OPINION FITNESS AND CONVERGENCE TO CONSENSUS IN HOMOGENEOUS AND HETEROGENEOUS POPULATIONS

Mayte Pérez-Llanos

Departamento de Ecuaciones Diferenciales y Análisis Numérico
Facultad de Matemáticas, C/ Tarfia s/n
and Instituto de Matemáticas Antonio de Castro Brzezicki
Edificio Celestino Mutis, Avda. de la Reina Mercedes s/n
Universidad de Sevilla
Campus de Reina Mercedes, (41012) Sevilla, Spain

Juan Pablo Pinasco and Nicolas Saintier

IMAS UBA-CONICET and Departamento de Matemática
Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires
Av Cantilo s/n, Ciudad Universitaria
(1428) Buenos Aires, Argentina

(Communicated by Xiaoping Xue)

Abstract. In this work we study the formation of consensus in homogeneous and heterogeneous populations, and the effect of attractiveness or fitness of the opinions. We derive the corresponding kinetic equations, analyze the long time behavior of their solutions, and characterize the consensus opinion.

1. **Introduction.** In the last few years, several physicists and mathematicians devoted their attention to opinion dynamics. Different techniques were used, depending on the specific problem. For instance, computer simulations are preferred when agents contact through networks, although they appear also in mean field problems [5, 15, 32, 33, 37], Markov processes and other probabilistic tools for finitely many opinions [17, 18, 21], large systems of ordinary differential equations for active particles, and the associated Boltzmann type equations for the evolution of several observables [12, 20, 27, 29, 31, 35, 38], together with their Fokker-Planck limits. We can cite the books [8, 9, 19, 28], and the surveys [1, 24] for details.

Despite these different tools, the microscopic interactions among agents are based on sociological theories like social impact and social pressure theory [3, 13, 23], where agents modify their opinions trying to fit in some social group, and the persuasive argument theory [11, 25], where the new opinion appears after an interchange of arguments among agents.

However, the space of opinions is typically assumed as homogeneous, without considering the advantages of holding some specific opinions. Recently, in [30] we have proposed a model where the opinions are weighted by a function $\lambda(w)$ representing the attractiveness, advantages, or fitness of opinion w. If we call w, w_* the actual opinions of two interacting agents, the new opinions w', w'_* are obtained from the following microscopic interaction rule

$$\begin{aligned} w' &= w + \gamma\lambda(w_*)(w_* - w) \\ w'_* &= w_* + \gamma\lambda(w)(w - w_*). \end{aligned} \tag{1}$$

where $0 < \gamma < 1/2$ is a fixed parameter related to the strength of the interaction. Notice that the new opinion w' is obtained moving w toward w_*, thus implementing the tendency to compromise. Moreover the magnitude of the change is proportional to $\lambda(w_*)$ i.e. to the attractiveness of opinion w_*.

Let us mention briefly that in [6, 7, 22] there are two competing opinions, say $\pm L$, and an *attitude spectrum* $A = \{\pm1, \pm2, \ldots, \pm L\}$ which represents the strength of the opinion of an individual or its degree of conviction. Now, the behavior of agents depends on their attitudes, and it is not a characteristic associated with an opinion. A similar mechanism can be found in [2, 38], where the microscopic interaction rule is given by

$$\begin{aligned} w' &= w + \gamma P(w)(w_* - w) \\ w'_* &= w_* + \gamma P(w_*)(w - w_*), \end{aligned} \tag{2}$$

where P is usually equal to zero at the boundary of the space of opinions, representing that extreme opinions are difficult to change. Despite the apparent similarity between (1) and (2), both dynamics are very different, see [30].

In this work we study the long time behavior of a population interacting through rule (1), and we add more heterogeneity on the agents by introducing two parameters p and q modeling the power of persuasion and the stubbornness of each agent, as in [31].

In Section §2 we introduce the model and the mean field first order equation, satisfied by the density of agents in the space of opinions. In Section §3 we focus on the effect of the fitness of opinions in an homogeneous population, and we derive the precise value of the consensus opinion m_∞, namely

$$m_\infty = \Lambda^{-1}\Big(\int \Lambda(w)df_0(w) \Big)$$

where Λ is a primitive of λ, and f_0 is the initial distribution of agents on the space of opinions. Let us mention that we need $\lambda > c > 0$ for some constant c. This hypotheses cannot be relaxed, due to the singular behavior of the population located near the zeros of λ, and a symmetry breaking phenomena when $\lambda(m_\infty) = 0$. In Section §4 we analyze the formation of consensus in a non-homogeneous population. We include in Section §5 some agent based simulations, showing a good agreement for a finite population of the results predicted by the mean field equation.

2. Description of the model.

2.1. Opinion, persuasion ability and stubbornness parameters. Let us introduce our model of opinion formation. We consider a population composed of infinitely many agents. The opinion of an agent with respect to some statement

is represented by a real number $w \in [-1, 1]$ (meaning -1 being completely in disagreement with the statement and 1 in complete agreement).

In addition, we take into account the ability (or difficulty) of an individual to persuade another agent, denoted by $p \in [0, 1]$; if $p = 0$ the agent has no persuasion at all, while $p = 1$ corresponds to perfect orators. We also take into account the stubbornness of an agent denoted by $q \in [0, 1]$, where $q = 0$ corresponds to a stubborn agent, who will never be affected by other's opinion, and $q = 1$ entails a very volatile agent who always takes into account other agents' opinions.

Each agent is thus characterized by the three parameters (w, p, q). Moreover, we assume that the power of persuasion p and the stubbornness q remain fixed for each of the agents, who may modify their opinion w after binary encounters.

2.2. Influence functions and microscopic interactions.
Our model also takes into account the possibility that the fitness of an individual opinion may affect other agents. We introduce an influence function $\lambda(w)$, with $w \in [-1, 1]$, representing the influence exerted by an individual with opinion w. Usual examples of influence functions are

- Quadratic : $\lambda(w) = |w|^2$;
- Linear : $\lambda(w) = |w|$;
- Uniform : $\lambda(w) = 1$;
- Co-Linear : $\lambda(w) = 1 - |w|$;
- Co-Quadratic : $\lambda(w) = (1 - |w|)^2$.

Individuals which held extreme opinions will have more influence under the linear and quadratic functions (*extremist influence functions*), while the co-linear and co-quadratic functions endow more influence to agents with moderate opinions, (*centrist influence functions*).

We now describe the up-dating rules of the opinions. Consider two interacting agents with parameters (w, p, q) and (w_*, p_*, q_*) before the encounter. Denote by (w', p', q'_*) and (w'_*, p'_*, q'_*) the new values for the parameters after the interaction, respectively. As we mentioned before, the parameters p, p_*, q, q_* will remain unchanged: $p' = p$, $p'_* = p_*$, $q' = q$ and $q'_* = q_*$. Regarding the up-dating of the opinion, we propose the following rule:

$$
\begin{aligned}
w' &= w + \gamma q p_* \lambda(w_*)(w_* - w), \\
w'_* &= w_* + \gamma\, q_* p \lambda(w)(w - w_*).
\end{aligned}
\tag{3}
$$

Observe that the term $\gamma \lambda(w^*) p_* (w_* - w)$ reflects the tendency to compromise, which is proportional to the power of persuasion, p_*, and the influence exerted by the other agent opinion through $\lambda(w^*)$, as well as her/his own stubbornness, q. Here, γ is a given real number in $(0, 1/2)$ modelling the strength of the tendency to compromise.

2.3. Macroscopic kinetic model: a first order, mean field, nonlocal transport equation.
First of all we establish some notations. Let $K = [-1, 1] \times [0, 1] \times [0, 1]$ be the space of the triple $\varpi = (w, p, q)$. We denote $P(K)$ the convex cone of probability measures on K. We endow $P(K)$ with the weak* convergence defined as $f_k \to f$ if $\int \phi \, df_k \to \int \phi \, df$ for any $\phi \in C(K)$. Since K is compact, $P(K)$ is also compact. It is well-known that this topology can be metricized in various ways. We will use the Monge-Kantorovich W_1 distance defined for any $\mu, \nu \in P(K)$ by

$$
W_1(\mu, \nu) = \sup_{\phi} \int \phi \, d(\mu - \nu),
\tag{4}
$$

where the supremum is taken among the 1-Lipschitz functions $\phi : K \to \mathbb{R}$;   namely, $|\phi(x) - \phi(y)| \le |x - y|$ for any $x, y \in K$. We refer to [39] for more details on this distance.

Denote as $f_t^\gamma(\varpi)$ the distribution of agents on the triple ϖ at time $t \ge 0$ when agents interact following the rule (3). Indeed, f_t^γ is a probability measure on K, denoted as well in the sequel as df_t^γ or $f_t^\gamma(\varpi)d\varpi$. Bear in mind that f_t^γ may not necessarily be absolutely continuous with respect to Lebesgue measure. In fact, f_t^γ could be a Dirac measure.

In the case of binary interactions and assuming, as usually done, that the joint distibution $f_t^\gamma(\varpi, \varpi_*)d\varpi\, d\varpi_* = f_t^\gamma(\varpi)f_t^\gamma(\varpi_*)d\varpi\, d\varpi_*$, it can be shown (see e.g. [28]) that f_t^γ is the unique solution of a Boltzmann type equation that takes into account a mean value of all possible interactions. Namely, for any observable $\phi \in C(K)$,

$$\frac{d}{dt} \int_K \phi(\varpi)\, df_t^\gamma(\varpi) = \int_K (\phi(\varpi') - \phi(\varpi))\, df_t^\gamma(\varpi_*)df_t^\gamma(\varpi).$$

A fixed point argument yields existence and uniqueness for this equation.

Our next purpose is to describe the time evolution of this density for a given initial condition $f_0 \in P(K)$. To this end, we perform a grazing limit in the Boltzmann type equation above, that is, as the parameter γ adjusting the strength of the interactions (3), goes to zero. Indeed we can then approximate

$$\phi(\varpi') - \phi(\varpi) \approx (w' - w)\partial_w\phi(\varpi) + \frac{1}{2}(w' - w)^2\partial_{ww}\phi(\varpi)$$

$$= \gamma q p_* \lambda(w_*)(w_* - w)\partial_w\phi(\varpi) + \frac{1}{2}(\gamma q p_* \lambda(w_*))^2(w_* - w)^2\partial_{ww}\phi(\varpi)$$

to obtain, after simplification,

$$\frac{1}{\gamma}\frac{d}{dt} \int_K \phi(\varpi)\, df_t^\gamma(\varpi) \approx \int_K (m^\gamma(t) - w)\langle p\lambda\rangle^\gamma q\partial_w\phi(\varpi)\, df_t^\gamma(\varpi)$$

$$+ \frac{\gamma}{2} \int_K \partial_{ww}\phi(\varpi)D_t^\gamma(\varpi)df_t^\gamma(\varpi)$$

with

$$\langle p\lambda\rangle^\gamma = \int_K p\lambda(w)df_t^\gamma(\varpi), \quad m^\gamma(t) = \int_K \frac{p\lambda(w)}{\langle p\lambda\rangle^\gamma}w\, df_t^\gamma(\varpi),$$

$$D_t^\gamma(\varpi) = q^2 \Big(\langle p^2w^2\lambda^2\rangle - 2w\langle p^2w\lambda^2\rangle + w^2\langle p^2\lambda^2\rangle\Big).$$

Notice that $\langle p\lambda\rangle^\gamma$ is the mean value at time t of $p\lambda(w)$, and $m^\gamma(t)$ the mean opinion weighted by the normalized actions of both the power of persuasion and the influence exerted by the agents.

Rescaling time considering $\tau = \gamma t$, so that $\frac{d}{d\tau} = \frac{1}{\gamma}\frac{d}{dt}$, and letting $f_\tau^\gamma := f_t^\gamma$, $m^\gamma(\tau) := m^\gamma(t)$, we obtain the approximation

$$\frac{d}{d\tau} \int_K \phi(\varpi)\, df_\tau^\gamma(\varpi) \approx \int_K (m^\gamma(\tau) - w)\langle p\lambda\rangle^\gamma q\partial_w\phi(\varpi)\, df_\tau^\gamma(\varpi)$$

$$+ \frac{\gamma}{2} \int_K \partial_{ww}\phi(\varpi)D_\tau^\gamma(\varpi)df_\tau^\gamma(\varpi). \tag{5}$$

This procedure can be justified showing that f_τ^γ converges as $\gamma \to 0$ to f_τ, the unique solution of the first order transport equation

$$\frac{d}{d\tau} \int_K \phi(\varpi)\, df_\tau(\varpi) = \int_K (m(\tau) - w)\langle p\lambda \rangle q \partial_w \phi(\varpi)\, df_\tau(\varpi).$$

This is the core of the following Theorem, whose proof (except for minor changes) can be found in [31].

Theorem 2.1. *There exists $f \in C([0, +\infty), P(K))$ (where $P(K)$ is endowed with the weak convergence) such that, as $\gamma \to 0$, $f_\tau^\gamma := f_t^\gamma$ with $\tau = \gamma t$ converges to f_τ in $C([0, T], P(K))$ for any $T > 0$. Moreover f is the unique solution in $C([0, +\infty), P(K))$ of*

$$\int_K \phi\, df_\tau = \int_K \phi\, df_0 + \int_0^\tau \int_K (m(s) - w)\langle p\lambda \rangle q \partial_w \phi(\varpi)\, df_s(\varpi) ds, \qquad (6)$$

for any $\tau \geq 0$ and any $\phi \in C^1(K)$. Here,

$$\langle p\lambda \rangle = \int_K p\lambda(w) df_\tau(\varpi) \qquad (7)$$

is the mean value at time τ of $p\lambda(w)$, and

$$m(\tau) = \int_K \frac{p\lambda(w)}{\langle p\lambda \rangle} w\, df_\tau(\varpi). \qquad (8)$$

is the mean opinion weighted by the normalized actions of both the power of persuasion and the influence exerted by the agents.

Remark 1. Notice that (6) is the weak formulation of the transport equation

$$\partial_\tau f + \partial_w \Big((m(\tau) - w)\langle p\lambda \rangle q f \Big) = 0.$$

The rest of the paper is devoted to the study of the long-time behaviour of f_τ. The next section deals with the simplest case of an homogeneous population i.e. where there are no parameter (p, q). The case of a hetereogeneous population with stubborn people is presented in section 4.

In both cases we will prove that f_τ converges as $\tau \to +\infty$ to an explicit measure f_∞ and provide an explicit estimation of the rate of convergence in term of the W_1-distance defined in (4). We obtain in particular that the distribution of opinion converges to a Dirac mass located at an explicit limit opinion w_∞. Since $f_\tau^\gamma \to f_\tau$ as $\gamma \to 0$ we deduce that for a fixed $\gamma \ll 1$, the solution f_τ^γ of the Boltzmann equation on the time scale τ will be very close to f_∞ for $\tau \gg 1$. However for a fixed γ the approximation of the Boltzmann equation by the first order equation (6) is valid up to an error term. Indeed (5) shows that the Boltzmann equation will be well-approximated by the first order equation plus an aditional diffusion term $\frac{\gamma}{2} \partial_{ww}(D_\tau^\gamma(\varpi) f_\tau^\gamma)$. We thus expect the solution f_τ^γ of the Boltzmann equation to be very close to a smoothed version of f_∞.

This indeed happens as shown by the numerical experiments at the end of the paper. We also refer to [28] for more details on this issue.

It is worth to emphasize here an important difference with respect to the model studied in [31], which corresponds to taking $\lambda(w) = 1$ as the influence function. Indeed, when $\lambda = 1$ then $m(t) = \frac{1}{\langle p \rangle} \int_K pw\, dg_t$,   and the weight $\langle p \rangle$ is *constant in time*. However, for a non-constant λ, $\langle p\lambda \rangle$ is a priori non-constant in time. This

forced us to develop new arguments in order to tackle this more delicate situation, when studying the asymptotic behaviour of $m(t)$.

From now on we denote $f_t := f_\tau$ for ease of notation.

Before going further we recall a useful trick concerning transport equations.

2.4. **An useful change of variables.** If $f \in P([-1,1])$ is a probablity measure on $[-1,1]$, and $F : \mathbb{R} \to [0,1]$ is its cumulative distribution function, (namely $F(x) = f((-\infty, x])$ - it is a non-decreasing and right-continuous function with left limit), one can consider the generalized inverse of F, defined as $F^{-1} : [0,1] \to [-1,1]$

$$F^{-1}(\rho) = \inf \{x \in [-1,1] \text{ such that } F(x) \geq \rho\}. \tag{9}$$

Observe that F^{-1} is also non-decreasing and left-continuous with right limit in $(0,1]$. Furthermore, the following inclusion holds

$$[F^{-1}(0^+), F^{-1}(1)] \supset \operatorname{supp} f. \tag{10}$$

In addition, for any $x \in [-1,1]$ and any $\rho \in [0,1]$ we have the inequalities

$$\text{If } F(x) > 0 \text{ then } F^{-1}(F(x)) \leq x \text{ while } F(F^{-1}(\rho)) \geq \rho. \tag{11}$$

See the note of Embrechts and Hofert [16] for the above (and further) properties of F^{-1}. Moreover, it can be proved that

$$\int_0^1 \phi(F^{-1}(r)) \, dr = \int_{-1}^1 \phi(w) \, df(w), \tag{12}$$

for any ϕ integrable.

This change of variables will be a key point in the subsequent arguments. More precisely, consider $f \in C([0,\infty); P([-1,1]))$ satisfying a transport equation of the form

$$\partial_t f_t + \partial_x(V(t,x) f_t) = 0. \tag{13}$$

Denote as F_t the cumulative distribution function of f_t, and $X_t = F_t^{-1}$ its generalized inverse. With the use of (12) we can rewrite the weak form of equation (13) in a much simpler form. This is the core of the next result:

Proposition 1. *Let $v : [0, +\infty) \times [-1,1] \to \mathbb{R}$ be continuous and globally Lipschitz with respect to the second variable. Then, $f \in C([0,+\infty), P([-1,1]))$ is a weak solution of (13) in the sense that for any $\phi \in C^1([-1,1])$ and any $t > 0$,*

$$\int_{-1}^1 \phi(x) \, df_t(x) = \int_{-1}^1 \phi(x) \, df_0(x) + \int_0^t \int_{-1}^1 \phi'(x) v(s,x) \, df_s(x) ds, \tag{14}$$

if and only if for any $r \in (0,1]$, $X_t(r)$ is a solution of

$$\partial_t X_t(r) = v(t, X_t(r)). \tag{15}$$

Here, X_0 is the generalized inverse of F_0 (the cumulative distribution function of f_0).

The proof can be found essentially in Theorem 3.1 in [2]. See also [31], where it is rewritten under the point of view of the ordinary differential equation for the flux (15).

3. **The effect of attractive opinions in an homogeneous population.** In this section we evaluate the impact of the attractiveness of the opinion independently of any other consideration. Thus, an agent is now completely characterized by its opinion $w \in [-1, 1]$, and when interacting with an agent of opinion $w_* \in [-1, 1]$ the resulting opinion w' is

$$w' = w + \lambda(w_*)(w_* - w). \tag{16}$$

This synergy after encounters was indeed introduced in [30]. A simpler form that takes the model with homogeneity, allowed us to find the kinetic equations with the use of the empirical distribution for N agents, $f_N(w, t) = \frac{1}{N} \sum_{i=1}^{N} \delta_{w_i(t)}$, and taking limits as $N \to \infty$.

The resultant continuous distribution $f_t \in P([-1, 1])$ verifies the mean field first order equation

$$\int_{-1}^{1} \phi \, df_t = \int_{-1}^{1} \phi \, df_0 \quad + \int_{0}^{t} \int_{-1}^{1} (m_t - w)\langle \lambda \rangle \, \phi'(w) \, df_s(w) ds. \tag{17}$$

where

$$m_t = \int_{-1}^{1} \frac{\lambda(w)}{\langle \lambda \rangle} w \, df_t(w), \qquad \langle \lambda \rangle = \int_{-1}^{1} \lambda(w) \, df_t(w),$$

which is consistent with (6) taking $p = q = 1$.

Moreover, in [30] we identified a conserved quantity for the evolution equation (17), whenever $\lambda \in C([-1, 1])$. We observed that the function $\int_{-1}^{1} \Lambda(w) \, df_t(w)$, being Λ an antiderivative of λ, remains constant in time. Indeed, according to (17),

$$\frac{d}{dt} \int_{-1}^{1} \Lambda(w) \, df_t \quad = \quad \int_{-1}^{1} (m_t - w)\langle \lambda \rangle \Lambda'(w) \, df_t(w)$$

$$= \quad \int_{-1}^{1} (m_t - w)\langle \lambda \rangle \lambda(w) \, df_t(w).$$

Recall that $\langle \lambda \rangle m_t = \int_{-1}^{1} w \lambda(w) \, df_t(w)$, thus the right hand side vanishes and then

$$\int_{-1}^{1} \Lambda \, df_t = \int_{-1}^{1} \Lambda \, df_0 \qquad \text{for any } t \geq 0.$$

Sending $t \to +\infty$ and assuming that consensus occurs in the sense that $f_t \to \delta_{m_\infty}$ for some m_∞, we deduce

$$\Lambda(m_\infty) = \int \Lambda \, df_0. \tag{18}$$

Hence, the candidate to be the value of consensus is

$$m_\infty = \Lambda^{-1}\Big(\int_{-1}^{1} \Lambda(w) \, df_0(w) \Big).$$

In the next result we perform a rigorous proof of the dynamics contraction towards this value.

Theorem 3.1. *Assume that $\lambda : [-1, 1] \to \mathbb{R}$ is a continuous function such that, for some $\underline{\lambda} > 0$,*

$$\lambda(w) \geq \underline{\lambda} \qquad \text{for any } w \in conv(supp(f_0)), \tag{19}$$

being $conv(supp(f_0))$ the convex hull of $supp(f_0)$.

Then, there exists $m_\infty := \lim_{t \to +\infty} m_t \in [-1, 1]$ such that

$$W_1(f_t, \delta_{m_\infty}) \leq |conv(supp(f_0))| e^{-\underline{\lambda} t}. \tag{20}$$

Moreover, the limit opinion m_∞ is given by

$$m_\infty = \Lambda^{-1}\Big(\int \Lambda(w)\,df_0(w)\Big), \tag{21}$$

where Λ is an antiderivative of λ.

Before the proof, some remarks are in order:

Remark 2. Notice that we just require λ to be positive on the convex hull of the support of the initial distribution f_0. This is due to the fact that the dynamic is contractive, in the sense that $supp(f_t) \subset supp(f_0)$, $t \geq 0$.

Remark 3. In the case $\lambda \equiv 1$ we have $\Lambda(w) = w$ so that the consensus opinion m_∞ is simply the initial mean opinion, see [34].

The proof of Theorem 3.1 goes as follows:

Proof of Theorem 3.1. Let X_t be the generalized inverse of the cumulative distribution function corresponding to f_t. According to Proposition 1 we can rewrite the equation satisfied by f_t as

$$\partial_t X_t(r) = (m_t - X_t(r))\langle\lambda\rangle.$$

Moreover,

$$m_t = \int_{-1}^{1} \frac{\lambda(w)}{\langle\lambda\rangle} w\,df_t(w) \ = \int_{0}^{1} X_t(r)\frac{\lambda(X_t(r))}{\langle\lambda\rangle}\,dr.$$

Since X_t is non-decreasing,

$$m_t \leq X_t(1)\int_{0}^{1} \frac{\lambda(X_t(r))}{\langle\lambda\rangle}\,dr$$

$$= X_t(1)\int_{-1}^{1} \frac{\lambda(w)}{\langle\lambda\rangle}\,df_t(w)$$

$$= X_t(1),$$

and similarly $m_t \geq X_t(0^+)$. Thus,

$$X_t(0^+) \leq m_t \leq X_t(1).$$

Note that $\lambda \geq 0$ implies that $\langle\lambda\rangle \geq 0$. As a result $\partial_t X_t(1) \leq 0$, thus $X_t(1)$ is non-increasing. An identical argument shows that $X_t(0^+)$ is non-decreasing. Since $[X_t(0^+), X_t(1)]$ is the convex hull of $supp(f_t)$ this proves that

$$\text{conv}(\text{supp}(f_t)) \subset \text{conv}(\text{supp}(f_0)) \qquad t \geq 0.$$

Indeed, $\lambda(w) \geq \underline{\lambda} > 0$ for any $w \in \text{supp}(f_t)$. Consequently,

$$\partial_t \Big[(X_t(1) - X_t(r))^2\Big] \ = -2(X_t(1) - X_t(r))^2\langle\lambda\rangle \ \leq -2\underline{\lambda}(X_t(1) - X_t(r))^2,$$

and Gronwall's Lemma gives, for any $r \in (0, 1]$ and any $t \geq 0$,

$$|X_t(1) - X_t(r)| \leq |X_0(1) - X_0(r)|e^{-\underline{\lambda}t}.$$

In particular,

$$|X_t(1) - X_t(0^+)| \leq |X_0(1) - X_0(0^+)|e^{-\underline{\lambda}t},$$

which reveals that the length of $supp(f_t)$ goes to 0.

Recalling that $X_t(0^+) \leq m_t \leq X_t(1)$ for any t, $X_t(0^+)$ increases and $X_t(1)$ decreases. Accordingly, supp(f_t) shrinks to a limit point $\lim_{t \to +\infty} m_t := m_\infty \in [X_t(0^+), X_t(1)]$ for any t. In fact, for any $r \in (0, 1]$,

$$|X_t(r) - m_\infty| \leq |X_t(1) - X_t(0^+)| \leq |X_0(1) - X_0(0^+)| e^{-\lambda t}.$$

Inequality (20) follows now by observing that

$$W_1(f_t, \delta_{m_\infty}) \leq \int_{-1}^1 |w - m_\infty| df_t(w)$$

$$= \int_0^1 |X_t(r) - m_\infty| dr$$

$$\leq |X_0(1) - X_0(0^+)| e^{-\lambda t}.$$

It remains to show (21). Notice that Λ is C^1, Λ is positive and increasing on conv(supp(f_0)). Thus, Λ defines a bijection from the interval conv(supp(f_0)) on its image

$$A := \Lambda(\text{conv}(\text{supp}(f_0))) = [\Lambda(X_0(0^+)), \Lambda(X_0(1))].$$

Furthermore,

$$\Lambda(X_0(0^+)) \leq \int_{-1}^1 \Lambda \, df_0(w) = \int_0^1 \Lambda(X_0(r)) \, dr \leq \Lambda(X_0(1)),$$

and consequently $\int \Lambda \, df_0 \in A$. Identity (18) entails that $m_\infty \in \text{conv}(\text{supp}(f_0))$ and the proof of (21) concludes. $\qquad\square$

We would like to close this section by emphasizing some relevant considerations of the result above.

3.1. **Positivity of the influence function.** First of all, the hypothesis of strict positivity of the influence function on the convex hull of the initial density is absolutely necessary to obtain Theorem 3.1.

Indeed in [30] we already noticed singular behaviour at the roots of λ in the simulations. Bear in mind that the value m_∞ specified in (21) is expected to hold in mean. That is the reason why, even starting from identical initial conditions, simulations could give different values for m_∞, due to the random fluctuations from realization to realization.

This fact becomes more evident the larger the order of the zero is. If λ has a high order zero at z, Λ is almost a constant function close to it, and the inverse function Λ is defined although it is very sensitive to small changes. If in addition we have that

$$z = \int_{-1}^1 \Lambda(w) df_0(w),$$

two phenomena occur. One is the slow formation of consensus due to a frozen dynamics due to the small values that λ takes. The second one is a symmetry rupture, and after a long time consensus is reached above or below z.

3.2. **Smoothness of the influence function.** Let us note that the first order equation is interpreted in a weak sense, so the differentiability of λ is not required. However, in the next section we need to impose the condition $\lambda \in W^{2,\infty}([-1, 1])$, which seems to be a technical hypothesis. Indeed, we can consider $\lambda(w) = |w|^{1/2}$ and given an uniform initial distribution of agents, the consensus is reached at $m_\infty = 0$, see [30]

3.3. **Comparison with other models.** It is worth to mention a related dynamics appearing in [2, 38], where agents are influenced by their own opinion when interact,

$$\begin{cases} w' = w + \gamma P(w)(w_* - w) \\ w'_* = w_* + \gamma P(w^*)(w - w_*). \end{cases} \tag{22}$$

Typical examples of function P are non-increasing on $|w|$, representing the fact that the extremist people are more likely to remain in their believes. In fact, the roots of the function P represent the reticence of the individual that eventually adopted that opinion, to change it after subsequent encounters. Let us observe that in this model stubborn agents appear dynamically when they approach the zeros of P. The qualitative differences in the dynamic between (22) and our model (1) were shown numerically in [30]. Moreover, when consensus occurs in (22), its value is completely different from the value (21) obtained here. In the case of (22), equation (17) reads

$$\int_{-1}^{1} \phi \, df_t = \int_{-1}^{1} \phi \, df_0 \; + \int_{0}^{t} \int_{-1}^{1} P(w)(\langle w \rangle - w) \; \phi'(w) \, df_s(w) ds, \tag{23}$$

where $\langle w \rangle = \int_{-1}^{1} w \, df_s(w)$ is the mean opinion at time s. If we assume that $P(w) \geq \underline{P} > 0$, $w \in [-1,1]$, it follows that an antiderivative Π of $1/P$ (and not of λ as in our model) is conserved. Indeed

$$\frac{d}{dt} \int \Pi(w) \, df_t(w) = \int_{-1}^{1} P(w)(\langle w \rangle - w) \; \frac{1}{P(w)} \, df_s(w) = \int_{-1}^{1} (\langle w \rangle - w) \; df_s(w)$$

which is equal to 0. Consequently, if consensus occurs in the sense that $f_t \to \delta_{\tilde{m}_\infty}$ for some limit opinion $\tilde{m}_\infty$, then $\tilde{m}_\infty$ must satisfy

$$\Pi(\tilde{m}_\infty) = \int \Pi \, df_0$$

so that $\tilde{m}_\infty = \Pi^{-1}\Big(\int \Pi \, df_0 \Big)$. We present in section §5 below some numerical experiments to validate the formula for $\tilde{m}_\infty$.

4. **Asymptotic behaviour of an heterogeneous population given the fitness of opinions.** This section is devoted to study the long-time behaviour of the unique solution to the transport equation (6), arising from an initial distribution f_0 of the form

$$f_{t=0} = \alpha_0 f_0^0 + (1 - \alpha_0) f_0^1, \tag{24}$$

being $\alpha_0 \in (0,1]$ the proportion of stubborn agents in the population, and $f_0^0, f_0^1 \in P(K)$ the initial distributions of agents with parameters (w, p, q) in the stubborn and non-stubborn population, respectively. Observe that since the stubborn agents do not change their opinion in an interaction, f_t will evolve as

$$f_t = \alpha_0 f_0^0 + (1 - \alpha_0) f_t^1, \qquad t \geq 0.$$

4.1. **Heuristic idea of the limit.** Before stating our main result, we give an informal deduction of the possible value to be the limiting opinion, presuming that the non-stubborn agents reach a consensus at m_∞.

In view of the transport term, whenever consensus is reached among the non-stubborn agents, it should take place at $m_\infty := \lim_{t \to \infty} m_t$, whenever this limit

exists. To search for the candidate to m_∞, we argue as follows. Accepting that consensus is reached at m_∞, then

$$f_t^1 \to f_0^1(p,q)dpdq \otimes \delta_{m_\infty} \qquad t \to +\infty,$$

where $f_0^1(p,q)dpdq$ is the distribution of the non-stubborn population on the (p,q)-parameters. This density is constant in time since (p,q)-parameters are unaffected by the dynamics. As a result, we can pass to the limit as $t \to +\infty$ in the definition of m_t, namely

$$\langle p\lambda \rangle \, m_t = \int_K p\lambda(w)w df_t(\varpi). \tag{25}$$

On the one hand,

$$\int_K p\lambda(w)w \, df_t(\varpi) = \alpha_0 \int_{-1}^1 \int_0^1 p\lambda(w)w df_0^0(w,p)$$

$$+ (1-\alpha_0) \int_K p\lambda(w)w df_t^1(\varpi)$$

$$\to \alpha_0 \int_{-1}^1 \int_0^1 p\lambda(w)w df_0^0(w,p)$$

$$+ (1-\alpha_0)\lambda(m_\infty)m_\infty \int_K p \, df_0^1(\varpi).$$

While on the other hand,

$$\langle p\lambda \rangle = \alpha_0 \int_{-1}^1 \int_0^1 p\lambda(w) \, df_0^0(w,p) + (1-\alpha_0) \int_K p\lambda(w) \, df_t^1(\varpi)$$

$$\to \alpha_0 \int_{-1}^1 \int_0^1 p\lambda(w) \, df_0^0(w,p) + (1-\alpha_0)\lambda(m_\infty) \int_K p \, df_0^1(\varpi).$$

Taking now limits in (25) we get

$$m_\infty \left(\alpha_0 \int_{-1}^1 \int_0^1 p\lambda(w) \, df_0^0(w,p) + (1-\alpha_0)\lambda(m_\infty) \int_K p \, df_0^1(\varpi) \right)$$

$$= \alpha_0 \int_{-1}^1 \int_0^1 p\lambda(w)w \, df_0^0(w,p) + (1-\alpha_0)\lambda(m_\infty)m_\infty \int_K p \, df_0^1(\varpi).$$

Since $\alpha_0 > 0$,

$$m_\infty \int_{-1}^1 \int_0^1 p\lambda(w) \, df_0^0(w,p) = \int_{-1}^1 \int_0^1 p\lambda(w)w \, df_0^0(w,p).$$

Conclusion: *If the non-stubborn population reaches consensus, then the consensus opinion m_∞ is specified by*

$$m_\infty := \int_{-1}^1 \int_0^1 \frac{p\lambda(w)w}{\langle p\lambda \rangle_0} df_0^0(w,p), \tag{26}$$

whenever the term

$$\langle p\lambda \rangle_0 := \int_{-1}^1 \int_0^1 p\lambda(w) \, df_0^0(w,p), \tag{27}$$

that stands for the mean value of $p\lambda$ within the stubborn population, does not vanish.

This shows that, admitting long-time consensus among the non-stubborn population, its shared opinion m_∞ is the mean opinion value weighted by the normalized $p\lambda$ (the power of conviction times the influence function). Observe that this mean

value is taken just within the stubborn population. Thus, if a common limit opinion exists, it is determined by the stubborn agents.

At this stage, it is crucial to underline a relevant fact about the candidate for consensus found above. If the influence function is constant, this model lays on our previous work studied in [31], and hence the value for m_∞ specified in (26) turns out to be

$$m_\infty := \int_{-1}^{1} \int_{0}^{1} \frac{pw}{\langle p \rangle_0} df_0^0(w,p).$$

4.2. **Statement and proof of the main result.** Our main result shows that non-stubborn agents indeed reach consensus asymptotically at the value m_∞ determined by (26). We also provide an estimate on the rate of convergence towards the consensus, in terms of the W_1-distance between f_t^1 and its limit $f_0^1(p,q)dpdq \otimes \delta_{m_\infty}$.

Before stating it recall that f_0^0 and f_t^1 denote the distribution of the stubborn and non-stubborn agents on (w,p,q), and we denote $\alpha_0 \in (0,1]$ the proportion of stubborn agents in the population. By $f_{t|(p,q)}^1 \in P([-1,1])$ we understand the distribution of opinions within the group of non-stubborn agents having parameters (p,q). Its existence is guaranteed by Jirina's Theorem. There are several classical references on this subject, for example [4, 36].

We are now ready to state our main result of this section

Theorem 4.1. *Assume $f_0^1 \in P(K)$ is supported in $\{q \geq \varepsilon_0\}$ for some $\varepsilon_0 > 0$ and that the map*

$$(p,q) \in [0,1] \times [\varepsilon_0,1] \rightarrow f_{0|(p,q)}^1 \in P([-1,1]),$$

is globally Lipschitz for the W_1-distance: there exists $L > 0$ such that for any $(p,q),(p',q') \in [0,1] \times [\varepsilon_0,1]$,

$$W_1(f_{0|(p,q)}^1, f_{0|(p',q')}^1) \leq L(|q-q'| + |p-p'|). \tag{28}$$

In addition, assume that $\lambda \in W^{2,\infty}([-1,1])$ verifies that there exists $\underline{\lambda} > 0$ such that $\lambda(w) \geq \underline{\lambda}$ for any $w \in [-1,1]$.

Then, for any $t \geq 0$,

$$W_1(f_t^1, f_0^1(p,q)dpdq \otimes \delta_{m_\infty}) \leq (4 + \kappa t)e^{-\varepsilon_0 \alpha_0 \langle p\lambda \rangle_0 t} \tag{29}$$

where m_∞ is specified in (26) and stands for the mean opinion weighted by the normalized power of persuasion multiplied by the influence function within the group of stubborn agents. The mean value of $p\lambda$ among the stubborn agents, $\langle p\lambda \rangle_0$ is defined in (27). Moreover

$$\kappa = \frac{8\|\lambda\|_\infty}{\underline{\lambda}}(\|\lambda'\|_\infty + \|\lambda''\|_\infty) + 4\|\lambda'\|_\infty.$$

The global idea of the proof of Theorem 4.1 follows the lines of the proof in [31], where the case $\lambda \equiv 1$ is treated. Notice however that when $\lambda \equiv 1$ then $\langle p\lambda \rangle = \langle p \rangle$ is constant in time whereas for an arbitrary λ it is time-varying quantity. This introduces new difficulties in many of the steps into which the proof is divided.

In the first step we consider $f_{t|(p,q)}^1 \in P([-1,1])$, the conditional distribution of opinion among the agents with parameter (p,q). This conditional distribution turns out to be the unique solution to the following transport equation. Furthermore, it is (p,q)-Lipschitz with respect to the Wasserstein distance. These facts are summarized below and the proof can be easily adapted from the case λ constant [31].

Step 4.1. For any $(p,q) \in \mathrm{supp}\ (f_0(p,q)dpdq)$, $f^1_{t|(p,q)}$ is the unique solution to

$$
\begin{cases}
\partial_t f^1_{t|(p,q)} + \partial_w((m_t - w)q\langle p\lambda\rangle\ f^1_{t|(p,q)}) = 0, \\
f^1_{t=0|(p,q)} = f_{0|(p,q)},
\end{cases}
\tag{30}
$$

in $C([0,+\infty), P([-1,1]))$.

Moreover, the function $(p,q) \to f^1_{t|(p,q)}$ is Lipschitz with respect to the Wasserstein distance W_1. Namely, for any $(p,q), (p',q') \in [0,1] \times [\varepsilon_0, 1]$,

$$
W_1(f^1_{t|(p,q)}, f^1_{t|(p',q')}) \leq\ C_t(|q - q'| + |p - p'|).
$$

Furthermore, it fulfils

$$
\int_K \phi\, df^1_t = \int_0^1 \int_0^1 \left(\int_{-1}^1 \phi\, df^1_{t|(p,q)}(w) \right) df^1_0(p,q), \qquad \forall \phi \in C(K).
\tag{31}
$$

In the next item, we take advantage of the tendency to compromise modeled by the interaction rules (3) to prove that non-stubborn agents with given (p,q) parameters tend to synchronize their opinions. Conditioning to values (p,q) we declare

$$
\langle \lambda \rangle_{(p,q)} = \int_{-1}^1 \lambda(w)\, df^1_{t|(p,q)}(w),
\tag{32}
$$

and

$$
m(t,p,q) = \int_{-1}^1 \frac{\lambda(w)}{\langle \lambda \rangle_{(p,q)}} w\, df^1_{t|(p,q)}(w),
\tag{33}
$$

the mean value of λ and the mean opinion among the agents with parameter $(p,q) \in [0,1] \times [\varepsilon_0, 1]$, respectively.

Step 4.2. For any $(p,q) \in [0,1] \times [\varepsilon_0, 1]$ there holds

$$
W_1(f^1_{t|(p,q)}, \delta_{m(t,p,q)}) \leq 2e^{-\varepsilon_0 \alpha_0 \langle p\lambda \rangle_0 t} \qquad t \geq 0,
\tag{34}
$$

being $\langle p\lambda \rangle_0$ defined in (27).

Proof. Let X_t be the generalized inverse of the cumulative distribution function corresponding to $f_{t|(p,q)}$. According to Proposition 1 we can rewrite (30) as

$$
\partial_t X_t(r) = (m_t - X_t(r))q\langle p\lambda \rangle.
$$

In particular

$$
\begin{aligned}
\partial_t \left[(X_t(1) - X_t(r))^2 \right] &= -2(X_t(1) - X_t(r))^2 q\langle p\lambda \rangle \\
&\leq -2(X_t(1) - X_t(r))^2 \varepsilon_0 \langle p\lambda \rangle.
\end{aligned}
$$

The nonnegativity of λ entails that

$$
\begin{aligned}
\langle p\lambda \rangle &= \int p\lambda(w)\, df_t \\
&= \alpha_0 \langle p\lambda \rangle_0 + (1 - \alpha_0)\ \int p\lambda(w)\, df^1_t \\
&\geq \alpha_0 \langle p\lambda \rangle_0,
\end{aligned}
$$

hence

$$
\partial_t \left[(X_t(1) - X_t(r))^2 \right]\ \leq\ -2\alpha_0 \varepsilon_0 (X_t(1) - X_t(r))^2 \langle p\lambda \rangle_0.
$$

Gronwall's Lemma implies now that

$$
|X_t(1) - X_t(r)| \leq |X_0(1) - X_0(r)|e^{-\alpha_0 \varepsilon_0 \langle p\lambda \rangle_0 t},
\tag{35}
$$

which sending $r \to 0^+$ gives

$$|X_t^i(1) - X_t^i(0^+)| \leq |X_0^i(1) - X_0^i(0^+)|e^{-\alpha_0 \varepsilon_0 \langle p\lambda \rangle_0 t}. \tag{36}$$

This shows that the length of the support of $f_{t|(p,q)}$ goes to 0.

Invoking (12), we express $m(t,p,q)$ in terms of the generalized inverse,

$$m(t,p,q) = \int_{-1}^{1} \frac{\lambda(w)}{\langle \lambda \rangle_{(p,q)}} w \, df_{t|(p,q)}(w)$$

$$= \int_0^1 X_t(r) \frac{\lambda(X_t(r))}{\langle \lambda \rangle_{(p,q)}} \, dr.$$

Since X_t is non-decreasing,

$$m(t,p,q) \leq X_t(1) \int_0^1 \frac{\lambda(X_t(r))}{\langle \lambda \rangle_{(p,q)}} \, dr$$

$$= X_t(1) \int_{-1}^{1} \frac{\lambda(w)}{\langle \lambda \rangle_{(p,q)}} \, df_{t|(p,q)}(w)$$

$$= X_t(1),$$

and similarly $m(t,p,q) \geq X_t(0^+)$. Thus,

$$X_t(0^+) \leq m_t^i \leq X_t^i(1). \tag{37}$$

Therefore, for any $r \in (0,1]$,

$$
\begin{aligned}
|X_t(r) - m(t,p,q)| &\leq |X_t(1) - X_t(0^+)| \\
&\leq |X_0(1) - X_0(0^+)|e^{-\alpha_0 \varepsilon_0 \langle p\lambda \rangle_0 t}.
\end{aligned}
$$

The proof now concludes by noticing that

$$
\begin{aligned}
W_1(f_{t|(p,q)}, \delta_{m(t,p,q)}) &\leq \int_{-1}^{1} |w - m(t,p,q)| df_{t|(p,q)}(w) \\
&= \int_0^1 |X_t(r) - m(t,p,q)| dr \\
&\leq |X_0(1) - X_0(0^+)|e^{-\alpha_0 \varepsilon_0 \langle p\lambda \rangle_0 t}.
\end{aligned}
$$

$\square$

As a result of the previous step, it is desirable to study the asymptotic behavior of the function $m(t,.)$ as $t \to +\infty$.

Step 4.3. For any $t \geq 0$ and any $(p,q) \in [0,1] \times [\varepsilon_0, 1]$ the function $m(t,p,q)$ declared in (33) satisfies

$$
\begin{aligned}
\partial_t m(t,p,q) \\
= \ & \alpha_0 q \langle p\lambda \rangle_0 [m_\infty - m(t,p,q)] \\
& + (1 - \alpha_0) q \int_{-1}^{1} p' \lambda(m(t,p',q'))[m(t,p',q') - m(t,p,q)] \, df_0^1(p',q') \\
& + R(t,p,q)
\end{aligned}
\tag{38}
$$

where m_∞ and $\langle p\lambda \rangle_0$ are given in (26) and (27) respectively, and

$$|R(t,p,q)| \leq \left\{ \frac{8\|\lambda\|_\infty}{\underline{\lambda}} (\|\lambda'\|_\infty + \|\lambda''\|_\infty) + 4\|\lambda'\|_\infty \right\} e^{-\varepsilon_0 \alpha_0 \langle p\lambda \rangle_0 t}. \tag{39}$$

Proof. Using equation (30), the evolution in time of the conditioned mean $\langle\lambda\rangle_{(p,q)}$ defined in (32) behaves as

$$\frac{d}{dt}\langle\lambda\rangle_{(p,q)} = \frac{d}{dt}\int_{-1}^{1}\lambda(w)\,df_{t|(p,q)}^{1}(w) = q\langle p\lambda\rangle\int_{-1}^{1}(m_t - w)\lambda'(w)\,df_{t|(p,q)}^{1}(w)$$

and

$$\frac{d}{dt}\int_{-1}^{1}w\lambda(w)\,df_{t|(p,q)}^{1} = q\langle p\lambda\rangle\int_{-1}^{1}(m_t - w)(w\lambda'(w) + \lambda(w))\,df_{t|(p,q)}^{1}(w).$$

Thus,

$$\frac{\partial}{\partial t}\,m(t,p,q) = \frac{\partial}{\partial t}\Big\{\frac{1}{\langle\lambda\rangle_{(p,q)}}\int_{-1}^{1}w\lambda(w)\,df_{t|(p,q)}^{1}(w)\Big\}$$

$$= \frac{q\langle p\lambda\rangle}{\langle\lambda\rangle_{(p,q)}}\Big\{\int_{-1}^{1}(m_t - w)\lambda'(w)(w - m(t,p,q))\,df_{t|(p,q)}^{1}(w)$$

$$+ \int_{-1}^{1}(m_t - w)\lambda(w)\,df_{t|(p,q)}^{1}(w)\Big\}.$$

Summing up,

$$\frac{\partial}{\partial t}\,m(t,p,q) = \frac{q\langle p\lambda\rangle}{\langle\lambda\rangle_{(p,q)}}\int_{-1}^{1}(m_t - w)\lambda'(w)(w - m(t,p,q))\,df_{t|(p,q)}^{1}(w)$$

$$+ q\langle p\lambda\rangle(m_t - m(t,p,q)). \tag{40}$$

Estimate (34) allows to bound the intergral in the r.h.s. as follows. Let $\phi(w) = (m_t - w)\lambda'(w)(w - m(t,p,q))$ so that

$$\int_{-1}^{1}(m_t - w)\lambda'(w)(w - m(t,p,q))\,df_{t|(p,q)}^{1}(w)$$

$$= \int_{-1}^{1}\phi(w) - \phi(m(t,p,q))\,df_{t|(p,q)}^{1}(w)$$

$$= \int_{-1}^{1}\phi(w)\,(df_{t|(p,q)}^{1}(w) - \delta_{m(t,p,q)}).$$

According to the definition of the W_1-distance we obtain

$$\int_{-1}^{1}(m_t - w)\lambda'(w)(w - m(t,p,q))\,df_{t|(p,q)}^{1}(w) \le Lip(\phi)W_1(f_{t|(p,q)}^{1}, \delta_{m(t,p,q)}).$$

It is easily seen that $Lip(\phi) \le 4(\|\lambda'\|_\infty + \|\lambda''\|_\infty)$. Accordingly to (34), we deduce

$$\int_{-1}^{1}(m_t - w)\lambda'(w)(w - m(t,p,q))\,df_{t|(p,q)}^{1}(w) \le 8(\|\lambda'\|_\infty + \|\lambda''\|_\infty)e^{-\varepsilon_0\alpha_0\langle p\lambda\rangle_0 t}.$$

Thus,

$$\frac{\partial}{\partial t}\,m(t,p,q) = q\langle p\lambda\rangle(m_t - m(t,p,q)) + \tilde{R}(t,p,q)$$

with

$$|\tilde{R}(t,p,q)| \le \frac{q\langle p\lambda\rangle}{\langle\lambda\rangle_{(p,q)}}8(\|\lambda'\|_\infty + \|\lambda''\|_\infty)e^{-\varepsilon_0\alpha_0\langle p\lambda\rangle_0 t}$$

$$\le \frac{8\|\lambda\|_\infty}{\underline{\lambda}}(\|\lambda'\|_\infty + \|\lambda''\|_\infty)e^{-\varepsilon_0\alpha_0\langle p\lambda\rangle_0 t},$$

given that the assumption $\lambda(w) \geq \underline{\lambda} > 0$ for any $w \in [-1, 1]$ entails that $\langle \lambda \rangle_{(p,q)} \geq \underline{\lambda} > 0$ for any (p, q).

We now focus on the second term in the right hand side of (40). First,

$$\langle p\lambda \rangle m_t \quad = \quad \langle pw\lambda \rangle \ = \alpha_0 \langle pw\lambda \rangle_0 \ + (1 - \alpha_0) \int_K p'w'\lambda(w') \, df_t^1(\varpi'),$$

where the integral can be written using (31) as

$$\int_0^1 \int_0^1 p' \Big(\int_{-1}^1 w'\lambda(w') \, df_{t|(p',q')}^1(w) \Big) \, df_0^1(p',q')$$

$$= \int_0^1 \int_0^1 p' \langle \lambda \rangle_{(p',q')} m(t, p', q') \ df_0^1(p', q').$$

Recalling in addition that $\langle pw\lambda \rangle_0 = \langle p\lambda \rangle_0 m_\infty$ with m_∞ given in (26), we obtain

$$\langle p\lambda \rangle m_t \ = \alpha_0 \langle p\lambda \rangle_0 m_\infty \ + (1 - \alpha_0) \int_0^1 \int_0^1 p' \langle \lambda \rangle_{(p',q')} m(t, p', q') \ df_0^1(p', q').$$

On the other hand,

$$\langle p\lambda \rangle \ = \alpha_0 \ \langle p\lambda \rangle_0 \ + (1 - \alpha_0) \int_K p'\lambda(w') \, df_t^1(\varpi')$$

$$= \alpha_0 \ \langle p\lambda \rangle_0 \ + (1 - \alpha_0) \int_0^1 \int_0^1 p' \Big(\int_{-1}^1 \lambda(w') \, df_{t|(p',q')}^1(w) \Big) \, df_0^1(p', q')$$

$$= \alpha_0 \ \langle p\lambda \rangle_0 \ + (1 - \alpha_0) \int_0^1 \int_0^1 p' \langle \lambda \rangle_{(p',q')} \ df_0^1(p', q').$$

Merging the former identities together,

$$\frac{\partial}{\partial t} m(t, p, q) = \alpha_0 q \langle p\lambda \rangle_0 [m_\infty - m(t, p, q)]$$

$$+ (1 - \alpha_0) q \int_0^1 \int_0^1 p' \langle \lambda \rangle_{(p',q')} [m(t, p', q') - m(t, p, q)] \, df_0^1(p', q')$$

$$+ \tilde{R}(t, p, q).$$

According to (34) we have for any (p, q) that

$$\left| \langle \lambda \rangle_{(p,q)} - \ \lambda(m(t, p, q)) \right| \quad = \quad |(f_{t|(p,q)}^1 - \delta_{m(t,p,q)}, \lambda)|$$

$$\leq \quad 2e^{-\varepsilon_0 \alpha_0 \langle p\lambda \rangle_0 t} Lip(\lambda)$$

$$\leq \quad 2e^{-\varepsilon_0 \alpha_0 \langle p\lambda \rangle_0 t} \|\lambda'\|_\infty.$$

As a result,

$$\frac{\partial}{\partial t} m(t, p, q) = \alpha_0 q \langle p\lambda \rangle_0 [m_\infty - m(t, p, q)]$$

$$+ (1 - \alpha_0) q \int p'\lambda(m(t, p', q'))[m(t, p', q') - m(t, p, q)] \, df_0^1(p', q')$$

$$+ \tilde{R}(t, p, q) + \hat{R}(t, p, q)$$

with

$$\hat{R}(t, p, q) \leq (1 - \alpha_0) q 2 e^{-\varepsilon_0 \alpha_0 \langle p\lambda \rangle_0 t} \|\lambda'\|_\infty$$

$$\times \int_0^1 \int_0^1 \ p'|m(t, p', q') - m(t, p, q)| \, df_0^1(p', q')$$

$$\leq 4e^{-\varepsilon_0\alpha_0\langle p\lambda\rangle_0 t}\|\lambda'\|_\infty$$

where we used that $m(t,p,q) \in [-1,1]$ for any (t,p,q). We deduce (38) taking $R(t,p,q) = \tilde{R}(t,p,q) + \hat{R}(t,p,q)$.   □

To understand intuitively the infinite system of equations (38) it is useful to consider the elementary situation where only a finite number of values for (p,q), $q \neq 0$, are present in the population, namely $(p_1,q_1),...,(p_N,q_N)$. Then, f_t^1 takes the simpler form

$$f_t^1 = \sum_{i=1}^N \alpha_i g_t^i(w)dw \otimes \delta_{p=p_i,q=q_i},$$

where $g_t^i := f_{t|(p_i,q_i)}$ is the distribution of opinion in the (p_i,q_i)-population and $\alpha_i \in (0,1]$ is the proportion of (p_i,q_i) agents in the non-stubborn population. Letting $m_t^i := m(t,p_i,q_i)$ we can rewrite the system (38) as

$$\frac{d}{dt}m_t^i = A - Bm_t^i + \sum_{j=1}^N c_j\lambda(m_t^j)(m_t^j - m_t^i) + R_i(t) \qquad (41)$$

where $R_t^i := R(t,p_i,q_i)$, $A = \alpha_0\langle pw\lambda\rangle_0$, $B = \alpha_0\langle p\lambda\rangle_0$, $c_j = (1-\alpha_0)\alpha_j p_j$. The right hand side of (41) is composed of three terms. The first one $A - Bm_t^i$ drives m_t^i towards $A/B = m_\infty$, which is the desired asymptotic state. The sum includes a coupling between m_t^i and all the m_t^j, $j = 1,\ldots,N$, whose effect contributes to synchronize them as in, e.g. the Cucker-Smale model [14]. The error term decreases exponentially fast to 0, thus it lacks relevance in the asymptotic behaviour. We thus expect that $\lim_{t\to+\infty} m_t^i = m_\infty$ for any $i = 1,..,N$. We will prove in the sequel that this intuition is indeed correct in general.

The regularity of $m(t,p,q)$ with respect to (p,q) plays an important role in the convergence of $m(t,p,q)$ to m_∞ as $t \to +\infty$.

Step 4.4. For any $t \geq 0$, the function $(p,q) \in supp\,(f_0^1(p,q)dpdq) \to m(t,p,q)$ is Lipschitz.

Proof. According to Step 4.1, $(p,q) \to f_{t|(p,q)}^1$ is Lipschitz with respect to the Wasserstein distance W_1: for any $(p,q),(p',q') \in [0,1] \times [\varepsilon_0,1]$,

$$W_1(f_{t|(p,q)}^1, f_{t|(p',q')}^1) \leq C_t(|q-q'| + |p-p'|).$$

The functions $\langle\lambda\rangle_{(p,q)}$ and $\langle w\lambda\rangle_{(p,q)}$ are also Lipschitz in (p,q) for a given t, whenever we assume that λ is Lipschitz. Indeed for a given $t \geq 0$ and any $(p,q),(p',q')$,

$$\begin{aligned}|\langle\lambda\rangle_{(p,q)} - \langle\lambda\rangle_{(p',q')}| &= |(f_{t|(p,q)}^1 - f_{t|(p',q')}^1, \lambda)| \\ &\leq Lip(\lambda)W_1(f_{t|(p,q)}^1, f_{t|(p',q')}^1) \\ &\leq C_t\|\lambda\|_\infty(|p-p'| + |q-q'|)\end{aligned}$$

and in the same way

$$|\langle w\lambda\rangle_{(p,q)} - \langle w\lambda\rangle_{(p',q')}| \leq C_t Lip(w\lambda(w))\,(|p-p'| + |q-q'|).$$

Consequently, for a given $t \geq 0$, the functions $\langle\lambda\rangle_{(p,q)}$ and $\langle w\lambda\rangle_{(p,q)}$ are continuous in (p,q). Since $\langle\lambda\rangle_{(p,q)} \geq \underline{\lambda} > 0$ for any (p,q), it follows that $m(t,p,q) = \langle w\lambda\rangle_{(p,q)}/\langle\lambda\rangle_{(p,q)}$ is continuous in (p,q) for any $t \geq 0$.   □

Step 4.5. For any $(p, q) \in \text{supp}\,(f_0^1(p, q)dpdq)$ and any $t \geq 0$ it holds that

$$|m(t, p, q) - m_\infty| \leq e^{-\varepsilon_0 \alpha_0 \langle p\lambda \rangle_0 t}(|m(0, p, q) - m_\infty| + \kappa t)$$

where

$$\kappa = \frac{8\|\lambda\|_\infty}{\underline{\lambda}}(\|\lambda'\|_\infty + \|\lambda''\|_\infty) + 4\|\lambda'\|_\infty.$$

Proof. Relation (38) implies that for any $q \in [\varepsilon_0, 1]$ and $t \geq 0$,

$$
\begin{aligned}
&\frac{1}{2}\frac{\partial}{\partial t}|m(t, p, q) - m_\infty|^2 \\
&= \partial_t m(t, p, q)[m(t, p, q) - m_\infty] \\
&= -q\alpha_0\langle p\lambda \rangle_0[m_\infty - m(t, p, q)]^2 \\
&\quad + q(1 - \alpha_0)[m(t, p, q) - m_\infty] \\
&\qquad \times \int_0^1 \int_0^1 p'\lambda(m(t, p', q'))[m(t, p', q') - m(t, p, q)]\, df_0^1(p', q') \\
&\quad + R(t, p, q)[m(t, p, q) - m_\infty].
\end{aligned}
\tag{42}
$$

Recall that $m(t, .)$ is continuous by the previous Step and $\text{supp}(f_0^1(p, q)dpdq)$ is compact. Take some (p^*, q^*) such that

$$\max_{\text{supp}(f_0^1(p,q)dpdq)} |m(t, .) - m_\infty| = |m(t, p^*, q^*) - m_\infty|.$$

In particular, we can write (42) at (p^*, q^*) as

$$
\begin{aligned}
&\frac{1}{2}\frac{\partial}{\partial t}|m(t, .) - m_\infty|^2_{|(p^*,q^*)} \\
&= -q^*\alpha_0\langle p\lambda \rangle_0[m_\infty - m(t, p^*, q^*)]^2 \\
&\quad + q^*(1 - \alpha_0)[m(t, p^*, q^*) - m_\infty] \\
&\qquad \times \int_0^1 \int_0^1 p'\lambda(m(t, p', q'))[m(t, p', q') - m_\infty]\, df_0^1(p', q') \\
&\quad - q^*(1 - \alpha_0)[m(t, p^*, q^*) - m_\infty]^2 \int_0^1 \int_0^1 p'\lambda(m(t, p', q'))\, df_0^1(p', q') \\
&\quad + R(t, p^*, q^*)[m(t, p^*, q^*) - m_\infty] \\
&=: I + II + III + IV.
\end{aligned}
$$

The choice of q^* assures that

$$
\begin{aligned}
II &\leq q^*(1 - \alpha_0)|m(t, p^*, q^*) - m_\infty|^2 \int_0^1 \int_0^1 p'\lambda(m(t, p', q'))\, df_0^1(p', q') \\
&= -III.
\end{aligned}
$$

The cancelation of these two terms and $q^* \geq \varepsilon_0$ gives

$$
\begin{aligned}
\frac{\partial}{\partial t}|m(t, .) - m_\infty|^2_{|(p^*,q^*)} \leq {}&- 2\varepsilon_0\alpha_0\langle p\lambda \rangle_0|m_\infty - m(t, p^*, q^*)|^2 \\
&+ 2R(t, p^*, q^*)|m(t, p^*, q^*) - m_\infty|.
\end{aligned}
\tag{43}
$$

Let $h(t; (p, q)) = |m(t, p, q) - m_\infty|^2$. Notice that $t \mapsto h(t; (p, q))$ is a C^1 function, since m is C^1 in time. Moreover, from (42) it follows that $|\partial_t h(t; (p, q))| \leq C$. Since $h(t, .)$ is continuous for any t by the previous Step, we obtain that h is continuous

in (t, p, q). According to the Measurable Selection Theorem (see e.g. §18.19 in [10]) we can choose (p^*, q^*) to be a measurable function of t.

The Envelope Theorem (see Theorem 4.2 below) ensures that the function $V(t)$ defined by

$$V(t) := \max_{(p,q)\in \text{supp}(f_0^1)} h(t; (p,q))$$

is absolutely continuous with derivative

$$V'(t) = \partial_t \left(|m(t, p^*, q^*) - m_\infty|^2 \right) \qquad a.e.$$

Furthermore, in view of (43) and (39)

$$V'(t) \leq -2\varepsilon_0 \alpha_0 \langle p\lambda\rangle_0 V(t) + 2\kappa \sqrt{V(t)} e^{-\varepsilon_0 \alpha_0 \langle p\lambda\rangle_0 t},$$

where

$$\kappa = \frac{8\|\lambda\|_\infty}{\underline{\lambda}}(\|\lambda'\|_\infty + \|\lambda''\|_\infty) + 4\|\lambda'\|_\infty.$$

Dividing by $2\sqrt{V}$ we obtain

$$(\sqrt{V})'(t) \leq -\varepsilon_0 \alpha_0 \langle p\lambda\rangle_0 \sqrt{V(t)} + \kappa e^{-\varepsilon_0 \alpha_0 \langle p\lambda\rangle_0 t}$$

which integrated gives

$$\sqrt{V(t)} \leq e^{-\varepsilon_0 \alpha_0 \langle p\lambda\rangle_0 t}\sqrt{V(0)} + \int_0^t e^{-\varepsilon_0 \alpha_0 \langle p\lambda\rangle_0 (t-s)} \kappa e^{-\varepsilon_0 \alpha_0 \langle p\lambda\rangle_0 s} \, ds$$

$$= e^{-\varepsilon_0 \alpha_0 \langle p\lambda\rangle_0 t}(\sqrt{V(0)} + \kappa t).$$

This shows that

$$|m(t, p, q) - m_\infty| \leq e^{-\varepsilon_0 \alpha_0 \langle p\lambda\rangle_0 t}(|m(0, p, q) - m_\infty| + \kappa t),$$

for any $(p, q) \in \text{supp}(f_0^1(p, q)dpdq)$, as desired. $\qquad\square$

In the course of the proof of the previous step we used the following envelope Theorem, due to Milgrom and Segal in [26]:

Theorem 4.2. *Consider the function $V(t) := \max_{x\in X} h(x, t)$, $t \in [0, 1]$ being X a set. Suppose that h is absolutely continuous with respect to t, for any x. Moreover, admit that there exists $b \in L^1([0, 1])$ such that $|\partial_t h(x, t)| \leq b(t)$ for any $x \in X$ and almost any $t \in [0, 1]$. Then V is absolutely continuous.*

Assuming further that h is differentiable in t, for any $x \in X$, and that for any $t \in [0, 1]$ the set $X(t) := argmax\ h(., t)$ is non-empty. Then, for any selection of $x^*(t) \in X(t)$ we have

$$V(t) = V(0) + \int_0^t \partial_t h(x^*(s), s) \, ds.$$

We are now in position to finish the proof of Theorem 4.1.

Step 4.6. There holds

$$W_1(f_t^1, f_0^1(p, q)dpdq \otimes \delta_{m_\infty}) \leq (4 + \kappa t)e^{-\varepsilon_0 \alpha_0 \langle p\lambda\rangle_0 t},$$

for any $t \geq 0$.

Proof. Step 4.5 ensures that for any $t \geq 0$ and any $(p,q) \in \mathrm{supp}\; f_0^1(p,q)dpdq$,

$$W_1\Big(\delta_{m(t,p,q)}, \delta_{m_\infty}\Big) = |m(t,p,q) - m_\infty| \leq (2 + \kappa t)e^{-\varepsilon_0\alpha_0\langle p\lambda\rangle_0 t},$$

while according to (34), we have that

$$W_1\Big(f_{t|(p,q)}^1, \delta_{m_\infty}\Big) \leq (4 + \kappa t)e^{-\varepsilon_0\alpha_0\langle p\lambda\rangle_0 t}.$$

In fact, we claim that

$$W_1\Big(f_t^1, \delta_{m_\infty} \otimes f_0^1(p,q)dpdq\Big) \leq (4 + \kappa t)e^{-\varepsilon_0\alpha_0\langle p\lambda\rangle_0 t}.$$

Let $\psi : K \to \mathbb{R}$ be an arbitrary 1-Lipschitz function. Then,

$$\int_K \psi\left(df_t^1 - \delta_{m_\infty} \otimes f_0^1(p,q)dpdq\right)$$
$$= \int_0^1 \int_0^1 \left(\int_{-1}^1 \psi(w,p,q)\left(df_{t|(p,q)} - \delta_{m_\infty}\right)\right) df_0(p,q).$$

The inner integral is bounded above by $W_1(f_{t|(p,q)}, \delta_{m_\infty})$ since $\psi(.,p,q)$ is 1-Lipschitz, which implies that

$$\int_K \psi\left(df_t^1 - \delta_{m_\infty} \otimes f_0^1(p,q)dpdq\right) \leq (4 + \kappa t)e^{-\varepsilon_0\alpha_0\langle p\lambda\rangle_0 t}\int_0^1 \int_0^1 df_0(p,q)$$
$$= (4 + \kappa t)e^{-\varepsilon_0\alpha_0\langle p\lambda\rangle_0 t}.$$

The claim follows by taking supremum among the functions ψ 1-Lipschitz. $\square$

5. **Numerical experiments.** In this section we perform some agent based simulations of the results studied along this work. We assess qualitatively the effect of the influence function

$$\lambda(w) = (w - 0.5)^2 + \varepsilon, \quad \varepsilon = 0.01,$$

on the dynamics. We first consider a homogeneous population and ratify that, indeed, the conclusions of Theorem 3.1 are true. Then, to illustrate the conclusion of Theorem 4.1, we add to this scenario several stubborn agents positioned at two specific values for ω. We finally analyze the evolution of a completely heterogeneous population with different values of q.

In all of the simulations presented here, we consider a population of $N = 1000$ agents. The opinion of the non-stubborn agents are initially uniformly distributed in $[-1,1]$. At every time slot, each one of the N agents interacts with a randomly selected agent and then, updates its opinion following the interaction rule (3) with $\gamma = 0.01$.

5.1. **Numerical experiments depicting the opinion attractiveness in a homogeneous population.** Since opinions are initially distributed uniformly in $[-1,1]$, i.e. $f_0 = \frac{1}{2}1_{[-1,1]}$, the theoretical value m_∞ of the consensus(given by (21) in Theorem 3.1), satisfies

$$\Lambda(m_\infty) = \frac{1}{2}\int_{-1}^1 \Lambda(w)\, dw,$$

where Λ is an antiderivative of $\lambda(w) = (w - 0.5)^2 + \varepsilon$, $\varepsilon = 0.01$. A numerical resolution of this nonlinear equation gives

$$m_\infty \approx -0.35.$$

In Figure 1 (left) we show the time evolution of the opinions of 10 agents of the population (blue curves). Consensus clearly occurs at the value m_∞, indicated by the horizontal red dashed line.

FIGURE 1. Evolution of the opinion of 10 agents (blue) from a homogeneous population of $N = 1000$ agents interacting according the interaction rule (16) studied in this paper (left) or the interaction rule (22) considered in [2, 38] (right). The red dashed line indicates the theoretical limit opinion in both cases ($m_\infty \approx -0.35$ for interaction rule (16), (left), and $\tilde{m}_\infty \approx 0.41$ for interaction rule (22), (right)) The early evolution is shown in inset.
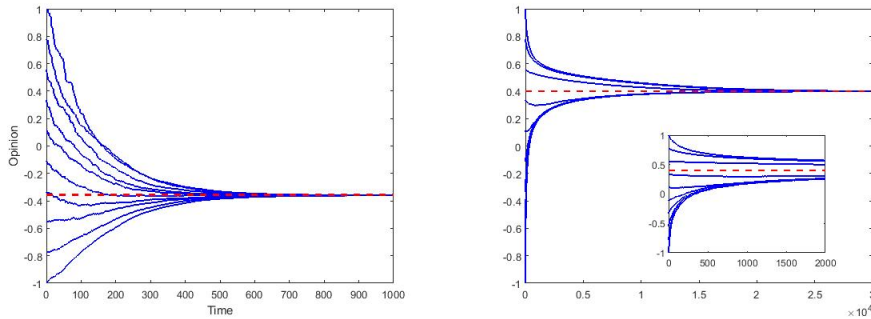
This is completely in contrast to consider that the influence is exerted, instead, by one's own opinion agent, see [2, 38]. Indeed, under the interaction rules (22), we prove in subsection §3.3 that if consensus is reached then, the consensus opinion $\tilde{m}_\infty$ must satisfy

$$\Pi(\tilde{m}_\infty) = \int_{-1}^{1} \Pi(w) \, df_0(w),$$

where Π is an antiderivative of $1/P$. Taking as f_0 the uniform distribution on $[-1, 1]$ and $P(w) = \lambda(w) = (w - 0.5)^2 + \varepsilon$, we numerically obtain

$$\tilde{m}_\infty \approx 0.41.$$

The result of the agent-based simulation shown in the right figure of Figure 1 confirms that consensus takes place at $\tilde{m}_\infty$ (indicated by the red dashed horizontal line).

We can also observe from the simulation that the consensus is attained much faster for the interaction rule (16) than for (22). Recall that the interaction rule (22) takes into account the attractiveness of the opinion of the agent one is interacting with. Intuitively, this is due to the fact that agents with opinion w such that $P(w) \approx 0$ are almost stubborn agents: they change opinion very slowly. This is clear from the figures in Figure 2, where we plot for the two interaction rules (16) (left) and (22) (right) the logarithm of the length of the convex hull of supp f_t, the support of the distribution f_t of opinion at time t. Namely, we depict

$$\ln \left(\max_{\text{supp} f_t} w - \min_{\text{supp} f_t} w \right).$$

In the rest of this section we examine the impact of stubborn agents combined with the influence function.

FIGURE 2. Evolution of $\ln(\max w - \min w)$, where the max and min are taken on the support of the distribution of opinions, for interaction rule (16) studied in this paper (left), and interaction rule (22) considered in [2, 38] (right).
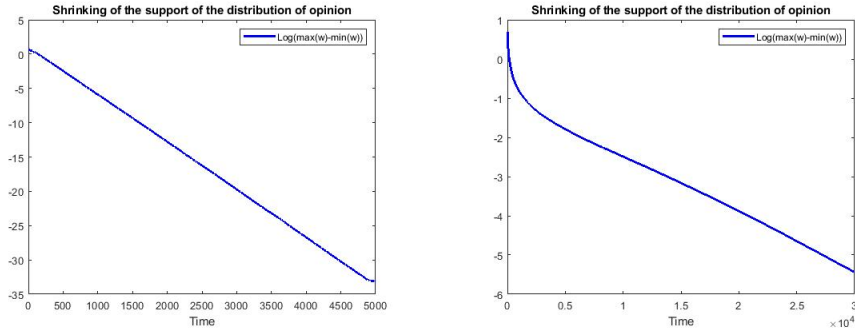
We consider a population of $N = 1000$ agents with a proportion α_0 of stubborn agents. To balance the effect between stubbornness and the attractiveness of the other's opinions, every agent has $p = 1$. As before the influence function is $\lambda(w) = (w - 0.5)^2 + \varepsilon$, $\varepsilon = 0.01$, and $\gamma = 0.01$. Assume that half of the stubborn agents have opinion $w = 1/4$ and the other half opinion $w = 3/4$, so that

$$f_0^0 = \frac{1}{2}\delta_{p=1} \otimes \delta_{w=\frac{1}{4}} + \frac{1}{2}\delta_{p=1} \otimes \delta_{w=\frac{3}{4}}.$$

Notice that, in particular

$$\langle p\lambda \rangle_0 = \int p\lambda(w)\, df_0^0(w, p) = \int_{-1}^1 \lambda(w)\, df_0^0(w) = \frac{1}{2}(\lambda(1/4) + \lambda(3/4)) = \frac{1}{16} + \varepsilon.$$

and

$$\langle p\lambda(w)w \rangle_0 = \int_{-1}^1 \lambda(w)w\, df_0(w) = \frac{1}{2}(\lambda(1/4)\frac{1}{4} + \lambda(3/4)\frac{3}{4}) = \frac{1}{2}(\frac{1}{16} + \varepsilon).$$

The theoretical limit opinion is then

$$m_\infty = \frac{\langle p\lambda(w)w \rangle_0}{\langle p\lambda \rangle_0} = \frac{1}{2}.$$

To evaluate the impact driven just by the proportion α_0 of stubborn agents on the dynamics, we suppose that all of the non-stubborn agents have $q = 1$. We show in Figure 3 the time evolution of the opinion of 10 agents belonging to the non-stubborn population (blue curves). We consider a proportion of stubborn agents $\alpha_0 = 2\%$ (left) and $\alpha_0 = 60\%$ (right). The theoretical limit opinion m_∞, is depicted on red dashed line.

We observe a perfect compliance between the agent-based simulations and the theoretical prediction. Furthermore, in the simulations the consensus is clearly achieved in two steps. First non-stubborn agents quickly reach a consensus, and then all together move slowly towards the final limit opinion m_∞.

This fact is specially well observed on the left figure, where a smaller proportion of stubborn population is considered ($\alpha_0 = 2\%$). At first the impact of the stubborn agents is almost negligible, so that the opinions of the non-stubborn population evolve first at a value close to the predicted consensus opinion in absence of stubborn

FIGURE 3. Evolution of the opinion of 10 agents (blue) from a population of $N = 1000$ agents with $\alpha_0 = 2\%$ (left) and $\alpha_0 = 60\%$ (right) stubborn agents. The red dashed line indicates the theoretical limit opinion $m_\infty = 1/2$, and the blue dotted lines the opinion of the stubborn agents (half with opinion $1/4$ and the other half with opinion $3/4$). The early evolution is shown in inset.
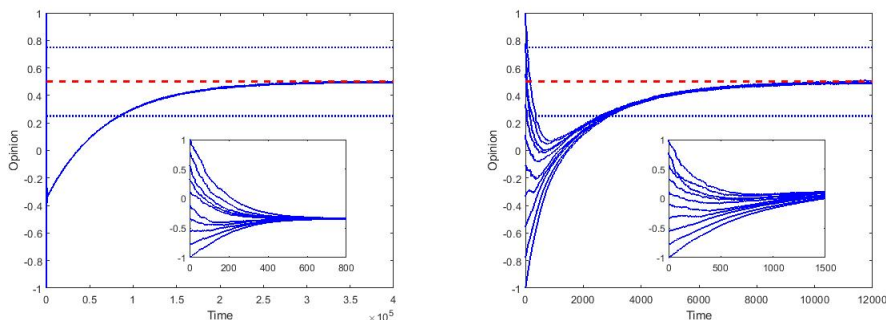
agents, namely here -0.355. In other words, the influence function λ drives the dynamics at early stages, and then is wiped out by the stubborn agents. Comparing both figures, we also notice that a high proportion of stubborn agents accelerates the convergence towards the consensus, in accordance with estimation (29).

Finally, we wish to appreciate the qualitative impact of the parameter q on the dynamics. We now consider that the values of q for the non-stubborn population are distributed uniformly in $(0.2; 1)$. The rest of the parameters are kept as before with $\alpha_0 = 0.6$. In Table 4 we show several snapshots at different times of the distribution of (w, q) in the non-stubborn population (w in the horizontal axis, q in the vertical axis). We can clearly appreciate that agents with a high q, i.e. the most volatile agents, are indeed changing opinion quicker than the rest.

Acknowledgments. The authors appreciated the fine comments of both referees which helped to enhance the previous version of the manuscript.

REFERENCES

[1] G. Aletti, A. K. Naimzada and G. Naldi., Mathematics and physics applications in sociodynamics simulation: The case of opinion formation and diffusion, *Mathematical Modeling of Collective Behavior in Socio-Economic and Life Sciences*, Birkhäuser Boston, (2010), 203–221.

[2] G. Aletti, G. Naldi and G. Toscani, First-order continuous models of opinion formation, *SIAM J. Appl. Math.*, **67** (2007), 837–853.

[3] S. E. Asch, Opinions and social pressure, *Scientific American*, **193** (1955), 31–35.

[4] R. B. Ash, *Real Analysis and Probability*, Probability and Mathematical Statistics, Academic Press, New York-London, 1972.

[5] P. Balenzuela, J. P. Pinasco and V. Semeshenko, The undecided have the key: Interaction driven opinion dynamics in a three state model, *PLoS ONE*, **10** (2016), e0139572, 1–21.

[6] B. O. Baumgaertner, R. C. Tyson and S. M. Krone, Opinion strength influences the spatial dynamics of opinion formation, *The Journal of Mathematical Sociology*, **40** (2016), 207–218.

[7] B. O. Baumgaertner, P. A. Fetros, R. C. Tyson and S. M. Krone, Spatial Opinion Dynamics and the Efects of Two Types of Mixing, *Phys Rev E.*, **98** (2018), 022310.

[8] N. Bellomo, *Modeling Complex Living Systems A Kinetic Theory and Stochastic Game Approach*, Birkhauser, 2008.

FIGURE 4. Evolution of the distribution of (w, q) among the non-stubborn population during one simulation (w in the horizontal axis and q in the vertical axis) with $\lambda(w) = (w - 0.5)^2 + \varepsilon$, $\varepsilon = 0.01$. From left to right and top to bottom, figures show the distribution of (w, q) at time $1, 500, 1000, 1500, 3000, 5000, 10000, 15000, 30000$.
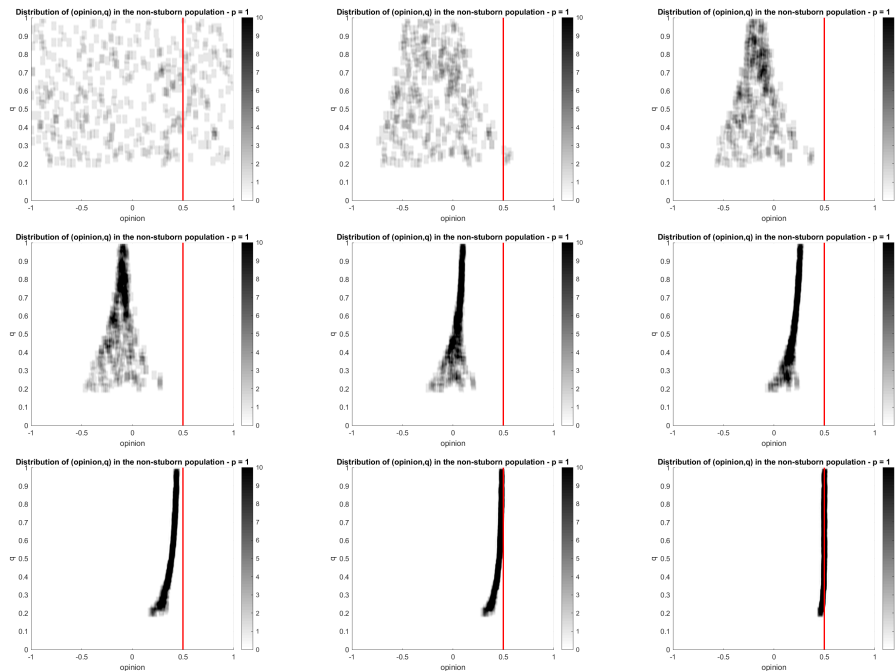
[9] N. Bellomo, G. Ajmone Marsan and A. Tosin, *Complex Systems and Society. Modeling and Simulation*, SpringerBriefs in Mathematics, 2013.

[10] K. C. Border and C. D. Aliprantis, *Infinite Dimensional Analysis - A Hitchhiker's Guide*, 3rd Edition, Springer, 2006.

[11] E. Burnstein and A. Vinokur, What a person thinks upon learning he has chosen differently from others: Nice evidence for the persuasive-arguments explanation of choice shifts, *Journal of Experimental Social Psychology*, **11** (1975), 412–426.

[12] J. A. Cañizo, J. A. Carrillo and J. Rosado, A well-posedness theory in measures for some kinetic models of collective motion, *Mathematical Models and Methods in Applied Sciences*, **21** (2011), 515–539.

[13] R. B. Cialdini and M. R. Trost, Social influence: Social norms, conformity and compliance, *The Handbook of Social Psychology*, McGraw-Hill, (1998), 151–192.

[14] F. Cucker and S. Smale, Emergent behavior in flocks, *IEEE Transactions on Automatic Control*, **52** (2007), 852–862.

[15] G. Deffuant, D. Neau, F. Amblard and G. Weisbuch, Mixing beliefs among interacting agents, *Advances in Complex Systems*, **3** (2000), 87–98.

[16] P. Embrechts and M. Hofert, A note on generalized inverse, *Mathematical Methods of Operations Research*, **77** (2013), 423–432.

[17] N. E. Friedkin and E. C. Johnsen, Social influence and opinions, *Journal of Mathematical Sociology*, **15** (1990), 193–206.

[18] T. Fujimoto, A simple model of consensus formation, *Okayama Economic Review*, **31** (1999), 95–100.

[19] S. Galam, *Sociophysics: A Physicist's Modeling of Psycho-Political Phenomena,* Springer Science & Business Media, 2012.

[20] R. Hegselmann and U. Krause, Opinion dynamics and bounded confidence: Models, analysis and simulation, *Journal of Artificial Societies and Social Simulation*, **5** (2002).

[21] R. A. Holley and T. M. Liggett, Ergodic theorems for weakly interacting infinite systems and the voter model, *The Annals of Probability*, **3** (1975), 643–663.

[22] C. La Rocca, L. A. Braunstein and F. Vázquez, The influence of persuasion in opinion formation and polarization, *Europhys. Letters*, **106** (2014), 40004.

[23] B. Latané, The psychology of social impact, *American Psychologist*, **36** (1981), 343–356.

[24] J. Lorenz, Continuous opinion dynamics under bounded confidence: A survey, *International Journal of Modern Physics C*, **18** (2007), 1819–1838.

[25] M. Mäs and A. Flache, Differentiation without distancing. Explaining bi-polarization of opinions without negative influence, *PloS one*, **8** (2013), e74516.

[26] P. Milgrom and I. Segal, Envelope theorems for arbitrary choice sets, *Econometrica*, **70** (2002), 583–601.

[27] R. Ochrombel, Simulation of Sznajd sociophysics model with convincing single opinions, *International Journal of Modern Physics C*, **12** (2001), 1091–1092.

[28] L. Pareschi and G. Toscani, *Interacting Multiagent Systems: Kinetic Equations and Monte Carlo Methods*, Oxford University Press, Oxford, 2014.

[29] L. Pedraza, J. P. Pinasco and N. Saintier, Measure-valued opinion dynamics, *M3AS: Mathematical Models and Methods in Applied Sciences*, **30** (2020), 225–260.

[30] M. Pérez-Llanos, J. P. Pinasco and N. Saintier Opinion attractiveness and its effect in opinion formation models, *Phys. A*, **559** (2020), 125017, 9 pp.

[31] M. Přez-Llanos, J. P. Pinasco, N. Saintier and A.Silva, Opinion formation models with heterogeneous persuasion and zealotry, *SIAM Journal on Mathematical Analysis*, **50** (2018), 4812–4837.

[32] F. Vazquez, N. Saintier and J. P. Pinasco, The role of voting intention in public opinion polarization, *Phys. Rev. E*, **101** (2020), 012101, 13pp.

[33] N. Saintier, J. P. Pinasco and F. Vazquez, A model for a phase transition between political mono-polarization and bi-polarization, *Chaos: An Interdisciplinary Journal of Nonlinear Science*, **30** (2020), 063146, 17 pp.

[34] J. P. Pinasco, V. Semeshenko and P. Balenzuela, Modeling opinion dynamics: Theoretical analysis and continuous approximation, *Chaos, Solitons & Fractals*, **98** (2017), 210–215.

[35] F. Slanina and H. Lavicka, Analytical results for the Sznajd model of opinion formation, *The European Physical Journal B*, **35** (2003) 279–288.

[36] D. W. Stroock, *Probability Theory, An Analytic View*, Cambridge University Press, 1993.

[37] K. Sznajd-Weron and J. Sznajd, Opinion evolution in closed community, *International Journal of Modern Physics - C*, **11** (2000), 1157–1165.

[38] G. Toscani, Kinetic models of opinion formation, *Communications in Mathematical Sciences*, **4** (2006), 481–496.

[39] C. Villani, Topics in optimal transportation, *Grad.Studies in Math.*, American Mathematical Soc., (2003).

E-mail address: mpperez@us.es

E-mail address: jpinasco@gmail.com

E-mail address: nsaintie@dm.uba.ar