–(3.10), the solutions  $y^1(t)$  and  $y^2(t)$  to the ODEs  $\dot{y}^i = b_{\Lambda^i}(y^i)$  with initial data  $y^1$  and  $y^2$ , respectively, satisfy

$$\begin{aligned} \frac{d}{dt} \|y^1 - y^2\|(t) &\leq \|b_{\Lambda^1}(y^1(t)) - b_{\Lambda^1}(y^2(t))\| + \|b_{\Lambda^1}(y^2(t)) - b_{\Lambda^2}(y^2(t))\| \\ &\leq L_R \|y^1 - y^2\|(t) + L_R \mathcal{W}_1(\Lambda_t^1, \Lambda_t^2). \end{aligned}$$

This gives, by means of a comparison argument, that

$$\|y^1(t) - y^2(t)\| \leq e^{L_R t} \|y^1 - y^2\| + L_R \int_0^t e^{L_R(t-\tau)} \mathcal{W}_1(\Lambda_\tau^1, \Lambda_\tau^2) \, d\tau;$$

equivalently,

$$(3.42) \quad \|\mathbf{Y}_{\Lambda^1}(t, 0, y^1) - \mathbf{Y}_{\Lambda^2}(t, 0, y^2)\| \leq e^{L_R t} \|y^1 - y^2\| + L_R \int_0^t e^{L_R(t-\tau)} \mathcal{W}_1(\Lambda_\tau^1, \Lambda_\tau^2) \, d\tau$$

for all  $t \in [0, T]$  and  $y_1$  and  $y_2 \in B_r^Y$ .

Now, let  $\Pi$  be an optimal coupling between  $\bar{\Lambda}^1$  and  $\bar{\Lambda}^2$ . Then clearly, by the definition of Lagrangian solutions,  $(\mathbf{Y}_{\Lambda^1}(t, 0, y^1), \mathbf{Y}_{\Lambda^2}(t, 0, y^2))_{\#} \Pi$  is a coupling between  $\Lambda_t^1$  and  $\Lambda_t^2$ . Therefore

$$\begin{aligned} \mathcal{W}_1(\Lambda_t^1, \Lambda_t^2) &\leq \int_{Y \times Y} \|\mathbf{Y}_{\Lambda^1}(t, 0, y^1) - \mathbf{Y}_{\Lambda^2}(t, 0, y^2)\| \, d\Pi(y^1, y^2) \\ &= \int_{B_r^Y \times B_r^Y} \|\mathbf{Y}_{\Lambda^1}(t, 0, y^1) - \mathbf{Y}_{\Lambda^2}(t, 0, y^2)\| \, d\Pi(y^1, y^2), \end{aligned}$$

where we also used that  $\bar{\Lambda}^1, \bar{\Lambda}^2 \in \mathcal{P}(B_r^Y)$ . Hence, using (3.42) we get

$$\begin{aligned} \mathcal{W}_1(\Lambda_t^1, \Lambda_t^2) &\leq e^{L_R t} \int_{B_r^Y \times B_r^Y} \|y^1 - y^2\| \, d\Pi(y^1, y^2) + L_R \int_0^t e^{L_R(t-\tau)} \mathcal{W}_1(\Lambda_\tau^1, \Lambda_\tau^2) \, d\tau \\ &= e^{L_R t} \mathcal{W}_1(\bar{\Lambda}^1, \bar{\Lambda}^2) + L_R \int_0^t e^{L_R(t-\tau)} \mathcal{W}_1(\Lambda_\tau^1, \Lambda_\tau^2) \, d\tau. \end{aligned}$$

With this and the Grönwall inequality, we get (3.40).

*Step 2: Existence and approximation of Lagrangian solutions.* We start by fixing a sequence of atomic measures  $\bar{\Lambda}^N \in \mathcal{P}(B_r^Y)$  such that

$$(3.43) \quad \lim_{N \rightarrow \infty} \mathcal{W}_1(\bar{\Lambda}^N, \bar{\Lambda}) = 0.$$

Such a sequence can be for instance constructed as follows: choose  $\bar{y}_i(\omega) \in Y$  independent and identically distributed, with law  $\bar{\Lambda}$ , so that the random measures  $\bar{\Lambda}^N(\omega) := \frac{1}{N} \sum_{i=1}^N \delta_{\bar{y}_i(\omega)}$  almost surely converge in  $\mathcal{P}_1(Y)$  to  $\bar{\Lambda}$ , and choose a realization  $\omega$  such that this convergence takes place. Now, let  $\Lambda_t^N$  be the empirical measures associated with the unique solution to (3.21) with initial datum  $\bar{y}^i, i = 1, \dots, N$ . As noticed in (3.32),  $\Lambda_t^N$  are Lagrangian solutions to (3.28) starting from  $\bar{\Lambda}^N$ . Hence, (3.40) provides a constant  $C := C(M, r, T)$  such that

$$\mathcal{W}_1(\Lambda_t^N, \Lambda_t^M) \leq C \mathcal{W}_1(\bar{\Lambda}^N, \bar{\Lambda}^M)$$

for all  $t \in [0, T]$  and  $N, M \in \mathbb{N}$ . It follows that  $\Lambda_t^N \in C([0, T]; (\mathcal{P}_1(B_R^Y), \mathcal{W}_1))$  is a Cauchy sequence. Let then  $\Lambda_t \in C([0, T]; (\mathcal{P}_1(B_R^Y), \mathcal{W}_1))$  be the limit of the sequence  $\Lambda_t^N$ .

For a given  $\bar{y} \in B_r^Y$ , consider now the solutions  $y^N(t)$  and  $y(t)$  to the ODEs  $\dot{y}^N = b_{\Lambda^N}(y^N)$  and  $\dot{y} = b_{\Lambda}(y)$ , respectively, with initial datum  $\bar{y}$ . Let  $R' \geq R$  be an upper bound<sup>3</sup> for  $\max_{t \in [0, T]} \|y(t)\|_{\bar{Y}}$ , which can be taken independent of  $\bar{y} \in B_r^Y$ . With (3.9)–(3.10) we obtain

$$\begin{aligned} \frac{d}{dt} \|y^N - y\|(t) &\leq \|b_{\Lambda^N}(y^N) - b_{\Lambda}(y^N)\| + \|b_{\Lambda}(y^N) - b_{\Lambda}(y)\| \\ &\leq L_{R'} \|y^N - y\| + L_R \mathcal{W}_1(\Lambda_t^N, \Lambda_t). \end{aligned}$$

Again by comparison, we deduce that

$$\|\mathbf{Y}_{\Lambda^N}(t, 0, \bar{y}) - \mathbf{Y}_{\Lambda}(t, 0, \bar{y})\| \leq L_R \int_0^t e^{L_{R'}(t-\tau)} \mathcal{W}_1(\Lambda_{\tau}^N, \Lambda_{\tau}) \, d\tau$$

which entails the uniform convergence of  $\mathbf{Y}_{\Lambda^N}(\cdot, 0, \cdot)$  to  $\mathbf{Y}_{\Lambda}(\cdot, 0, \cdot)$  in  $[0, T] \times B_r^Y$ . For each  $t \in [0, T]$  this implies, together with (3.43) and the fact that  $\mathbf{Y}_{\Lambda}(t, 0, \cdot)$  is a Lipschitz map on  $B_r^Y$ , that

$$\Lambda_t^N = \mathbf{Y}_{\Lambda^N}(t, 0, \cdot)_{\#} \bar{\Lambda}^N \rightarrow \mathbf{Y}_{\Lambda}(t, 0, \cdot)_{\#} \bar{\Lambda}$$

in  $\mathcal{P}_1(Y)$ , which gives (3.31).

*Step 3: Uniqueness and conclusion.* Uniqueness of Lagrangian solutions, given the initial datum, follows now from (3.40). Existence and uniqueness of Eulerian solutions is now a consequence of Remark 3.9, and Theorem 3.12, respectively. The same argument used in the second step of this proof gives also part (ii) of the statement.  $\square$

**4. A leader-follower dynamics.** We want to discuss the application of our results to the relevant scenario of two interacting populations, one consisting of leaders and the other one of followers, with a switching rate between the two. In this setting, the set  $U$  consists of two elements, that is,  $U := \{F, L\}$  and is endowed with a two-valued distance

$$0 = \text{dist}(F, F) = \text{dist}(L, L), \quad 1 = \text{dist}(F, L) = \text{dist}(L, F).$$

The space  $\text{Lip}(\{F, L\})$  is a two-dimensional linear space spanned by the two indicator functions  $\mathbb{1}_F$  and  $\mathbb{1}_L$ ; accordingly, the space  $\mathcal{F}(\{F, L\})$  is the two-dimensional space of

<sup>3</sup>Observe that at this point of the proof we cannot a priori exclude that  $R' > R$ , since we still do not know that  $\Lambda_t$  is a Lagrangian solution, hence we cannot apply (3.41) (which holds instead for  $\Lambda^N$ ).

signed Borel measures on the discrete set  $\{F, L\}$ , whose generic element  $\xi$  is completely described by the two values  $\xi_F := \xi(\{F\})$  and  $\xi_L := \xi(\{L\})$ .

Throughout this section, we will restrict our attention to transition operators  $\mathcal{T}$  which act linearly on  $\xi$  of the type considered in Lemma 3.2. A simple characterization of operators of this kind complying with our set of assumptions (T0)–(T3) is given in the following proposition.

**PROPOSITION 4.1.** *Let  $U := \{F, L\}$  and  $\mathcal{T}_\Psi: Y \rightarrow \mathcal{F}(\{F, L\})$  be of the form  $\mathcal{T}_\Psi(y) = \mathcal{Q}^*(x, \Psi)\lambda$  for  $\mathcal{Q}(x, \Psi) \in \mathcal{L}(\text{Lip}(\{F, L\}); \text{Lip}(\{F, L\}))$ . Then  $\mathcal{T}_\Psi$  satisfies (T0)–(T3) if and only if there exist two functions  $\alpha_F, \alpha_L: \mathbb{R}^d \times \mathcal{P}_1(Y) \rightarrow [0, +\infty)$  such that*

( $\alpha 0$ ) for every  $(x, \Psi) \in \mathbb{R}^d \times \mathcal{P}_1(Y)$  and  $\lambda \in \mathcal{P}_1(\{F, L\})$  it holds that

$$\begin{aligned} (\mathcal{Q}^*(x, \Psi)\lambda)_F &= -\alpha_F(x, \Psi)\lambda_F + \alpha_L(x, \Psi)(1 - \lambda_F), \\ (\mathcal{Q}^*(x, \Psi)\lambda)_L &= \alpha_F(x, \Psi)\lambda_F - \alpha_L(x, \Psi)(1 - \lambda_F); \end{aligned}$$

( $\alpha 1$ ) for every  $(x, \Psi) \in \mathbb{R}^d \times \mathcal{P}_1(Y)$ , there exists  $M_{\mathcal{Q}} > 0$  such that

$$0 \leq \alpha_\bullet(x, \Psi) \leq M_{\mathcal{Q}}(1 + |x| + m_1(\Psi)) \quad \text{for } \bullet = F, L;$$

( $\alpha 2$ ) for every  $R > 0$  there exists  $L_{\mathcal{Q}, R} > 0$  such that, for every  $(x^1, \Psi^1), (x^2, \Psi^2) \in B_R \times \mathcal{P}(B_R^Y)$ ,

$$|\alpha_\bullet(x^1, \Psi^1) - \alpha_\bullet(x^2, \Psi^2)| \leq L_{\mathcal{Q}, R}(|x^1 - x^2| + \mathcal{W}_1(\Psi^1, \Psi^2)) \quad \text{for } \bullet = F, L.$$

*Proof.* Since  $\text{Lip}(\{F, L\})$  is a two-dimensional linear space, we can identify the operator  $\mathcal{Q}(x, \Psi)$  with its matrix representation with respect to the canonical basis  $\{\mathbb{1}_F, \mathbb{1}_L\}$  and endow the space  $\mathcal{L}(\text{Lip}(\{F, L\}), \text{Lip}(\{F, L\}))$  with the Frobenius norm of such a matrix representation. Accordingly, the transpose matrix will be the representation of  $\mathcal{Q}^*(x, \Psi)$  with respect to the canonical basis of  $\mathcal{F}(\{F, L\})$  consisting of the two Dirac masses  $\delta_F$  and  $\delta_L$ .

For  $\mathcal{T}_\Psi(y)$  of the form  $\mathcal{Q}^*(x, \Psi)\lambda$ , conditions (T0)–(T3) are equivalent to conditions (S0)–(S3) of Lemma 3.2. Condition (Q0) is equivalent to the fact that

$$(4.1) \quad \mathcal{Q}(x, \Psi) = \begin{pmatrix} -\alpha_F(x, \Psi) & \alpha_F(x, \Psi) \\ \alpha_L(x, \Psi) & -\alpha_L(x, \Psi) \end{pmatrix}$$

with  $\alpha_F, \alpha_L: \mathbb{R}^d \times \mathcal{P}_1(Y) \rightarrow [0, +\infty)$ . In turn, this is equivalent to ( $\alpha 0$ ) by a direct computation. Again by a direct computation, (Q2) is equivalent to ( $\alpha 2$ ), while (Q1) is equivalent to the inequality

$$|\alpha_\bullet(x, \Psi)| \leq M_{\mathcal{Q}}(1 + |x| + m_1(\Psi)) \quad \text{for } \bullet = F, L.$$

In particular,  $\alpha_F$  and  $\alpha_L$  are uniformly bounded for  $(x, \Psi) \in B_R \times \mathcal{P}_1(B_R^Y)$ , hence (Q3) holds if and only if the nondiagonal elements  $\alpha_F(x, \Psi)$  and  $\alpha_L(x, \Psi)$  of the matrix representation (4.1) are nonnegative. This implies that (Q1) and (Q3) together are equivalent to ( $\alpha 1$ ), which concludes the proof.  $\square$

To each element  $\Psi$  of  $\mathcal{P}_1(\mathbb{R}^d \times \mathcal{P}_1(\{F, L\}))$  we can associate a followers' and a leaders' distribution in a natural way, as we are going to discuss in the next definition.

**DEFINITION 4.2.** *Let  $\Psi \in \mathcal{P}_1(\mathbb{R}^d \times \mathcal{P}_1(\{F, L\}))$ . The followers distribution  $\mu_\Psi^F$  associated with  $\Psi$  is the positive Borel measure on  $\mathbb{R}^d$  defined by*

$$(4.2a) \quad \mu_\Psi^F(B) := \int_{B \times \mathcal{P}_1(\{F, L\})} \lambda_F \, d\Psi(x, \lambda)$$

for each Borel set  $B \subset \mathbb{R}^d$ . Similarly, the leaders distribution  $\mu_\Psi^L$  associated with  $\Psi$  is the positive Borel measure on  $\mathbb{R}^d$  defined by

$$(4.2b) \quad \mu_\Psi^L(B) := \int_{B \times \mathcal{P}_1(\{F,L\})} (1 - \lambda_F) d\Psi(x, \lambda).$$

Both the measures defined above have a simple interpretation. For instance,  $\mu_\Psi^F(B)$  is the expected value of the number of leaders in a region  $B$  for a probability distribution  $\Psi$  on  $\mathbb{R}^d \times \mathcal{P}_1(\{F, L\})$ . We also observe that the sum of the two measures  $\mu_\Psi^F$  and  $\mu_\Psi^L$  is exactly the  $x$ -marginal of  $\Psi(x, \lambda)$ . From a practical point of view, the integrals appearing in (4.2) can be computed identifying the metric space  $(\mathcal{P}_1(\{F, L\}), \mathcal{W}_1)$  with the one-dimensional simplex  $[0, 1]$  endowed with the Euclidean distance (this is indeed an isometry). We also point out the following inequalities, whose proof is straightforward:

$$(4.3) \quad \begin{aligned} m_1(\mu_\Psi^F) + m_1(\mu_\Psi^L) &\leq m_1(\Psi) \quad \text{for all } \Psi \in \mathcal{P}_1(\mathbb{R}^d \times \mathcal{P}_1(\{F, L\})), \\ \|\mu_{\Psi_1}^F - \mu_{\Psi_2}^F\|_{BL} &\leq 2\mathcal{W}_1(\Psi_1, \Psi_2) \quad \text{for all } \Psi_1, \Psi_2 \in \mathcal{P}_1(\mathbb{R}^d \times \mathcal{P}_1(\{F, L\})), \\ \|\mu_{\Psi_1}^L - \mu_{\Psi_2}^L\|_{BL} &\leq 2\mathcal{W}_1(\Psi_1, \Psi_2) \quad \text{for all } \Psi_1, \Psi_2 \in \mathcal{P}_1(\mathbb{R}^d \times \mathcal{P}_1(\{F, L\})). \end{aligned}$$

Using the previous definition, we can also provide some relevant examples of velocity fields  $v_\Psi$  complying with assumptions (v1)–(v3).

PROPOSITION 4.3. *Let  $U = \{F, L\}$ . For  $\Psi \in \mathcal{P}_1(\mathbb{R}^d \times \mathcal{P}_1(\{F, L\}))$ , consider the velocity field*

$$(4.4) \quad v_\Psi(x, \lambda) := \lambda_F (K^{FF} \star \mu_\Psi^F + K^{LF} \star \mu_\Psi^L) + (1 - \lambda_F) (K^{FL} \star \mu_\Psi^F + K^{LL} \star \mu_\Psi^L),$$

where  $\mu_\Psi^F$  and  $\mu_\Psi^L$  are defined in (4.2) and the interaction kernels  $K^{ij}: \mathbb{R}^d \rightarrow \mathbb{R}^d$  satisfy, for  $i, j \in \{F, L\}$ ,

$$\begin{aligned} |K^{ij}(x)| &\leq M(1 + |x|) \quad \text{for all } x \in \mathbb{R}^d, \\ |K^{ij}(x_1) - K^{ij}(x_2)| &\leq L_R |x_1 - x_2| \quad \text{for all } x_1, x_2 \in B_R. \end{aligned}$$

Then,  $v_\Psi(x, \lambda)$  satisfies (v1)–(v3).

*Proof.* The result follows by a direct computation. □

Remark 4.4. The velocity field (4.4) corresponds to a particle model where each follower experiences a velocity  $K^{FF} \star \mu_\Psi^F + K^{LF} \star \mu_\Psi^L$ , which combines the action of the overall followers and leaders distribution. Similarly, each leader is under the action of the velocity field  $K^{FL} \star \mu_\Psi^F + K^{LL} \star \mu_\Psi^L$ . Hence, (4.4) is an *average* velocity of the system, weighed by the probability  $\lambda$  that a particle at  $x$  has of being a leader or a follower.

In a similar spirit to the previous proposition, transition rates  $\alpha_F$  and  $\alpha_L$  depending on  $\mu_\Psi^F$  and  $\mu_\Psi^L$  can also be considered in our setting. This is for instance the point of view taken in [2, Assumption (H4) and Appendix A], where some explicit examples were also provided. Our assumptions are actually more general than those considered there: in particular, we can allow for an explicit dependence on the space variable  $x$ .

PROPOSITION 4.5. *Let  $U = \{F, L\}$  and consider two functions  $\alpha_F, \alpha_L: \mathbb{R}^d \times \mathcal{M}_+(\mathbb{R}^d) \times \mathcal{M}_+(\mathbb{R}^d) \rightarrow [0, +\infty)$  satisfying the following assumptions:*

- there exists a constant  $M$  such that, for  $\bullet = F, L$ ,

$$(4.5) \quad 0 \leq \alpha_\bullet(x, \mu, \nu) \leq M(1 + |x| + m_1(\mu) + m_1(\nu))$$

for all  $x \in \mathbb{R}^d$  and  $(\mu, \nu) \in \mathcal{M}_+(\mathbb{R}^d) \times \mathcal{M}_+(\mathbb{R}^d)$ ;

- for all  $R > 0$ , there exist a constant  $L_R$  such that, for  $\bullet = F, L$ ,

$$(4.6) \quad \begin{aligned} & |\alpha_\bullet(x_1, \mu_1, \nu_1) - \alpha_\bullet(x_2, \mu_2, \nu_2)| \\ & \leq L_R (|x_1 - x_2| + \|\mu_1 - \mu_2\|_{BL} + \|\nu_1 - \nu_2\|_{BL}) \end{aligned}$$

for all  $x_1, x_2 \in B_R$  and  $(\mu_1, \nu_1), (\mu_2, \nu_2) \in \mathcal{M}_+(B_R) \times \mathcal{M}_+(B_R)$ .

For  $\Psi \in \mathcal{P}_1(\mathbb{R}^d \times \mathcal{P}_1(\{F, L\}))$ , define  $\mu_\Psi^F$  and  $\mu_\Psi^L$  as in (4.2). Then, the functions

$$(4.7) \quad \alpha_F(x, \Psi) := \alpha_F(x, \mu_\Psi^F, \mu_\Psi^L) \quad \text{and} \quad \alpha_L(x, \Psi) := \alpha_L(x, \mu_\Psi^F, \mu_\Psi^L)$$

satisfy Assumptions  $(\alpha 0)$ – $(\alpha 2)$  in Proposition 4.1.

*Proof.* The result follows from (4.5)–(4.6) from the inequalities in (4.3). □

The main result of this section is an existence and uniqueness result for the system of equations considered in [2], which we are going to deduce from Theorem 3.5. By doing this, we will extend the result in [2, Proposition 3.2] to the case were the transition rates  $\alpha_F, \alpha_L$  are allowed to explicitly depend on  $x$ , which was not considered there. The equations we consider are, namely,

$$(4.8) \quad \begin{cases} \partial_t \mu_t^F = -\operatorname{div}((K^F \star \mu_t^F + K^L \star \mu_t^L)\mu_t^F) - \alpha_F(x, \mu_t^F, \mu_t^L)\mu_t^F + \alpha_L(x, \mu_t^F, \mu_t^L)\mu_t^L, \\ \partial_t \mu_t^L = -\operatorname{div}((K^F \star \mu_t^F + K^L \star \mu_t^L)\mu_t^L) + \alpha_F(x, \mu_t^F, \mu_t^L)\mu_t^F - \alpha_L(x, \mu_t^F, \mu_t^L)\mu_t^L \end{cases}$$

to be solved by two positive Borel measures  $\mu^F$  and  $\mu^L$  attaining, for  $t = 0$ , an initial datum  $\bar{\mu}^F$ , and  $\bar{\mu}^L$ , respectively, which we assume to have compact support in  $\mathbb{R}^d$ . As is customary in these kinds of models, we will assume that the initial total population is normalized to 1, i.e.,

$$\bar{\mu}^F(\mathbb{R}^d) + \bar{\mu}^L(\mathbb{R}^d) = 1.$$

*Remark 4.6.* The velocity field in (4.8) corresponds to the choice  $K^{FF} = K^{FL} = K^F$  and  $K^{LF} = K^{LL} = K^L$  in (4.4). The key observation is that, in this case, the field  $v_\Psi(x, \lambda)$  shows no explicit dependence on  $\lambda$  and is simply given by

$$(4.9) \quad v_\Psi(x) = K^F \star \mu_\Psi^F + K^L \star \mu_\Psi^L.$$

This will eventually allow us to decouple (3.28) into the simpler system (4.8). Such an analysis is not possible if more than two different kernels are considered in (4.4). In that general case, the mean-field limit of the associated particle system must be formulated in terms of a solution  $\Lambda$  to (3.28), defined in the product space  $\mathbb{R}^d \times \mathcal{P}_1(\{F, L\})$ .

To proceed to the announced result, we need to recall the definition of a solution to (4.8) which has been considered in [2]. Below, the shortcut  $\mathcal{M}_c(\mathbb{R}^d)$  is used to denote a positive Borel measure having compact support in  $\mathbb{R}^d$ .

**DEFINITION 4.7** (solution of system (4.8)). *Let  $(\bar{\mu}^F, \bar{\mu}^L) \in \mathcal{M}_c(\mathbb{R}^d) \times \mathcal{M}_c(\mathbb{R}^d)$  be given, as well as  $\mu^F, \mu^L : [0, T] \rightarrow \mathcal{M}_c(\mathbb{R}^d)$ . We say that the couple  $(\mu_t^F, \mu_t^L)$  is a solution of system (4.8) with initial datum  $(\bar{\mu}^F, \bar{\mu}^L)$  when*

- (i)  $\mu_0^F = \bar{\mu}^F$  and  $\mu_0^L = \bar{\mu}^L$ ;

- (ii) for each  $i \in \{F, L\}$ , the function  $t \rightarrow \mu_t^i$  is continuous with respect to the topology of weak convergence of measures;
- (iii) there exists  $R_T > 0$  such that  $\bigcup_{t \in [0, T]} \text{supp}(\mu_t^i) \subseteq B_{R_T}$  for every  $i \in \{F, L\}$ ;
- (iv) for every  $\varphi \in C_c^1(\mathbb{R}^d)$  and  $i \in \{F, L\}$  it holds that

$$\begin{aligned} \frac{d}{dt} \int_{\mathbb{R}^d} \varphi(x) d\mu_t^i(x) &= \int_{\mathbb{R}^d} \nabla \varphi(x) \cdot \left[ \sum_{j \in \{F, L\}} (K^j \star \mu_t^j)(x) \right] d\mu_t^i(x) \\ &\quad - \int_{\mathbb{R}^d} \varphi(x) \alpha_i(x, \mu_t^F, \mu_t^L) d\mu_t^i(x) \\ &\quad + \int_{\mathbb{R}^d} \varphi(x) \alpha_{\neg i}(x, \mu_t^F, \mu_t^L) d\mu_t^{\neg i}(x) \end{aligned}$$

for almost every  $t \in [0, T]$  with

$$\neg i := \begin{cases} L & \text{if } i = F, \\ F & \text{if } i = L. \end{cases}$$

We can now state the existence and uniqueness result for system (4.8).

**THEOREM 4.8.** *Let  $U = \{F, L\}$ . Consider two functions  $\alpha_F, \alpha_L: \mathbb{R}^d \times \mathcal{M}_+(\mathbb{R}^d) \times \mathcal{M}_+(\mathbb{R}^d) \rightarrow [0, +\infty)$  satisfying (4.5)–(4.6) and two kernels  $K^F, K^L: \mathbb{R}^d \rightarrow \mathbb{R}^d$  with*

(4.10)

$$|K^F(x)| + |K^L(x)| \leq M(1 + |x|) \quad \text{for all } x \in \mathbb{R}^d,$$

$$|K^F(x_1) - K^F(x_2)| + |K^L(x_1) - K^L(x_2)| \leq L_R |x_1 - x_2| \quad \text{for all } x_1, x_2 \in B_R.$$

For  $\Psi \in \mathcal{P}_1(\mathbb{R}^d \times \mathcal{P}_1(\{F, L\}))$ , define  $\mu_\Psi^F$  and  $\mu_\Psi^L$  as in (4.2). For  $x \in \mathbb{R}^d$  and  $\Psi \in \mathcal{P}_1(\mathbb{R}^d \times \mathcal{P}_1(\{F, L\}))$ , let  $v_\Psi(x)$  and  $\mathcal{Q}(x, \Psi)$  be given by (4.9) and (4.1) respectively, and, for  $\mathcal{T}_\Psi(y) = \mathcal{Q}^*(x, \Psi)\lambda$ , consider the corresponding velocity field  $b_\Psi$  as in (3.1).

Then, if  $\Lambda \in C([0, T]; \mathcal{P}_1(\mathbb{R}^d \times \mathcal{P}_1(\{F, L\})), \mathcal{W}_1)$  is the unique solution to (3.28) starting from  $\bar{\Lambda} \in \mathcal{P}_c(\mathbb{R}^d \times \mathcal{P}_1(\{F, L\}))$ , the measures  $\mu_t^F := \mu_{\Lambda_t}^F$  and  $\mu_t^L := \mu_{\Lambda_t}^L$  are the unique solutions to (4.8) with initial data  $\bar{\mu}^F = \mu_{\bar{\Lambda}}^F$  and  $\bar{\mu}^L = \mu_{\bar{\Lambda}}^L$ .

*Proof.* We start by observing that, under our assumptions on  $\alpha_F, \alpha_L, K^F$ , and  $K^L$ , the previous results of this section ensure that the field  $b_\Psi$  complies with the requirements of Theorem 3.5, hence existence and uniqueness of the solution  $\Lambda$  to (3.28) starting from  $\bar{\Lambda}$  is guaranteed. We split the proof into two parts, proving first existence and then uniqueness of the solutions.

*Existence.* For  $\Lambda$  as above, define  $\mu_t^F := \mu_{\Lambda_t}^F$  and  $\mu_t^L := \mu_{\Lambda_t}^L$ . By definition and Lemma 3.13, conditions (i) and (iii) in Definition 4.7 are satisfied. The continuity property (ii) is instead a direct consequence of (4.3) and the continuity of  $\Lambda$  as a function of the time. We therefore only have to check (iv) in Definition 4.7. We perform the required computation only for  $\mu_t^F$ , since the one for  $\mu_t^L$  follows along similar lines. We take  $\varphi \in C_c^1(\mathbb{R}^d)$  and we define, for all  $(x, \lambda) \in \mathbb{R}^d \times \mathcal{F}(\{F, L\})$ , the test function

$$\phi(x, \lambda) = \lambda_F \varphi(x).$$

We notice that the  $(\mathbb{R}^d)^*$ -component of the Fréchet differential of  $\phi$  at  $(x, \lambda)$  is given by  $\lambda_F D\varphi(x)$ , while the action of the other component is independent of  $\lambda$  by linearity and is given by

$$\xi \mapsto \varphi(x) \xi_F.$$



We apply the definition (3.29) of Eulerian solution to the above test function  $\phi$ , which does not depend on  $t$ , and we get

$$(4.11) \quad \begin{aligned} \frac{d}{dt} \int_{\mathbb{R}^d \times \mathcal{P}_1(\{F,L\})} \lambda_F \varphi(x) \, d\Lambda_t(x, \lambda) &= \int_{\mathbb{R}^d \times \mathcal{P}_1(\{F,L\})} \lambda_F \nabla \varphi(x) \cdot v_{\Lambda_t}(x) \, d\Lambda_t(x, \lambda) \\ &+ \int_{\mathbb{R}^d \times \mathcal{P}_1(\{F,L\})} \varphi(x) (\mathcal{Q}^*(x, \Lambda_t)\lambda)_F \, d\Lambda_t(x, \lambda) \end{aligned}$$

for all  $t \in [0, T]$ . We observe now that (4.2a)–(4.2b) are equivalent to the duality relationships

$$(4.12) \quad \begin{aligned} \int_{\mathbb{R}^d} \zeta(x) \, d\mu_{\Psi}^F(x) &= \int_{\mathbb{R}^d \times \mathcal{P}_1(\{F,L\})} \lambda_F \zeta(x) \, d\Psi(x, \lambda), \\ \int_{\mathbb{R}^d} \zeta(x) \, d\mu_{\Psi}^L(x) &= \int_{\mathbb{R}^d \times \mathcal{P}_1(\{F,L\})} (1 - \lambda_F) \zeta(x) \, d\Psi(x, \lambda), \end{aligned}$$

for all  $\zeta \in C_b(\mathbb{R}^d)$ . Applying the first one to the functions  $\varphi(x)$  and  $\nabla \varphi(x) \cdot v_{\Lambda_t}(x) \in C_b(\mathbb{R}^d)$  gives<sup>4</sup>

$$(4.13) \quad \begin{aligned} \int_{\mathbb{R}^d \times \mathcal{P}_1(\{F,L\})} \lambda_F \varphi(x) \, d\Lambda_t(x, \lambda) &= \int_{\mathbb{R}^d} \varphi(x) \, d\mu_t^F(x), \\ \int_{\mathbb{R}^d \times \mathcal{P}_1(\{F,L\})} \lambda_F \nabla \varphi(x) \cdot v_{\Lambda_t}(x) \, d\Lambda_t(x, \lambda) &= \int_{\mathbb{R}^d} \nabla \varphi(x) \cdot v_{\Lambda_t}(x) \, d\mu_t^F(x). \end{aligned}$$

With Proposition 4.1, (4.13), and applying (4.12) to the functions  $-\varphi(x)\alpha_F(x, \mu_t^F, \mu_t^L)$  and  $\varphi(x)\alpha_L(x, \mu_t^F, \mu_t^L)$ , respectively, we get

$$\begin{aligned} &\int_{\mathbb{R}^d \times \mathcal{P}_1(\{F,L\})} \varphi(x) (\mathcal{Q}^*(x, \Lambda_t)\lambda)_F \, d\Lambda_t(x, \lambda) \\ &= - \int_{\mathbb{R}^d} \varphi(x) \alpha_F(x, \mu_t^F, \mu_t^L) \, d\mu_t^F(x) + \int_{\mathbb{R}^d} \varphi(x) \alpha_L(x, \mu_t^F, \mu_t^L) \, d\mu_t^L(x). \end{aligned}$$

Hence, also using (4.13) and the explicit expression (4.9) of the field  $v_{\Lambda_t}(x)$ , (4.11) is equivalent to equality (iv) in Definition 4.7 for  $i = F$ , as required. This proves the existence of a solution to (4.8).

*Uniqueness.* The proof is divided into two steps.

*Uniqueness—Step 1: Continuous dependence for an auxiliary equation.* For a given  $Z > 0$  we fix the class of Carathéodory vector fields

$$\mathcal{V}_Z := \{v \in L^\infty([0, T], \text{Lip}(\mathbb{R}^d)) : \|v\|_{L^\infty([0, T], \text{Lip}(\mathbb{R}^d))} \leq Z\}$$

and we denote as usual by  $\mathbf{Y}_v(t, s, \cdot)$  the associated transition maps, which satisfy the equalities

$$(4.14) \quad \mathbf{Y}_v(s, s, x) = x, \quad \frac{d}{dt} \mathbf{Y}_v(t, s, x) = v_t(\mathbf{Y}_v(t, s, x))$$

<sup>4</sup>Here, it is crucial that the velocity field  $v_{\Lambda_t}(x)$  only depends on  $x$ .

for every  $0 \leq s \leq t \leq T$  and  $x \in \mathbb{R}^d$ . From these, we can deduce the Grönwall-type estimates

$$(4.15) \quad \begin{aligned} |\mathbf{Y}_v(t, s, x_1) - \mathbf{Y}_v(t, s, x_2)| &\leq e^{Z(t-s)}|x_1 - x_2|, \\ |\mathbf{Y}_v(t, s, x) - \mathbf{Y}_w(t, s, x)| &\leq (t - s)e^{Z(t-s)} \sup_{s \in [0, t]} \|w_s - v_s\|_{C_b(\mathbb{R}^d)} \end{aligned}$$

for every  $0 \leq s \leq t \leq T$ ,  $x, x_1, x_2 \in \mathbb{R}^d$  and  $v, w \in \mathcal{V}_Z$ . For  $v \in \mathcal{V}_Z$  and a given narrowly continuous family  $\xi_t$  of signed measures satisfying  $|\xi_t| \leq R$  for all  $t \in [0, T]$ , we consider the inhomogeneous equation

$$(4.16) \quad \partial_t \mu_t + \operatorname{div}(v_t \mu_t) = \xi_t.$$

We claim that, for a given initial datum  $\bar{\mu} \in \mathcal{M}(\mathbb{R}^d)$ , the unique solution of (4.16) starting from  $\bar{\mu}$  is given by the variation of constants formula

$$(4.17) \quad \mu_t = \mathbf{Y}_v(t, 0, \cdot) \# \bar{\mu} + \int_0^t \mathbf{Y}_v(t, s, \cdot) \# \xi_s \, ds.$$

A direct computation using (4.14) proves indeed that the above formula provides a solution to (4.16), while uniqueness follows by taking the difference of two solutions and using the comparison principle [7, Proposition 8.1.7] for the continuity equation.

Now, if  $\sigma_t$  is another narrowly continuous family of signed measures, using (4.15) for all test functions  $\varphi$  with  $|\varphi| \leq 1$  and  $\operatorname{Lip}(\varphi) \leq 1$  we have

$$\begin{aligned} &\left| \int_0^t \int_{\mathbb{R}^d} \varphi(\mathbf{Y}_v(t, s, x)) d(\xi_s - \sigma_s) \, ds \right| \\ &\leq \left( \int_0^t e^{Z(t-s)} \, ds \right) \sup_{s \in [0, t]} \|\xi_s - \sigma_s\|_{BL} \leq t C_{T,Z} \sup_{s \in [0, t]} \|\xi_s - \sigma_s\|_{BL}. \end{aligned}$$

With this, (4.15), and (4.17), we also deduce that, if  $\nu$  solves (4.16) for another velocity field  $w \in \mathcal{V}_Z$ , another narrowly continuous family  $\sigma_t$  with  $|\sigma_t| \leq R$ , and the same initial datum  $\bar{\mu}$ , the following estimate holds true:

$$(4.18) \quad \|\mu_t - \nu_t\|_{BL} \leq t C_{\bar{\mu}, T, R, Z} \left( \sup_{s \in [0, t]} \|w_s - v_s\|_{C_b(\mathbb{R}^d)} + \sup_{s \in [0, t]} \|\xi_s - \sigma_s\|_{BL} \right),$$

where the constant  $C_{\bar{\mu}, T, R, Z}$  is given by

$$C_{\bar{\mu}, T, R, Z} := |\bar{\mu}| e^{ZT} + (1 + RT) C_{T, Z}.$$

*Uniqueness—Step 2: Conclusion.* Consider two solutions  $(\mu_t^F, \mu_t^L)$  and  $(\nu_t^F, \nu_t^L)$  of (4.8) starting from the same initial datum  $(\bar{\mu}^F, \bar{\mu}^L)$ . Observe that the equation preserves the total mass, that is,

$$\mu_t^F(\mathbb{R}^d) + \mu_t^L(\mathbb{R}^d) = \nu_t^F(\mathbb{R}^d) + \nu_t^L(\mathbb{R}^d) = 1$$

for all  $t \in [0, T]$ . This can be rewritten as

$$(4.19) \quad |\mu_t^F| + |\mu_t^L| = |\nu_t^F| + |\nu_t^L| = 1,$$

since all the above measures are positive. Now, on the one hand,  $\mu_t^F$  solves (4.16) with  $v_t = K^F \star \mu_t^F + K^L \star \mu_t^L$  and  $\xi_t = -\alpha_F(\cdot, \mu_t^F, \mu_t^L) \mu_t^F + \alpha_L(\cdot, \mu_t^F, \mu_t^L) \mu_t^L$ . On

the other hand,  $\nu_t^F$  solves (4.16) with  $v_t$  replaced by  $w_t = K^F \star \nu_t^F + K^L \star \nu_t^L$  and  $\sigma_t = -\alpha_F(\cdot, \nu_t^F, \nu_t^L)\nu_t^F + \alpha_L(\cdot, \nu_t^F, \nu_t^L)\nu_t^L$  in place of  $\xi_t$ . Since  $\mu_t^F, \mu_t^L, \nu_t^F$ , and  $\nu_t^L$  have compact support contained in a fixed ball  $B_{R_T}$  by Definition 4.7, we can assume (up to multiplying  $v_t$  and  $w_t$  by a suitable cutoff function not affecting (4.16)) that  $v$  belongs to  $\mathcal{V}_Z$  for a constant  $Z$  only depending on  $T, R_T$ , and the constant  $M$  in (4.10). Using (4.5), (4.6), and (4.19), we also get the estimates

$$\begin{aligned} |\xi_t| &\leq 2M(1 + 2R_T), \quad |\sigma_t| \leq 2M(1 + 2R_T), \\ \|\sigma_t - \xi_t\|_{BL} &\leq (M(1 + 2R_T) + 2LR_T) (\|\mu_t^F - \nu_t^F\|_{BL} + \|\mu_t^L - \nu_t^L\|_{BL}). \end{aligned}$$

In a similar way, using (4.10) we obtain

$$\|v_t - w_t\|_{C_b(\mathbb{R}^d)} \leq (M(1 + 2R_T) + L_{2R_T}) (\|\mu_t^F - \nu_t^F\|_{BL} + \|\mu_t^L - \nu_t^L\|_{BL}).$$

Combining the previous inequalities with (4.18) and (4.19), we get that there exists a constant  $C$ , only depending on  $T, R_T$ , and  $M$ , such that

$$\|\mu_\tau^F - \nu_\tau^F\|_{BL} \leq \tau C \left( \sup_{s \in [0, \tau]} \|\mu_s^F - \nu_s^F\|_{BL} + \sup_{s \in [0, \tau]} \|\mu_s^L - \nu_s^L\|_{BL} \right)$$

for all  $0 \leq \tau \leq t \leq T$ . A similar estimate also holds for  $\|\mu_\tau^L - \nu_\tau^L\|_{BL}$ . Taking the supremum in the left-hand side and summing up the resulting inequalities we obtain

$$\begin{aligned} \sup_{s \in [0, t]} \|\mu_s^F - \nu_s^F\|_{BL} + \sup_{s \in [0, t]} \|\mu_s^L - \nu_s^L\|_{BL} \\ \leq 2t C \left( \sup_{s \in [0, t]} \|\mu_s^F - \nu_s^F\|_{BL} + \sup_{s \in [0, t]} \|\mu_s^L - \nu_s^L\|_{BL} \right) \end{aligned}$$

for all  $0 \leq t \leq T$ . This implies that  $\mu_t^F = \nu_t^F$ , as well as  $\mu_t^L = \nu_t^L$  for all  $0 \leq t \leq \frac{1}{2C}$ . As  $C$  only depends on  $T, R_T$ , and  $M$ , iterating the argument a finite number of times yields uniqueness on all  $[0, T]$ .  $\square$

*Remark 4.9.* For all choices of the initial data  $\bar{\mu}^F$  and  $\bar{\mu}^L$  whose sum is a probability measure  $\bar{\mu}$  with compact support, we can construct  $\bar{\Lambda} \in \mathcal{P}_c(Y)$  so that  $\mu_{\bar{\Lambda}}^F = \bar{\mu}^F$  and  $\mu_{\bar{\Lambda}}^L = \bar{\mu}^L$ . This can be done, for instance, as follows: if  $g^F(x)$  is the Radon–Nikodym derivative of  $\bar{\mu}^F$  with respect to  $\bar{\mu}$ , we can define, for  $\bar{\mu}$ -a.e.  $x \in \mathbb{R}^d$ , the measure  $\lambda_x \in \mathcal{P}_1(\{F, L\})$  via  $\lambda_x := g^F(x)\delta_F + (1 - g^F(x))\delta_L$ . Then, the measure  $\bar{\Lambda}$ , defined by duality through

$$\int_Y \phi(y) \, d\bar{\Lambda} := \int_{\mathbb{R}^d} \phi(x, \lambda_x) \, d\bar{\mu}(x)$$

for all  $\phi \in C_b(Y)$ , satisfies the claim. Hence, Theorem 4.8 is a full existence result for system (4.8).

To conclude the section, we observe that combining the above remark with Theorems 3.5 and 4.8 we obtain a mean-field derivation of system (4.8) which extends the results of [2, section 4].

**5. Further applications.** In this section we present two situations which fit in the theory presented in section 3 and are suitable to describe possible applications. In subsection 5.1 we present the generalization of the results in section 4 to the case of a discrete and finite label space  $U$ ; in subsection 5.2 we focus on a continuum of labels and compare it with the results of [5].

**5.1. Discrete and finite spaces of labels  $U$ .** In this case, we will identify the discrete space of labels  $U = \{u_1, \dots, u_H\}$  with the set of the indices of the labels, so that our model will be  $U = \{1, \dots, H\}$ . We will endow  $U$  with the Euclidean distance (restricted to  $U$ ), namely,

$$(5.1) \quad \text{dist}(h, k) = |h - k| \quad \text{for } h, k \in U.$$

Then the space  $\text{Lip}(U)$  is an  $H$ -dimensional space spanned by the indicator functions  $\mathbb{1}_h$  for  $h \in \{1, \dots, H\}$ . Consequently, the free space  $\mathcal{F}(U)$  is the space of signed Borel measures on  $U$  whose generic element  $\xi$  is characterized by the values  $\xi_h := \xi(\{h\})$ .

Also in this section we will focus on transition operators which act linearly on  $\lambda$ . Analogously to Proposition 4.1, we have the following characterization of the operators  $\mathcal{Q}$  satisfying assumptions (Q0)–(Q3).

**PROPOSITION 5.1.** *Let  $U = \{1, \dots, H\}$  and  $\mathcal{T}_\Psi: Y \rightarrow \mathcal{F}(U)$  be of the form  $\mathcal{T}_\Psi(y) = \mathcal{Q}^*(x, \Psi)\lambda$ , for  $\mathcal{Q}(x, \Psi) \in \mathcal{L}(\text{Lip}(U); \text{Lip}(U))$ . Then  $\mathcal{T}_\Psi$  satisfies (T0)–(T3) if and only if there exist  $H^2$  functions  $\alpha_{hk}: \mathbb{R}^d \times \mathcal{P}_1(U) \rightarrow [0, +\infty)$  such that*

( $\bar{\alpha}0$ ) *for every  $(x, \Psi) \in \mathbb{R}^d \times \mathcal{P}_1(Y)$  and  $\lambda \in \mathcal{P}_1(U)$  it holds*

$$(\mathcal{Q}^*(x, \Psi)\lambda)_h = -\alpha_{hh}(x, \Psi)\lambda_h + \sum_{k \neq h} \alpha_{kh}(x, \Psi)\lambda_k \quad \text{for all } h \in U$$

with

$$(5.2) \quad \alpha_{hh}(x, \Psi) = \sum_{k \neq h} \alpha_{hk}(x, \Psi) \quad \text{for all } h \in U;$$

( $\bar{\alpha}1$ ) *for every  $(x, \Psi) \in \mathbb{R}^d \times \mathcal{P}_1(Y)$ , there exists  $M_{\mathcal{Q}} > 0$  such that*

$$0 \leq \alpha_{hk}(x, \Psi) \leq M_{\mathcal{Q}}(1 + |x| + m_1(\Psi)) \quad \text{for all } h, k \in U;$$

( $\bar{\alpha}2$ ) *for every  $R > 0$  there exists  $L_{\mathcal{Q}, R} > 0$  such that, for every  $(x^1, \Psi^1), (x^2, \Psi^2) \in B_R \times \mathcal{P}(B_R^Y)$ ,*

$$|\alpha_{hk}(x^1, \Psi^1) - \alpha_{hk}(x^2, \Psi^2)| \leq L_{\mathcal{Q}, R}(|x^1 - x^2| + \mathcal{W}_1(\Psi^1, \Psi^2)) \quad \text{for } h, k \in U.$$

*Proof.* The proof is analogous to that of Proposition 4.1. □

We notice that a matrix representation of the operator  $\mathcal{Q}$  analogous to that in (4.1) holds:

$$(5.3) \quad \mathcal{Q}(x, \Psi) = \begin{pmatrix} -\alpha_{11}(x, \Psi) & \alpha_{12}(x, \Psi) & \cdots & \alpha_{1H}(x, \Psi) \\ \alpha_{21}(x, \Psi) & -\alpha_{22}(x, \Psi) & \cdots & \alpha_{2H}(x, \Psi) \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{H1}(x, \Psi) & \alpha_{H2}(x, \Psi) & \cdots & -\alpha_{HH}(x, \Psi) \end{pmatrix},$$

where, as a consequence of (5.2), the sum of the elements on each row must give zero. In this discrete setting, the operator  $\mathcal{Q}^*(x, \Psi)$  has a matrix representation given by the transpose matrix of that representing  $\mathcal{Q}(x, \Psi)$ , so that

$$(5.4) \quad \mathcal{Q}^*(x, \Psi) = \begin{pmatrix} -\alpha_{11}(x, \Psi) & \alpha_{21}(x, \Psi) & \cdots & \alpha_{H1}(x, \Psi) \\ \alpha_{12}(x, \Psi) & -\alpha_{22}(x, \Psi) & \cdots & \alpha_{H2}(x, \Psi) \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{1H}(x, \Psi) & \alpha_{2H}(x, \Psi) & \cdots & -\alpha_{HH}(x, \Psi) \end{pmatrix}$$

and condition (5.2) implies that the sum of the elements on each column must vanish.

Definition 4.2 is adapted to the following definition.

DEFINITION 5.2. Let  $\Psi \in \mathcal{P}_1(\mathbb{R}^d \times \mathcal{P}_1(U))$  The distribution  $\mu_\Psi^h$  of the agents with label  $h \in U$  associated with  $\Psi$  is the positive Borel measure on  $\mathbb{R}^d$  defined by

$$(5.5) \quad \mu_\Psi^h(B) := \int_{B \times \mathcal{P}_1(U)} \lambda_h \, d\Psi(x, \lambda)$$

for each Borel set  $B \subset \mathbb{R}^d$ .

Also in this case, upon choosing suitable interaction kernels, the measures  $\mu_\Psi^h$  defined in (5.5) can be used to construct velocity fields (of the type in (4.4))  $v_\Psi$  satisfying (v1)–(v3).

PROPOSITION 5.3. Let  $U = \{1, \dots, H\}$ . For  $\Psi \in \mathcal{P}_1(\mathbb{R}^d \times \mathcal{P}_1(U))$ , consider the velocity field

$$(5.6) \quad v_\Psi(x, \lambda) := \sum_{h,k \in U} \lambda_k (K^{hk} \star \mu_\Psi^h),$$

where  $\mu_\Psi^h$ , for  $h \in U$ , are defined in (5.5) and the interaction kernels  $K^{hk}: \mathbb{R}^d \rightarrow \mathbb{R}^d$ , for  $h, k \in U$ , satisfy

$$\begin{aligned} |K^{hk}(x)| &\leq M(1 + |x|) \quad \text{for all } x \in \mathbb{R}^d, \\ |K^{hk}(x_1) - K^{hk}(x_2)| &\leq L_R |x_1 - x_2| \quad \text{for all } x_1, x_2 \in B_R. \end{aligned}$$

Then,  $v_\Psi(x, \lambda)$  satisfies (v1)–(v3).

*Proof.* The result follows by a direct computation. □

Remark 4.4 applies in this case as well, so that the velocity field  $v_\Psi$  defined in (5.3) can be interpreted as an average velocity of the system, weighted by the probability that a particle at  $x$  has of having label  $k$ .

Proposition 4.5 concerning transition rates  $\alpha_h$ 's depending on the  $\mu_\Psi^h$ 's and explicitly on the space variable  $x$  is generalized to the following.

PROPOSITION 5.4. Let  $U = \{1, \dots, H\}$  and consider  $H^2$  functions  $\alpha_{hk}: \mathbb{R}^d \times (\mathcal{M}_+(\mathbb{R}^d))^H \rightarrow [0, +\infty)$  for  $h, k \in U$ , satisfying the following assumptions:

- there exists a constant  $M$  such that, for all  $h, k \in U$ ,

$$(5.7) \quad 0 \leq \alpha_{hk}(x, \mu^1, \dots, \mu^H) \leq M \left( 1 + |x| + \sum_{l \in U} m_1(\mu^l) \right)$$

for all  $x \in \mathbb{R}^d$  and  $(\mu^1, \dots, \mu^H) \in (\mathcal{M}_+(\mathbb{R}^d))^H$ ;

- for all  $R > 0$ , there exist a constant  $L_R$  such that, for all  $h, k \in U$ ,

$$(5.8) \quad \begin{aligned} &|\alpha_{hk}(x_1, \mu_1^1, \dots, \mu_1^H) - \alpha_{hk}(x_2, \mu_2^1, \dots, \mu_2^H)| \\ &\leq L_R \left( |x_1 - x_2| + \sum_{l \in U} \|\mu_1^l - \mu_2^l\|_{BL} \right) \end{aligned}$$

for all  $x_1, x_2 \in B_R$ , and  $(\mu_1^1, \dots, \mu_1^H), (\mu_2^1, \dots, \mu_2^H) \in (\mathcal{M}_+(B_R))^H$ .

For  $\Psi \in \mathcal{P}_1(\mathbb{R}^d \times \mathcal{P}_1(U))$ , define  $\mu_\Psi^h$  as in (5.5), for  $h \in U$ . Then, the functions

$$(5.9) \quad \alpha_{hk}(x, \Psi) := \alpha_{hk}(x, \mu_\Psi^1, \dots, \mu_\Psi^H) \quad \text{for } h, k \in U$$

satisfy assumptions  $(\bar{\alpha}0)$ – $(\bar{\alpha}2)$  in Proposition 5.1.

*Proof.* The result follows from (5.7) and (5.8) by means of the following inequalities:

$$(5.10) \quad \begin{aligned} \sum_{l \in U} m_1(\mu_\Psi^l) &\leq m_1(\Psi) \quad \text{for all } \Psi \in \mathcal{P}_1(\mathbb{R}^d \times \mathcal{P}_1(U)), \\ \|\mu_{\Psi_1}^h - \mu_{\Psi_2}^h\|_{BL} &\leq 2\mathcal{W}_1(\Psi_1, \Psi_2) \quad \text{for all } \Psi_1, \Psi_2 \in \mathcal{P}_1(\mathbb{R}^d \times \mathcal{P}_1(U)), h \in U, \end{aligned}$$

which are a straightforward generalization of those in (4.3). □

Theorem 4.8 can be generalized to the case of a finite discrete space of labels  $U = \{1, \dots, H\}$ . To obtain system (4.8) in the current context, we have to assume that the interaction kernels  $K^{hk}: \mathbb{R}^d \rightarrow \mathbb{R}^d$  for  $h, k \in U$  introduced in Proposition 5.3 are such that

$$(5.11) \quad K^{hk} = K^h \quad \text{for all } h \in U$$

for  $H$  kernels  $K^h: \mathbb{R}^d \rightarrow \mathbb{R}^d$ . In this case, analogously to the case with two labels, the velocity field  $v_\Psi$  defined in (5.6) does not depend on  $\lambda$  anymore and has the form

$$(5.12) \quad v_\Psi(x) = \sum_{h \in U} K^h \star \mu_\Psi^h,$$

where the  $\mu_\Psi^h$ 's are defined in (5.5). Then, system (4.8) becomes a set of  $H$  equations (5.13)

$$\partial_t \mu_t^h = -\operatorname{div} \left( \left( \sum_{k \in U} K^k \star \mu_t^k \right) \mu_t^h \right) - \alpha_{hh}(x, \mu_t^1, \dots, \mu_t^H) \mu_t^h + \sum_{k \neq h} \alpha_{kh}(x, \mu_t^1, \dots, \mu_t^H) \mu_t^k,$$

to be solved for Borel measures  $\mu_h$  such that, at the initial time  $t = 0$ ,  $\mu_0^h = \bar{\mu}^h$  for  $h \in U$ , where  $\bar{\mu}^1, \dots, \bar{\mu}^H$  are given Borel measures satisfying

$$(5.14) \quad \sum_{h \in U} \bar{\mu}^h(\mathbb{R}^d) = 1.$$

We give the following definition.

**DEFINITION 5.5** (solution to system (5.13)). *Let  $(\bar{\mu}^1, \dots, \bar{\mu}^H) \in (\mathcal{M}_c(\mathbb{R}^d))^H$  be given such that (5.14) is satisfied, as well as  $\mu^1, \dots, \mu^H: [0, T] \rightarrow \mathcal{M}_c(\mathbb{R}^d)$ . We say that  $(\mu_t^1, \dots, \mu_t^H)$  is a solution to system (5.13) with initial datum  $(\bar{\mu}^1, \dots, \bar{\mu}^H)$  when*

- (i)  $\mu_0^h = \bar{\mu}^h$  for each  $h \in U$ ;
- (ii) for each  $h \in U$ , the function  $t \rightarrow \mu_t^h$  is continuous with respect to the topology of weak convergence of measures;
- (iii) there exists  $R_T > 0$  such that  $\bigcup_{t \in [0, T]} \operatorname{supp}(\mu_t^h) \subseteq B_{R_T}$  for every  $h \in U$ ;
- (iv) for every  $\varphi \in C_c^1(\mathbb{R}^d)$  and  $h \in U$  it holds that

$$\begin{aligned} \frac{d}{dt} \int_{\mathbb{R}^d} \varphi(x) d\mu_t^h(x) &= \int_{\mathbb{R}^d} \nabla \varphi(x) \cdot \left( \sum_{k \in U} (K^k \star \mu_t^k)(x) \right) d\mu_t^h(x) \\ &\quad - \int_{\mathbb{R}^d} \varphi(x) \alpha_{hh}(x, \mu_t^1, \dots, \mu_t^H) d\mu_t^h(x) \\ &\quad + \int_{\mathbb{R}^d} \varphi(x) \sum_{k \neq h} \alpha_{kh}(x, \mu_t^1, \dots, \mu_t^H) d\mu_t^k(x) \end{aligned}$$

for almost every  $t \in [0, T]$ .

We are now in a position to prove the most important result of this section, namely, how a solution to (5.13) follows from Theorem 3.5.

**THEOREM 5.6.** *Let  $U = \{1, \dots, H\}$ . Consider functions  $\alpha_{hk} : \mathbb{R}^d \times (\mathcal{M}_+(\mathbb{R}^d))^H \rightarrow [0, +\infty)$  satisfying (5.7) and (5.8) and  $H$  kernels  $K^h : \mathbb{R}^d \rightarrow \mathbb{R}^d$  with*

$$\sum_{h \in U} |K^h(x)| \leq M(1 + |x|) \quad \text{for all } x \in \mathbb{R}^d,$$

$$\sum_{h \in U} |K^h(x_1) - K^h(x_2)| \leq L_R |x_1 - x_2| \quad \text{for all } x_1, x_2 \in B_R.$$

For  $\Psi \in \mathcal{P}_1(\mathbb{R}^d \times \mathcal{P}_1(U))$ , define the  $\mu_\Psi^h$ 's as in (5.5). For  $x \in \mathbb{R}^d$  and  $\Psi \in \mathcal{P}_1(\mathbb{R}^d \times \mathcal{P}_1(U))$ , let  $v_\Psi(x)$  and  $\mathcal{Q}(x, \Psi)$  be given by (5.12) and (5.3), respectively, and, for  $\mathcal{T}_\Psi(y) = \mathcal{Q}^*(x, \Psi)\lambda$ , consider the corresponding velocity field  $b_\Psi$  as in (3.1).

Then, if  $\Lambda \in C([0, T]; \mathcal{P}_1(\mathbb{R}^d \times \mathcal{P}_1(U), \mathcal{W}_1))$  is the unique solution to (3.28) starting from  $\bar{\Lambda} \in \mathcal{P}_c(\mathbb{R}^d \times \mathcal{P}_1(U))$ , the measures  $\mu_t^h := \mu_{\Lambda_t}^h$  for  $h \in U$ , are the unique solutions to (5.13) with initial data  $\bar{\mu}^h = \mu_{\bar{\Lambda}}^h$ , for  $h \in U$ .

*Proof.* The proof is analogous to that of Theorem 4.8. □

It is useful to recall the notion of  $Q$ -matrix from the literature of Markov chains (see, e.g., [37, Chapter 2]). Let  $Q$  be an  $H \times H$  matrix; the element  $q_{hk}$  represents the transition rate from state  $h$  to  $k$ . Such a matrix is called a  $Q$ -matrix if satisfies the following conditions:

- (q1)  $q_{hk} \geq 0$ , for all  $h, k \in U, h \neq k$ ;
- (q2)  $\sum_{k \in U} q_{hk} = 0$  for all  $h \in U$ .

It is customary, in the literature on Markov chains, to complement these conditions by

- (q0)  $0 \leq -q_{hh} < +\infty$  for all  $h \in U$ ,

even though condition (q0) is a consequence of (q1) and (q2).

*Remark 5.7.* It is easy to see that the matrix in (5.3), representing the operator  $\mathcal{Q}(x, \Psi)$  for all  $(x, \Psi) \in \mathbb{R}^d \times \mathcal{P}_1(Y)$  is a  $Q$ -matrix; indeed, condition (5.2) is equivalent to (q2) and  $(\bar{a}1)$  is equivalent to (q1).

For  $(x, \Psi) \in \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d \times \mathcal{P}_1(U))$ , and for fixed initial conditions  $(x^0, \lambda^0) \in Y$ , the dynamics described by the vector field  $b_\Psi$  defined in (3.1) is

$$(5.15) \quad \begin{pmatrix} \dot{x} \\ \dot{\lambda} \end{pmatrix} = b_\Psi(x, \lambda) = \begin{pmatrix} v_\Psi(x) \\ \mathcal{Q}^*(x, \Psi)\lambda \end{pmatrix}, \quad \begin{pmatrix} x \\ \lambda \end{pmatrix}(0) = \begin{pmatrix} x^0 \\ \lambda^0 \end{pmatrix}.$$

Concerning the second equation,  $\dot{\lambda} = \mathcal{Q}^*(x, \Psi)\lambda$ , this is a linear equation in the space of measures, which can be easily integrated to give

$$(5.16) \quad \lambda(t) = \lambda_t = \mathcal{S}_t(x, \Psi)\lambda^0 \quad \text{with} \quad \mathcal{S}_t(x, \Psi) := e^{t\mathcal{Q}(x, \Psi)}.$$

A classical result in the theory of Markov chains, [37, Theorem 2.1.2] grants that the fact that  $\mathcal{Q}(x, \Psi)$  satisfies (q1)–(q2) is equivalent to the matrix  $\mathcal{S}_t(x, \Psi)$  being a *stochastic matrix* for all times  $t \geq 0$ , namely, a matrix  $S \in \mathbb{R}^{H \times H}$  that satisfies

- (s1)  $0 \leq S_{kh} < +\infty$  for all  $h, k \in U$ ;
- (s2)  $\sum_{h \in U} S_{kh} = 1$  for all  $k \in U$ .

We remark that the explicit solution given by (5.16) is available only for fixed  $(x, \Psi)$ , so that the matrix (5.3) is constant and its exponential can be computed. This would restrict the dynamics to constant solutions  $x(t) = x^0$  for all  $t$  and to a constant distribution  $\Psi \in \mathcal{P}(Y)$ .

**5.2. A continuum of labels.** We now turn to the case in which  $U$  is a compact metric space, as in the general theory developed in section 3, and that it is a continuum. We will shortly present some possible expressions of the velocity field  $b_\Psi$  defined in (3.1) in which both components feature a two-player game that determines the evolution.

Let  $\bar{y} \in Y$  and let  $\mathcal{K}: Y \times Y \rightarrow \mathbb{R}^d \times \mathcal{F}(U)$  be an interaction kernel such that, for  $\Psi \in \mathcal{P}(Y)$ ,

$$(5.17) \quad b_\Psi(y) = \int_Y \mathcal{K}(y, y') \, d\Psi(y').$$

Our aim is to study system (3.30) where the evolution is driven by the field  $b_\Psi$  defined in (5.17). Recalling (3.1), let  $\mathcal{K}_x: Y \times Y \rightarrow \mathbb{R}^d$  and  $\mathcal{K}_\lambda: Y \times Y \rightarrow \mathcal{F}(U)$  be the components of  $\mathcal{K}$ . Then (5.17) gives

$$(5.18) \quad v_\Psi(y) = \int_Y \mathcal{K}_x(y, y') \, d\Psi(y'), \quad \text{and} \quad \mathcal{T}_\Psi(y) = \int_Y \mathcal{K}_\lambda(y, y') \, d\Psi(y').$$

We furthermore assume that each of  $\mathcal{K}_x$  and  $\mathcal{K}_\lambda$  accounts for an averaging over all the strategies in  $U$ . To this aim, let  $V: (\mathbb{R}^d \times U)^2 \rightarrow \mathbb{R}^d$  be such that

$$(5.19) \quad \mathcal{K}_x(y, y') = \int_U \int_U V(x, u, x', u') \, d\lambda'(u') d\lambda(u).$$

The interpretation of the field  $V$  is the following: the quantity

$$V(x, u, x', u') \in \mathbb{R}^d$$

gives the direction that the payer at  $x$  should follow if it has label/strategy  $u$  when playing a game against a player at  $x'$  with label/strategy  $u'$ . Keeping the structure (3.30), (5.18), and (5.19) in mind, the full law of motion for the variable  $x$  is given by

$$(5.20) \quad \dot{x} = v_\Psi(y) = \int_U \int_Y \int_U V(x, u, x', u') \, d\lambda'(u') d\Psi(x', \lambda') d\lambda(u),$$

which can be interpreted in the following way: the player at  $x$  with label/strategy  $u$  plays a game with all the other players at  $x'$  with their labels/strategies  $u'$  and the velocity  $\dot{x}_t$  is determined by averaging over all the possible strategies of the opponent (the integral with respect to  $d\lambda'(u')$ ), over all the possible distributions of opponents with their distributions of strategies (the integral with respect to  $d\lambda(x', \lambda')$ ), and finally over all the possible strategies that  $x$  have at their disposal (the integral with respect to  $d\lambda(u)$ ).

We present two difference choices of  $\mathcal{T}_\Psi$  complying with our set of assumptions (T0)–(T3). The first one will be a natural generalization of the operators considered in the discrete setting of section 4, while the second one will be exactly the transition operators considered in [5] for the replicator equation.

In the former case, for  $(x, \Psi) \in \mathbb{R}^d \times \mathcal{P}_1(U)$  define the linear operator  $\mathcal{Q}(x, \Psi): \text{Lip}(U) \rightarrow \text{Lip}(U)$  by

$$(5.21) \quad (\mathcal{Q}(x, \Psi)f)(u) := \int_Y \int_U J(x, u, x', u') f(u') \, d\lambda'(u') d\Psi(x', \lambda') - g(x, \Psi)(u) f(u),$$

and set  $\mathcal{T}_\Psi(y) = \mathcal{Q}^*(x, \Psi)\lambda$ . In (5.21),  $J: (\mathbb{R}^d \times U)^2 \rightarrow \mathbb{R}$ , and the function  $g(x, \Psi) \in C(U)$  can be interpreted as a *global departure rate* from  $u$ .



In the latter case, define

$$(5.22) \quad \mathcal{T}_\Psi(y) := \left( \int_Y \int_U J(x, u, x', u') \, d\lambda'(u') d\Psi(x', \lambda') - \int_U \int_Y \int_U J(x, w, x', u') \, d\lambda'(u') d\Psi(x', \lambda') d\lambda(w) \right) \lambda,$$

which, up to the change in notation adapted to our purposes, is exactly the continuous analogue of the second equation in (1.2) considered in [5].

In the following proposition, we summarize the assumptions on  $V$ ,  $J$ , and  $g$ , so that (v1)–(v3) and (T0)–(T3) are satisfied.

PROPOSITION 5.8. *Let  $\Psi \in \mathcal{P}_1(Y)$ , let  $V : (\mathbb{R}^d \times U)^2 \rightarrow \mathbb{R}^d$ ,  $J : (\mathbb{R}^d \times U)^2 \rightarrow \mathbb{R}$ , and  $g(\cdot, \Psi) \in C(U)$  be such that*

(V1)  *$V$  is locally Lipschitz with respect to all of its variables;*

(V2)  *$V$  is sublinear in the spatial variables, namely, there exists  $C_V > 0$  such that*

$$|V(x, u, x', u')| \leq C_V(1 + |x| + |x'|) \quad \text{for all } (x, u), (x', u') \in (\mathbb{R}^d \times U)^2;$$

(J1)  *$J$  is locally Lipschitz with respect to all of its variables;*

(J2)  *$J$  is sublinear in the spatial variables, namely, there exists  $C_J > 0$  such that*

$$|J(x, u, x', u')| \leq C_J(1 + |x| + |x'|) \quad \text{for all } (x, u), (x', u') \in (\mathbb{R}^d \times U)^2;$$

(g) *for  $(x, \Psi) \in \mathbb{R}^d \times \mathcal{P}_1(Y)$ ,*

$$(5.23) \quad g(x, \Psi)(u) = \int_Y \int_U J(x, u, x', u') \, d\lambda'(u') d\Psi(x', \lambda').$$

Then

(i) *the field  $v_\Psi : Y \rightarrow \mathbb{R}^d$  given by (5.18) via (5.19) satisfies (v1)–(v3);*

(ii) *the operator  $\mathcal{T}_\Psi : Y \rightarrow \mathcal{F}(U)$  given by (5.21) satisfies (T0)–(T3) and is represented as in (5.18) by means of  $\mathcal{K}_\lambda : Y \times Y \rightarrow \mathcal{F}(U)$  such that, for  $f \in \text{Lip}(U)$ ,*

$$(5.24) \quad \langle \mathcal{K}_\lambda(y, y'), f \rangle = \int_U \int_U J(x, u, x', u') [f(u') - f(u)] \, d\lambda'(u') d\lambda(u);$$

(iii) *the operator  $\mathcal{T}_\Psi : Y \rightarrow \mathcal{F}(U)$  given by (5.22) satisfies (T0)–(T3) and is represented as in (5.18), by means of  $\mathcal{K}_\lambda : Y \times Y \rightarrow \mathcal{F}(U)$  given by*

$$(5.25) \quad \mathcal{K}_\lambda(y, y') = \left( \int_U J(x, \cdot, x', u') \, d\lambda'(u') - \int_{U \times U} J(x, w, x', u') \, d\lambda'(u') d\lambda(w) \right) \lambda.$$

*Proof.* (i) Let  $R > 0$  be fixed and let  $L_{V,R}$  be the Lipschitz constant of  $V$  on  $(B_R \times U)^2$ . Properties (v1)–(v2) are obtained from (V1) by standard estimates keeping the definition of the BL norm and the structure (5.18)–(5.19) into account; the constant  $L_{v,R}$  of (v1)–(v2) is determined by  $L_{V,R}$  and  $\text{diam } U$ . Property (v3) follows from (V1) and (V2) and the constant  $M_v$  of (v3) is determined by  $C_V$ .

(ii) Analogously to (i), let  $L_{J,R}$  be the Lipschitz constant of  $J$  on  $(B_R \times U)^2$ . Property (T2) is obtained from (J1) by standard estimates keeping the definition of the BL norm and the structure (5.21) into account; the constant  $L_{\mathcal{T},R}$  of (T2) is determined by  $L_{J,R}$  and  $\text{diam } U$ . Property (T1) is obtained also using (J2), and the constant

$M_{\mathcal{T}}$  of (T1) is determined by  $C_J$ . Property (T3) follows from the boundedness of  $J(\cdot, \cdot, x', u')$  and  $g(\cdot, \Psi)$  on  $B_R \times U$  and  $U$ , respectively.

Finally, a simple computation shows that (g) and (T0) are equivalent, so that, using (5.23), we obtain the following expression for (5.21):

$$(5.26) \quad (\mathcal{Q}(x, \Psi)f)(u) = \int_Y \int_U J(x, u, x', u')[f(u') - f(u)] d\lambda'(u') d\Psi(x', \lambda'),$$

for  $f \in \text{Lip}(U)$ . Recalling, from (3.1) and (3.30), that  $\dot{\lambda} = \mathcal{Q}^*(x, \Psi)\lambda$ , for  $f \in \text{Lip}(U)$  we have, using (5.18),

$$\begin{aligned} \langle \dot{\lambda}, f \rangle &= \langle \mathcal{Q}^*(x, \Psi)\lambda, f \rangle = \langle \lambda, \mathcal{Q}(x, \Psi)f \rangle \\ &= \left\langle \lambda, \int_Y \int_U J(x, u, x', u')[f(u') - f(u)] d\lambda'(u') d\Psi(x', \lambda') \right\rangle \\ &= \int_U \int_Y \int_U J(x, u, x', u')[f(u') - f(u)] d\lambda'(u') d\Psi(x', \lambda') d\lambda(u) \\ &= \int_Y \left[ \int_U \int_U J(x, u, x', u')[f(u') - f(u)] d\lambda'(u') d\lambda(u) \right] d\Psi(x', \lambda'), \end{aligned}$$

which is (5.24).

(iii) A direct computation shows that  $\mathcal{T}_{\Psi}$  is of the form (5.18) for  $\mathcal{K}_{\lambda}$  given by (5.25). For fixed  $y' \in Y$  we now have

$$\begin{aligned} \langle K_{\lambda}(\cdot, y'), 1 \rangle_{\mathcal{F}(U), \text{Lip}(U)} \\ = \int_{U \times U} J(x, u, x', u') d\lambda'(u') d\lambda(u) - \int_{U \times U} J(x, w, x', u') d\lambda'(u') d\lambda(w) = 0, \end{aligned}$$

which clearly implies (T0). Arguing as in [5, Proposition 3.5] we can prove that, given  $R > 0$ , for  $\theta$  sufficiently small, one has

$$\lambda + \theta \mathcal{K}_{\lambda}(y, y') \geq 0$$

whenever  $y = (x, \lambda)$  and  $y' = (x, \lambda') \in B_R^Y$ , which implies (T3). Property (T1) can be deduced from (J2), similarly to the derivation in (ii). (T2) is equivalent to Lipschitz estimates on  $J$  which can be deduced, again, arguing exactly as in [5, Proposition 3.5].  $\square$

*Remark 5.9.* Let us assume that the hypotheses of Proposition 5.8 hold; the ODE in (3.30), written componentwise, reads

$$(5.27) \quad \begin{pmatrix} \dot{x} \\ \dot{\lambda} \end{pmatrix} = \begin{pmatrix} v_{\Psi}(y) \\ \mathcal{T}_{\Psi}(y) \end{pmatrix}$$

for  $y = (x, \lambda) \in Y = \mathbb{R}^d \times \mathcal{P}(U)$  and  $\Psi \in \mathcal{P}_1(Y)$ .

A comparison with [5] is in order. In [5, equations (1.3), (1.4), and (1.8)] the ODE studied is (keeping the notations of [5])

$$(5.28) \quad \begin{pmatrix} \dot{x} \\ \dot{\sigma} \end{pmatrix} = \begin{pmatrix} a(y) \\ \Delta_{\Sigma,y}\sigma \end{pmatrix},$$

where  $y = (x, \sigma) \in C := \mathbb{R}^d \times \mathcal{P}(U)$  and  $\Sigma \in \mathcal{P}_1(C)$ . The right-hand sides are

$$(5.29) \quad \begin{aligned} a(y) &= \int_U e(x, u) \, d\sigma(u), \\ \Delta_{\Sigma,y}\sigma &= \left( \int_C \int_U J(x, u, x', u') \, d\sigma'(u') \, d\Sigma(x', \sigma') \right. \\ &\quad \left. - \int_U \int_C \int_U J(x, w, x', u') \, d\sigma'(u') \, d\Sigma(x', \sigma') \, d\sigma(w) \right) \sigma, \end{aligned}$$

where  $e: \mathbb{R}^d \times U \rightarrow \mathbb{R}^d$  and  $J: (\mathbb{R}^d \times U)^2 \rightarrow \mathbb{R}$ . The equation for  $\sigma$  in (5.28) is known in the literature of evolutionary games as the replicator equation (see, e.g., [30]).

We notice that the theory developed in section 3 includes the case in (5.29), but it also allows us to deal with a broader class of velocity fields that includes agent interaction. Indeed, the velocity field  $a: C \rightarrow \mathbb{R}^d$  in (5.29) is linear in the mixed strategy  $\sigma$  and depends only on the position  $x$  of the agent; it depends neither on the global distribution  $\Sigma$  of the system nor on any interaction between the agents. The velocity field  $v_\Psi: Y \rightarrow \mathbb{R}^d$  in (5.20), instead, additionally takes into account the global distribution  $\Psi$  of the system as well as the interaction between players through the kernel  $V$ .

**Acknowledgments.** The authors thank the hospitality of the departments of mathematics of the Politecnico di Torino, of the Università di Napoli, and of the E. Schrödinger Institute in Vienna, where this research was developed. The authors are members of the Gruppo Nazionale per l'Analisi Matematica, la Probabilità e le loro Applicazioni (GNAMPA) of the Istituto Nazionale di Alta Matematica “Francesco Severi” (INdAM).

#### REFERENCES

- [1] G. ALBI, M. BONGINI, E. CRISTIANI, AND D. KALISE, *Invisible control of self-organizing agents leaving unknown environments*, SIAM J. Appl. Math., 76 (2016), pp. 1683–1710.
- [2] G. ALBI, M. BONGINI, F. ROSSI, AND F. SOLOMBRINO, *Leader formation with mean-field birth and death models*, Math. Models Methods Appl. Sci., 29 (2019), pp. 633–679.
- [3] G. ALBI, L. PARESCHI, AND M. ZANELLA, *Boltzmann-type control of opinion consensus through leaders*, Philos. Trans. Roy. Soc. A, 372 (2014), 20140138.
- [4] G. ALBI, L. PARESCHI, AND M. ZANELLA, *Opinion dynamics over complex networks: Kinetic modelling and numerical methods*, Kinet. Relat. Models, 10 (2017), pp. 1–32.
- [5] L. AMBROSIO, M. FORNASIER, M. MORANDOTTI, AND G. SAVARÉ, *Spatially Inhomogeneous Evolutionary Games*, preprint, arXiv:1805.04027, 2018.
- [6] L. AMBROSIO AND W. GANGBO, *Hamiltonian ODEs in the Wasserstein space of probability measures*, Comm. Pure Appl. Math., 61 (2008), pp. 18–53.
- [7] L. AMBROSIO, N. GIGLI, AND G. SAVARÉ, *Gradient flows in metric spaces and in the space of probability measures*, 2nd ed., Lectures in Math. ETH Zürich. Birkhäuser, Basel, 2008.
- [8] L. AMBROSIO AND D. PUGLISI, *Linear extension operators between spaces of Lipschitz maps and optimal transport*, J. Reine Angew. Math., to appear, <https://doi.org/10.1515/crelle-2018-0037>.
- [9] L. AMBROSIO AND D. TREVISAN, *Well-posedness of Lagrangian flows and continuity equations in metric measure spaces*, Anal. PDE, 7 (2014), pp. 1179–1234.

- [10] R. F. ARENS AND J. EELLS, JR., *On embedding uniform and topological spaces*, Pacific J. Math., 6 (1956), pp. 397–403.
- [11] M. BALLERINI, N. CABIBBO, R. CANDELIER, A. CAVAGNA, E. CISBANI, I. GIARDINA, V. LECOMTE, A. ORLANDI, G. PARISI, A. PROCACCINI, M. VIALE, AND V. ZDRAVKOVIC, *Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study*, Proc. Natl. Acad. Sci. USA, 105 (2008), pp. 1232–1237.
- [12] M. BONGINI AND G. BUTTAZZO, *Optimal control problems in transport dynamics*, Math. Models Methods Appl. Sci., 27 (2017), pp. 427–451.
- [13] M. BONGINI, M. FORNASIER, M. HANSEN, AND M. MAGGIONI, *Inferring interaction rules from observations of evolutive systems I: The variational approach*, Math. Models. Methods Appl. Sci., 27 (2017), pp. 909–951.
- [14] B. BONNET AND F. ROSSI, *The Pontryagin maximum principle in the Wasserstein space*, Calc. Var. Partial Differential Equations, 58 (2019), 11.
- [15] T. BÖRGERS AND R. SARIN, *Learning through reinforcement and replicator dynamics*, J. Econom. Theory, 77 (1997), pp. 1–14.
- [16] H. BRÉZIS, *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, North-Holland, Amsterdam, 1973.
- [17] M. BURGER, R. PINNAU, C. TOTZECK, O. TSE, AND A. ROTH, *Instantaneous control of interacting particle systems in the mean-field limit*, J. Comput. Phys., 405 (2020), 109181.
- [18] P. CARDALIAGUET AND S. HADIKHANLOO, *Learning in mean field games: The fictitious play*, ESAIM Control Optim. Calc. Var., 23 (2017), pp. 569–591.
- [19] J. A. CARRILLO, J. A. CAÑIZO, AND J. ROSADO, *A well-posedness theory in measures for some kinetic models of collective behavior*, Math. Models Methods Appl. Sci., 21 (2011), pp. 515–539.
- [20] J. A. CARRILLO, Y.-P. CHOI, AND M. HAURAY, *The derivation of swarming models: Mean-field limit and Wasserstein distances*, in Collective Dynamics from Bacteria to Crowds, Springer, Vienna, 2014, pp. 1–46.
- [21] H. CARTAN, *Calcul différentiel*, Hermann, Paris, 1967.
- [22] M. CIRANT, *Multi-population mean field games systems with Neumann boundary conditions*, J. Math. Pures Appl. (9), 103 (2015), pp. 1294–1315.
- [23] J. DIEUDONNÉ, *Foundations of Modern Analysis*, Pure Appl. Math., Vol. X, Academic Press, New York, 1960.
- [24] M. DI FRANCESCO AND S. FAGIOLI, *Measure solutions for non-local interaction PDEs with two species*, Nonlinearity, 26 (2013), pp. 2777–2808.
- [25] R. DOBRUSHIN, *Vlasov equations*, Funct. Anal. Appl., 13 (1979), pp. 115–123.
- [26] M. DORIGO AND C. BLUM, *Ant colony optimization theory: A survey*, Theoret. Comput. Sci., 344 (2005), pp. 243–278.
- [27] B. DÜRING, P. MARKOWICH, J.-F. PIETSCHMANN, AND M.-T. WOLFRAM, *Boltzmann and Fokker–Planck equations modelling opinion formation in the presence of strong leaders*, Proc. R. Soc. A Math. Phys. Eng. Sci., 465 (2009), pp. 3687–3708.
- [28] M. FORNASIER, B. PICCOLI, AND F. ROSSI, *Mean-field sparse optimal control*, Philos. Trans. Roy. Soc. A, 372 (2014), 20130400.
- [29] F. GOLSE, *The mean-field limit for the dynamics of large particle systems*, in Journées “Équations aux dérivées partielles,” Forges-les-Eaux, France, 2003, Exposés Nos. I–XV, Université de Nantes, Nantes, 2003, pp. 1–47.
- [30] J. HOFBAUER AND K. SIGMUND, *Evolutionary Games and Population Dynamics*, Cambridge University Press, Cambridge, 1998.
- [31] H. HUANG, J.-G. LIU, AND J. LU, *Learning interacting particle systems: Diffusion parameter estimation for aggregation equations*, Math. Models. Methods Appl. Sci., 29 (2019), pp. 1–29.
- [32] M. KAC, *Foundations of kinetic theory*, in Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 3: Contributions to Astronomy and Physics, University of California Press, Berkeley, CA, 1956, pp. 171–197.
- [33] J. KENNEDY, *Particle swarm optimization*, in Encyclopaedia of Machine Learning, Springer, New York, 2011, pp. 760–766.
- [34] T. S. LIM, Y. LU, AND J. NOLEN, *Quantitative propagation of chaos in the bimolecular chemical reaction-diffusion model*, SIAM J. Math. Anal., to appear.
- [35] H. P. MCKEAN, *Propagation of chaos for a class of non-linear parabolic equations*, in Stochastic Differential Equations, in Lecture Ser. Differential Equations, Catholic University, Washington D.C., Air Force Office of Scientific Research, Arlington, VA, 1967, pp. 41–57.
- [36] P. MOZGUNOV, M. BECCUTI, A. HORVATH, T. JAKI, R. SIROVICH, AND E. BIBBONA, *A review of the deterministic and diffusion approximations for stochastic chemical reaction networks*, React. Kinet. Mech. Catal., 123 (2018), pp. 289–312.

- [37] J. R. NORRIS, *Markov Chains*, Cambridge University Press, New York, 1997.
- [38] K. OELSCHLÄGER, *On the derivation of reaction-diffusion equations as limit dynamics of systems of moderately interacting stochastic processes*, *Probab. Theory Related Fields*, 82 (1989), pp. 565–586.
- [39] B. PICCOLI AND F. ROSSI, *Transport equation with nonlocal velocity in Wasserstein spaces: Convergence of numerical schemes*, *Acta Appl. Math.*, 124 (2013), pp. 73–105.
- [40] B. PICCOLI AND F. ROSSI, *Measure-theoretic models for crowd dynamics*, in *Crowd Dynamics Volume 1 - Theory, Models, and Safety Problems*, N. Bellomo and L. Gibelli eds., Birkhäuser, Cham, Switzerland, 2018, pp. 137–165.
- [41] M.-N. THAI, *Birth and Death process in mean field type interaction*, *Bernoulli* 2020, to appear.
- [42] G. TOSCANI, *Kinetic models of opinion formation*, *Commun. Math. Sci.*, 4 (2006), pp. 481–496.
- [43] C. VILLANI, *Topics in Optimal Transportation*, *Grad. Stud. Math.* 58, American Mathematical Society, Providence, RI, 2003.
- [44] N. WEAVER, *Lipschitz algebras*, 2nd ed., World Scientific, Hackensack, NJ, 2018.