Social Influence, Binary Decisions and Collective Dynamics*

Dunia López-Pintado

Institute for Social and Economic Research and Policy, Columbia University.

Duncan J. Watts

Columbia University and Santa Fe Institute

March 26, 2006

Adress for correspondence:

Duncan Watts

Department of Sociology

Columbia University

815 International Affairs Building,

420 West 118 Street.

New York, NY, 10027, USA.

Email: djw24@columbia.edu.

Phone:1-212-854-4343

FAX: 1-212-854-2963

*We would like to thank Peter Dodds, Steven Durlauf and Peyton Young, as well as the members of the Collective Dynamics Group at Columbia University for valuable comments and suggestions.

Abstract

In this paper we address the general question of how social influence determines collective outcomes for large populations of individuals faced with binary decisions. First, we define conditions under which the behavior of individuals making binary decisions can be described in terms of what we call an influence-response function: a one-dimensional function of the (weighted) number of individuals choosing each of the alternatives. And second, we demonstrate that, under the assumptions of global and anonymous interactions, general knowledge of the influence-response functions is sufficient to compute equilibrium (and even non-equilibrium) properties of the collective dynamics. In general, we find that collectives making very different kinds of decisions can exhibit surprisingly similar behavior; and conversely, that very similar types of decisions can yield collective behavior that is dramatically different.

Keywords: social influence, influence-response function, collective outcomes, best-response dynamics.

JEL Classification Numbers: C73, O31, O33, L14.

1 Introduction

The role of social influence in individual and collective decision making is a pervasive but still poorly understood feature of the social world. Although individuals often perceive themselves to be acting independently, regarding their subsequent actions and choices as accurate reflections of stable, intrinsic preferences, sociologists, psychologists, and economists, have come to accept that many decisions made by individuals depend critically on their observations, or anticipations, of similar decisions made by others. In this paper, we refer to the resulting interdependence between individual decision makers as "social influence." We emphasize, however, that social influence is not a singular phenomenon, but rather exhibits multiple social, psychological, and economic origins (Cialdini 2001). Individuals, for example, may be susceptible to social influence out of a desire to identify with certain social groups (Festinger, Schachter and Back, 1950) or to differentiate oneself from them (Simmel, 1957), in order to avoid sanctions over non-conformity (Asch, 1953); as a socially conditioned response to authority (Milgram, 1969), as a means of reducing the complexity of the decision making process (Gigerenzer, Todd and Group, 1999), as a way of inferring otherwise inaccessible information about the world (Bikhchandani, Hirshleifer and Welch, 1998; Goldstein and Gigerenzer, 2002), in order to gain access to a particular network (Katz and Shapiro, 1985), or to reap the benefits of coordinated action (Oliver, 1980; Oliver and Marwell, 1985).

Consistent with its multifaceted origins, social influence also manifests itself in a wide range of social phenomena, including local variability in crime rates (Edward L. Glaeser et al., 1996 and Dan M. Kahan, 1997) and economic conventions (H. Peyton Young and Mary A Burke, 2001), diffusion of innovations (Thomas W. Valente, 1995), "bystander inactivity" (Christina Bicchieri and Yoshitaka Fukui, 1999 and Robert B. Cialdini, 2001), residential segregation (Thomas C. Schelling, 1971), herd behavior in financial markets (Robert J. Shiller, 2000), the success and failure of social movements (Hyojoung Kim and Peter S. Bearman, 1997), political uprisings (Timur Kuran, 1991 and Susanne Lohmann, 1994), contributions to public goods (Elinor Ostrom et al., 1999), and other forms of collective action (David Strang and Sarah A. Soule, 1998). Furthermore, while many discussions of social influence tend to be concerned with non-market behavior (e.g. social movements, conformity to reference groups, and fashion), social influence can play an important role in markets as well. For example, a related body of work that has attracted the attention of economists deals with a class of technology markets that exhibit what have been called "network externalities" (Michael L. Katz and Carl Shapiro, 1985) or somewhat more generally "network effects" (Stanley J. Liebowitz and Stephen E. Margolis, 1998, 1994). Both terms are meant to imply that the utility to an individual of a particular product (e.g. a fax machine) or skill (e.g. a language) is positively related the size of the relevant "network" associated with the product/skill; thus it is effectively a function of the previous decisions of others.

The pervasive and multifarious nature of social influence makes it a topic of great relevance to much of social science; however, it also renders the concept somewhat difficult to describe precisely. In order to gain some analytical traction on a potentially diffuse problem, we therefore restrict our discussion of social influence to the class of binary decisions (i.e. choices between precisely two discrete alternatives) that exhibit what we will call "decision externalities", by which we mean simply that the likelihood of each individual choosing one alternative over another is a function of the number of others choosing each alternative.[1] At the cost of some precision, previous models of binary decisions with externalities can be divided into three broad categories to which we refer as heuristic models, mechanistic models, and social utility models respectively.[2]

Heuristic models, which dominate the diffusion of innovations and collective action literature (Roy M. Anderson and Robert M. May, 1991; Edward L. Glaeser et al., 1996; M. S. Granovetter, 1978, and Duncan J. Watts, 2002), begin from the presumption that individual actors "adopt" some practice (which can be variously an item, a procedure, or an idea) as a consequence of being exposed to it by other actors (who can be individuals or organizations). Although these models can differ considerably in their specific assumptions about how exposure leads to adoption, all of them either explicitly or implicitly assume the logic of contagion; that is, that "susceptible" individuals can be "infected" purely by exposure to existing "infectives". Heuristic models have the great advantage that once the infection rule is specified, the exercise of computing the equilibrium, or even non-equilibrium, behavior of the collective dynamics is relatively straightforward[3] (P. S. Dodds and D. J. Watts, 2005, 2004; M. S. Granovetter, 1978; M. S. Granovetter and R. Soong, 1986; Mark. S. Granovetter and R. Soong, 1983; D. López-Pintado 2004, 2005, and D. J. Watts, 2002). Unfortunately, heuristic models are usually proposed in the absence of any precise psychological or economic rationalization; thus many specific choices of heuristics may seem equally plausible, or implausible. These ambiguities, furthermore, turn out to be important (P. S. Dodds

[1] Our use of the term "externality" is therefore roughly analogous to its standard usage, where it refers to an external effect generated by an individual's action (e.g. downstream pollutions), but is specific to the context of social influence. By contrast our use of the term is somewhat more general than Schelling (1973) who consideres only one particular class of what we will call "explicit" externalities.

[2] Young (2005) also distinguishes between different types of social influence models, namely, contagion models and learning models. He studies how the overall shape of the adoption curve depends on the particular mechanism that describes why individuals adopt, whereas we are more interested in simply describing the long-run behavior of the dynamics.

[3] In this sense, heuristic models of social influence are related to a vast literature in mathematical epidemiology, which is concerned with onset, size, duration, and control of epidemics of infectious disease (Roy M. Anderson and Robert M. May, 1991; N.M. Ferguson et al., 2003, among others).

and D. J. Watts, 2004 and D. López- Pintado, 2005); thus which model is appropriate to what specific applications would seem to be an important question, but not one that can be resolved without careful attention to the micro-mechanical details.

By contrast with heuristic models, mechanistic models have the virtue that the externalities in question arise directly from assumptions about the psychological or economic details of the decision making process itself; hence are grounded theoretically in the application used to motivate them. Parameters are therefore interpretable and policy implications are, at least in principle, clear. For example, Bikhchandani et al. (1998, 1992) propose a sequential decision making model in which individuals update their own private signals regarding a choice between two alternatives by observing the choices (but not the private signals) of previous decision makers. The result is that under certain conditions, individuals will ignore their own private signals and follow the majority, triggering what Bikhchandani et al. (1992) label an "information cascade".[4] An alternative strand of mechanistic models, following Schelling (1973), focuses on the decision externalities inherent to particular classes of N-player games. A number of authors have subsequently investigated the relationship between individual best responses and the dynamics of collective decisions in the context of social dilemmas (N. S. Glance and B. A. Huberman, 1993, Geoffrey Heal and Howard Kunreuther, 2003, Pamela E. Oliver, 1980, 2001), anti-coordination games (Bramoullé, 2001), and coordination games (G. Ellison, 1993; M. Kandori et al., 1993; S. Morris, 2000 and P. Young, 1993, 1998). Finally, Milgrom and Roberts (1990) analyze the properties of the Nash equilibrium in very general games. In particular, they study games with strategic complementarity, which broadly corresponds with the positive externality games in our setting.[5]

The downside of mechanistic models is that they are difficult to generalize to examples other than those from which their micro-mechanical description is derived. For example, while it is clear that in each of the above examples, individual decisions are subject to externalities, it is unclear how the information cascades discussed by Bikhchandani et al. (1992) are relevant to problems involving public goods. Presumably some kind of cascadelike dynamics are possible under some conditions (Geoffrey Heal and Howard Kunreuther, 2003), regardless of whether the externality in question derives from the assumption that others have information that one lacks, or from the decisions makers utility being contingent on the contributions of others. But short of a unifying framework under which both kinds of externalities can be subsumed, it is impossible to say what those conditions might be, or

[4] Banerjee (1992) has independently proposed a similar model, with similar results, and Arthur and Lane (1993) have extended the approach to decisions in which risk-averse agents observe private signals of predecessors as well as their actions.

[5] We extend Milgrom and Roberts (1990) since apart from strategic complentarity games we also consider sbustitute gams and combinations of both types of games. In other aspects, however, Milgrom and Roberts (1990) is more general because it allows for a larger strategy space.

how lessons learned in one context might be applied in another.

One possible compromise between the generality of heuristic models and the economic rationality of mechanistic models is the class of “social utility” models, first introduced by Leibenstein (1950), and later applied specifically to the case of binary decisions by Brock and Durlauf (2001) and others.[6] Brock and Durlauf (2001) identify two specific functional forms that social utility might hypothetically take: “proportional spillovers”, according to which social utility increases linearly with the expected number of others; and “pure conformity”, according to which social utility diminishes quadratically away from the mean response. They then derive expressions governing the equilibrium states of the collective decision dynamics, corresponding to the proportional spillovers and pure conformity utility functions respectively. Other work in the same tradition also introduces one of these two functional forms for social utility in their respective utility functions with analogous consequences. Young and Burke (2001), for example, introduce a proportional spillover term, and Bernheim (1994), Bicchieri and Fukui (1999), and Blume and Durlauf (2003) all invoke terms that correspond to pure conformity.

Although less arbitrary than heuristic models and more general than mechanistic models, social utility models are in some ways an awkward compromise. It remains ultimately unclear, for example, on what basis individuals should derive utility from the choices or attitudes of others; thus one cannot say with confidence why one particular functional form for the “social” part of the utility function should be preferred over another, and which functions are appropriate to which applications. Furthermore, unlike heuristic models which at least in principle, could be estimated empirically, it is unclear how one would go about measuring social utility. And finally, social utility models, although formulated intuitively, still present considerable analytical difficulties with respect to computation of collective dynamics, especially for heterogeneous populations.

Acknowledging the difficulty that social utility models have attempted to overcome—that is, combining the economic rationality of mechanistic models with the generality of heuristic models—this paper adopts an approach to the problem that omits the explicit formulation of social utility altogether. We achieve this goal in two, analytically distinct stages. First, we define conditions under which the myopic-best response of individuals making binary decisions in the presence of social influence can be described in terms of what we call an influence-response function: a one-dimensional (i.e. scalar) function of the (weighted) number of others choosing each of the alternatives. As we will discuss below, our formulation of influence-response functions, although restricted in some important ways, encompasses a

[6] Brock and Durlauf's formulation of social utility corresponds roughly to one part of what Leibenstein calls non-functional utility, a category under which Leibenstein also includes "speculative utility" and "irrational utility". Here we follow Brock and Durlauf 's specification.

wide range of interesting applications. And second, we demonstrate that, under some additional conditions, reasonably general knowledge of individual influence-response functions is sufficient to compute the equilibrium, and even non-equilibrium, properties of the collective dynamics. Importantly, our method for computing collective dynamics—which is standard in the field of nonlinear dynamics, but which appears to be little known in economics[7]—is reasonably simple to perform in practice, and generalizes easily to allow for heterogeneous populations, synchronous and asynchronous updating, and stochastic as well as deterministic best-responses.

Our analytically distinct treatment of individual and collective behavior is central to our approach, and has a number of advantages. First, it permits us to treat in a unified way a number of applications, which at face value seem quite different. For example, we can compute the equilibrium behavior of a large population engaged in certain kinds of public goods games in the same way that we would for certain kinds of social learning problems, possibly even with the same equilibria, even though the two "games" arise from very different micro-origins. Second, and conversely, our approach also permits us to differentiate between applications, such as public goods games with slightly different shaped production functions, which are superficially very similar but which result in qualitatively distinct collective behavior. Finally, as we discuss briefly in Section 3, in cases where it is not possible to derive the relevant influence-response functions from first principles, one may nevertheless be able to measure them directly, perhaps through experiments. Because, in such cases, it would still be possible to compute collective equilibria, the separation of individual from collective behavior may have empirical advantages as well.

The remainder of this paper is organized as follows. In Section 2, we outline our analytical framework for characterizing binary decision externalities, and then derive the type of influence-response functions for a number of specific cases. In Section 3, we introduce some additional assumptions regarding how individuals in a population of decision makers interact with each other, and then show how under these conditions, our framework of influence-response functions leads to particularly simple methods for computing collective equilibria, outlining solutions both for discrete (synchronized) and continuous (asynchronized) updating of decisions. In Section 4, we outline some extensions to the basic model, including stochastic best responses; and in Section 5, we conclude.

2 Binary decisions and externalities

Our first goal is to outline a framework for relating the externalities that pertain to various decision making scenarios to individual-level decision rules that is both general, in the sense

[7] Schelling 1973 introduces the same idea, but does not develop it in detail.

that it encompasses all the particular examples discussed above, yet precise in the sense that the parameters which appear in the decision making rule can be interpreted in terms of the motivating application, and hence are amenable to measurement. In order to proceed, however, we require some analytical distinctions, both with respect to the origins of decision externalities, and also their form.

With regard to origins, we distinguish between what we call implicit and explicit decision externalities. *Explicit externalities* arise whenever the utility ascribed to one alternative over another is a direct function of the (absolute or relative) number of others choosing that alternative. *Implicit externalities*, however, arise indirectly as a result of inferences that individuals make about information regarding the decision that is held by others, and that they obtain through some sampling procedure. Both explicit and implicit externalities can arise in a variety of circumstances, and within each broad class there exist additional subclasses; however, all obey this basic distinction. Implicit and explicit externalities, notably cannot be derived from any common framework: although they may or may not have the same effective implications for decision making, they are simply of different origins.

With respect to form, we identify three kinds of externalities—strictly positive, strictly negative, and non-monotonic—where the first two kinds correspond loosely to Leibenstein's (1950) characterization of "bandwagon" and "snob" effects, and the third can be regarded as a mixture of the other two (M. S. Granovetter and R. Soong, 1986). In our scheme, positive externalities pertain whenever the probability of choosing a particular alternative increases with respect to the number of others choosing it, and negative externalities imply the opposite. In our usage, the presence of non-monotonic externalities implies that externalities are positive when the number of adopters is low, but that they become negative as the number of adopters increases (where clearly other combinations are possible). For example, in fashion, one may not wish to adopt a new style until it has been adopted by sufficiently many others (a positive externality), but one may also lose interest in the same style once it has been adopted by too many others (a negative externality).[8]

Depending on the details of the decision, all three kinds of externalities may arise either from explicit or from implicit origins, where the same form of externality may arise in very different scenarios, and by contrast different forms may arise from scenarios that superfi-

[8]The definitions of positive, negative and non-monotonic externalities introduced here, do not coincide with the common usage of these terms in Economics. In this paper the form of the externality depends on how the behavior of others affect ones *probability* of choosing an action Elsewhere, however, the distinction between positive, negative and non-monotonic externalities generally depends on whether the utility increases, decreases or is non-monotonic with respect to the behavior of others. To clarify this point, consider for instance a public goods game. If we rely on the standard definitions public goods always exhibit positive externalities. However, when the production function is concave, ones probability of contributing decreases as the number of contributors increase, which would imply negative externalities, using our nomenclature.

cially may appear hard to distinguish (for example, between two public goods with different shaped production functions). Furthermore, as we will show in Section 3, the distinction between monotonic and non-monotonic externalities is consequential, as the collective dynamics corresponding to each case will generally display qualitatively distinct properties.

Before proceeding to derive influence-response functions for binary decisions, we require one additional restriction on our model—that social influence satisfies what we call the assumption of “independent effects”. By this assumption, we mean that the influence an individual experiences as result of another individual ’s action is independent of the influence experienced from any other individual. To put it differently, independent effects assumes that one considers the action of other individuals as decisions taken simultaneously, and therefore independently of each other.[9] Although this assumption is quite general—it does not require, for example, that all influences be felt equally, or that the resulting influence-response function be linear—it does exclude some important cases. For example, in the model of Bikhchandani et al. (1992) individuals not only condition their choice on the previous choices of others, but also on the order in which those choices were made; thus the influence exerted by the second individual in the sequence on the third individual will be different depending on whether or not the second individual made the same decision as the first individual. No such conditioning is possible in our model, which is equivalent to saying that an individuals observations of others are unordered.[10] We do not believe, however, that this simplifying assumption overly restricts the applicability of our model, as for most of the situations in which we are interested (typically including large number of individuals), the knowledge and computational capacity required to condition influence on some exogenously defined ordering would seem to exceed even quite generous estimators of human rationality.

In what follows, we explain why, under this assumption, individuals’ behavioral rules (or best responses) can be described by simple one-dimensional functions of an aggregate "social signal". To formalize matters, let $N = \{1, ...n\}$ be a finite but large set of individuals and $A = \{0, 1\}$ be the common set of actions. That is, every individual makes a binary decision $a_i \in A$ (e.g. whether or not to adopt a certain behavior, purchase a product, participate in a riot, contribute to a public good, etc.). For every i, let $\widehat{a}_{-i} = (a_1, ..., a_{i-1}, a_{i+1}, ...a_n) \in A^{n-1}$ be the action profile representing the behavior of the remaining individuals in the population. We define a function, $R_i : A^{n-1} \to [0, 1]$ such that $R_i(\widehat{a}_{-i})$ is i's probability of choosing action 1, given $\widehat{a}_{-i}$. Notice that this function is contingent exclusively on the action profile

[9] This assumption is related to the phenomenon of *pluralistic ignorance,* studied in the field of social psychology; that is, even when individuals’ decisions are influenced by others they may ignore that the decisions of others are also subject to social influence.

[10] Other kinds of dependecies may be possible as well, and it may be possible to generalize our model to include some of them. However, for the sake of simplicity, we ignore all such dependencies here.

of others, therefore the order in which one samples is irrelevant and the actions of others are considered as simultaneous decisions. We note that if the best response is deterministic then $R_i(\widehat{a}_{-i}) \in \{0,1\}$; otherwise $R_i(\widehat{a}_{-i}) \in [0,1]$.

The next result shows how $R_i(\widehat{a}_{-i})$ can be expressed in terms of a one-dimensional function.

Proposition 1 *For every $i \in N$, given $R_i(\widehat{a}_{-i})$, there exists a vector of weights $\{w_{ij}\}_{j\in N}$, such that $w_{ij} \in \mathbb{R}_+$, $w_{ii} = 0$ and $\sum_{j\in N} w_{ij} \equiv n_i$, as well as a one-dimensional map r_i : $[0, n_i] \to [0,1]$, denoted as the influence-response function hereafter, such that*

$$R_i(\widehat{a}_{-i}) = r_i(k_i(\widehat{a}_{-i}))$$

where $k_i : \mathbb{A}^{n-1} \to [0, n_i]$, the social signal, is defined given the vector of weights $\{w_{ij}\}_{j\in N\setminus\{i\}}$ as follows:

$$k_i(\widehat{a}_{-i}) = \sum_{j\in N\setminus\{i\}} a_j w_{ij}.$$

As illustrated in Figure 1, to prove this proposition, we have to show that R_i can be expressed as a composition of the functions k_i and r_i. Obviously, it is enough to prove that there exists a vector of weights $\{w_{ij}\}_{j\in N\setminus\{i\}}$ such that k_i is an injective function (i.e., $k_i(\widehat{a}_{-i}) = k_i(\widehat{b}_{-i})$ if and only if $\widehat{a}_{-i} = \widehat{b}_{-i}$), since this would then allow us to define r_i as a function satisfying that $r_i(k_i(\widehat{a}_{-i})) = R_i(\widehat{a}_{-i})$. A constructive proof of the existence of these weights is presented in the appendix.

Insert Figure 1 about here

Although, in principle, the weights could simply be considered as abstract parameters used to construct a one-dimensional map, we concentrate here on situations where the weights are actually interpretable and reflect the relevance (or importance) that each interaction impinges on the decision under consideration. To be precise, w_{ij} represents the importance that i attributes to j's action. Hence, the influence structure determined by the weights can be interpreted as a weighted social network where, $N_i = \{j \in N \text{ s.t. } w_{ij} > 0\}$ is the set of individuals with whom i interacts (or whom he cares about). If $w_{ij} \in \{0,1\}$ for all $i, j \in N$, the social network is unweighted although directed. The network is undirected if in addition $w_{ij} = w_{ji}$ for all $i, j \in N$.[11]

In Section 4 we illustrate how stochastic influence-response functions might arise. Nevertheless, for concreteness, we focus here exclusively on the deterministic case. Notice that any deterministic influence-response function $r_i(k_i)$ is characterized by a finite number of thresholds, as illustrated in Figure 2. Formally, given $r_i(k_i)$, there exists $k_i^1, k_i^2, ...k_i^M \in [0, n_i]$ such

[11] Following the representation of a weighted graph, one could consider that some individuals have a negative impact on your decision. One way of achieving this would be to allow for negative weights . For simplicity, however, this extension is not included in this paper.

that for every $m \in \{1, ...M-1\}$, either $r_i(k_i) = 0$ for all $k_i \in [k_i^m, k_i^{m+1}]$ or $r_i(k_i) = 1$ for all $k_i \in [k_i^m, k_i^{m+1}]$.

Insert Figure 2 about here

One of the advantages of representing the influence-response function as a one-dimensional map is that we can study its monotonicity. In other words, we can analyze the effects of increasing what we refer as the social signal k_i. For instance, as illustrated in Figure 3.a, if $r_i(k_i)$ is increasing (positive externalities) then, given that it only takes the values 0 or 1, it can be characterized by an upward threshold $k_i^U \in [0, n_i]$ such that action 1 is chosen if and only if $k_i^U \leq k_i$.[12] Notice that, in the case of unweighted networks, positive externalities simply imply that the higher the number of individuals choosing action 1 in ones neighborhood, the higher the probability of choosing 1. And finally, in the extreme case where the network is complete and unweighted, positive externalities simply mean that the higher the *overall* number of adopters of action 1, the higher the probability of adopting 1. As illustrated in Figure 3.b, if the influence-response function is decreasing (negative externalities), it is characterized by a downward threshold $k_i^D \in [0, n_i]$ such that action 1 is chosen if and only if $k_i < k_i^D$.[13] [14] Finally, a non-monotonic influence-response function is typically characterized by multiple thresholds. For concreteness, in this paper we focus on the case with two thresholds; an upward threshold $k_i^U \in [0, n_i]$ and a downward threshold $k_i^D \in [0, n_i]$, such that $0 \leq k_i^U < k_i^D \leq n_i$, where action 1 is chosen if and only if $k_i^U \leq k_i < k_i^D$. This case is illustrated in Figure 3.c.

Insert Figure 3 about here

We emphasize that there may be many real world examples leading to these different forms of externalities. For example, if everybody expects nobody to applaud between the movements of a quartet, hardly anybody will, which makes this situation akin to the upward threshold influence-response function case. By contrast, when attending to a party, if you expect everybody to bring drinks you may think of bringing food. This also represents a critical

[12] This situation corresponds with a version of positive externalities where, if we compare the probability of choosing action 1 when a set of individuals, say S, are choosing 1, with the corresponding probability when there is a larger set of individuals choosing 1 (say S') which contains S (i.e. $S \subset S'$), then the latter probability is higher or equal than the former one.

[13] Notice that we are assuming that the influence-response function $r_i(k_i)$ evalutated at the threshold always takes the value of its right lateral limit. For example, $r_i(k_i^D) = \lim_{k_i \to k_{i+}^D} r_i(k_i)$.

[14] This specification may also be interpreted as a general definition for negative externalities that coincides with the standard one if the network is complete and unweighted. Specifically, if we compare the probability of choosing action 1 when a set of individuals, say S, are choosing 1, with the corresponding probability when there is a larger set of individuals choosing 1 (say S') which contains S (i.e. $S \subset S'$), then the latter probability is lower or equal than the former one.

mass phenomenon but with the opposite effect, that is, where there is a tendency to anti-coordinate or, in other words, to coordinate on different actions. Finally, upward-downward influence-response functions are also possible. In fashion, for example, people are often unwilling to purchase a product until some number of others have done so, but the same product can be once more unappealing when it becomes too popular.

As discussed above, influence-response functions, in addition to exhibiting multiple functional forms, can also arise in many circumstances and for different reasons. In the following sections we propose two frameworks to account for what we refer to as explicit and implicit externalities; that is, two fundamentally different origins for social influence that nevertheless can lead to very similar behavior. In order to formally explain the difference between explicit and implicit externalities, we must pay careful attention to the micro-mechanical details of the adoption process. To do so, we follow a mechanistic (or social utility) approach and incorporate the costs and benefits of choosing one action or the other in the analysis. Therefore the payoffs (or utilities) of individuals become part of the model. Although, this may be unnecessary for obtaining predictions about the collective outcomes, one clear advantage of tracing incentives to primitives related with costs and benefits is that it enables us to ask questions of whether or not we obtain efficient outcomes in equilibrium (in the sense of maximizing the total benefit to society) or, if instead, there is some tension between individual and collective incentives. Furthermore, the parameters of the models become interpretable and policy implications are, in principle, understandable.

2.1 Explicit externalities

Explicit externalities arise when the utility assigned to an action depends explicitly on the absolute or relative number of individuals choosing the action. For example, the benefits of using a particular computer language depend on the size of the population using the same language. In this example, there are at least two reasons for such effect. First, due to compatibility issues, the more people adopting the product, the easier it becomes to exchange information with them. Second, services and accessories are more readily and cheaply available the larger the population of users, where this second effect also applies to less obvious examples of network externalities such as choosing among brands of cars, digital cameras, etc.

To formalize matters, assume that each individual $i \in N$ is characterized by a utility function. Invoking the same argument used for the construction of the one-dimensional influence-response function, the utility of an individual can be represented as a function of the social signal k_i. That is, every $i \in N$ is characterized by a utility function $u_i : A \times \mathbb{R}_+ \to \mathbb{R}_+$ such that $u_i(a_i, k_i) \in \mathbb{R}_+$, where $k_i = \sum_{j \in N \setminus \{i\}} w_{ij} a_j$.

Let $\bigtriangledown u_i(\cdot, k_i)$ represent the difference in utility between choosing action 1 or action 0 given a particular value of k_i. More precisely, $\bigtriangledown u_i(\cdot, k_i) = u_i(1, k_i) - u_i(0, k_i)$. Assuming that individuals are utility maximizers, i will choose action 1 if $\bigtriangledown u_i(\cdot, k_i) \geq 0$, and action 0 otherwise, (where for simplicity we consider that if $\bigtriangledown u_i(\cdot, k_i) = 0$, action 1 is chosen). If $\bigtriangledown u_i(\cdot, k_i)$ is increasing with respect to k_i (u_i satisfies increasing differences) then i is subject to positive externalities, which in turn implies that her influence-response function is characterized by an upward threshold. Similarly, if $\bigtriangledown u_i(\cdot, k_i)$ is decreasing with respect to k_i (u_i satisfies decreasing differences) the externalities are negative, and thus, i's influence-response function is characterized by a downward threshold.[15] Finally, influence-response functions with multiple thresholds are only possible if the utilities have non-monotonic differences, i.e. $\bigtriangledown u_i(\cdot, k_i)$ is non-monotonic (or similarly, the externalities are non-monotonic). Note that, two individuals with the same thresholds do no necessarily need to have the same utilities, whereas obviously two individuals with the same utilities always have the same thresholds. Furthermore, the population will typically be characterized by a distribution of thresholds and, as shown later in the paper (Section 3), this information is sufficient to compute the equilibrium, and even non-equilibrium, properties of the collective dynamics. In contrast, if we want to evaluate the efficiency of the outcomes, information about utilities is also needed.[16]

To summarize, one can determine the form of the externality simply by analyzing the monotonicity with respect to the social signal of the difference in utilities of choosing action 1 and action 0. We next illustrate how this analysis can be performed in two different scenarios: technology adoption and public goods.

2.1.1 Technology adoption

Consider the model of technological adoption proposed by Katz and Shapiro (1985).[17] Consider two technologies $a = 0$ and $a = 1$, where the "intrinsic" utility of adopting technology a for individual i is given by $b_a(i)$. Letting p_a denote the price of purchasing technology a, then the overall utility of individual i is equal to

$$u_i(a, k_i) = b_a(i) - p_a + \nu_a(k_i)$$

[15] Obviously, individuals might have degenerate thresholds that would imply the precence of a dominant strategy.

[16] Schelling (1973) has also emphasized this point.

[17] The objective of Katz and Shapiro (1985) is to model the strategic behavior of two firms producing competing network goods by the endogenous decision of the price. In this example we consider their basic framework but we address a different question; we consider the prices of each product as given and study the form of the externality that arises.

where $\nu_a(k_i)$ represents the network effects for $a \in \{0,1\}$ and

$$k_i = \sum_{j \in N \setminus \{i\}} w_{ij} a_j$$

is the social signal. Obviously, the standard assumptions in the context of technology adoption are that $\nu_1(k_i)$ increases with respect to k_i, whereas $\nu_0(k_i)$ decreases with respect to k_i; hence the difference in utility between choosing 1 or 0 is

$$\nabla u_i(\cdot, k_i) = b_1(i) - p_1 + \nu_1(k_i) - b_0(i) + p_0 - \nu_0(k_i)$$

from which it follows that

$$\frac{d\nabla u_i(\cdot, k_i)}{dk_i} = \frac{d\nu_1(k_i)}{dk_i} - \frac{d\nu_0(k_i)}{dk_i} > 0.$$

As a consequence, the externalities are positive and thus individuals' influence-response functions are characterized by an upward threshold that depends on the parameters of the model as well as the specific forms of the functions $\nu_a(k_i)$, where $a \in \{0,1\}$. Notice that one can consider other markets (apart from technology) such as clothing or art where non-monotonic externalities may arise. These examples would correspond to non-monotonic functions $\nu_1(k_i)$ and $\nu_0(k_i)$. We note also that this model encompasses as a particular case (when $\nu_a(k_i)$ is a linear function of k_i) the standard social interaction models where individuals play a bilateral game with each neighbor. If the game is a coordination game the externalities are positive whereas if the game is an anti-coordination game the externalities are negative. Finally, we observe that if $\nu_1(k_i)$ and $\nu_0(k_i)$ are determined from primitive assumptions of the economic and psychological decision process, our model would therefore fit into the category of mechanistic models, whereas if these functions are determined as simple heuristics of how the behavior of others affects ones utility, our model would follow a social utility approach.

2.1.2 Public good games

Public good games represent another class of games from which influence-response functions can be simply derived.[18] Public goods appear in many different contexts such as health insurance, public transportation, environmental issues, and innovation, among others. In all these cases, each individual has to decide whether to invest ($a_i = 1$) or not, ($a_i = 0$) in some public good where the cost of investing is $c_i > 0$. Let $G_i(a_i, \widehat{a}_{-i})$ be the production function from the perspective of player i, meaning that $G_i(a_i, \widehat{a}_{-i})$ determines the benefits

[18] A very similar example has been described in Peterhansl and Watts (2005). Also Bramoulle and Kranton (2005) have recently studied a model of public good games in a social network framework.

obtained by player i given his own action a_i and the action taken by the remaining players, $\widehat{a}_{-i}$. For simplicity assume that the social network is unweighted; that is, $w_{ij} \in \{0,1\}$ for all $i, j \in N$, and therefore n_i denotes the size of i's neighborhood. It is straightforward to show that the production function can be expressed in terms of a function (also referred as the production function for i) $g_i : [0, n_i] \to R_+$ satisfying

$$G_i(a_i, a_{-i}) = g_i(k_i + a_i)$$

where

$$k_i = \sum_{j \in N \setminus \{i\}} w_{ij} a_j$$

is the number of individuals that have invested in i's neighborhood.
Then, for any player $i \in N$, her utility or payoff is

$$u_i(a_i, k_i) = g_i(k_i + a_i) - c_i a_i,$$

and player i will choose to invest if and only if

$$\nabla u_i(\cdot, k_i) = g_i(k_i + 1) - g_i(k_i) - c_i \geq 0.$$

It is always the case in public good games that an individual's utility increases with the number of contributions of others; that is, the production function is always increasing (i.e. $g_i'(k_i) \geq 0$). In this case, however, it is the monotonicity of g_i' what determines the form of the externality. Thus only if g_i is convex (positive externalities) will the influence-response function be an upward threshold function; and only if g_i is concave (negative externalities) will it be a downward threshold function. Finally, if g_i is convex for low values of k_i and concave for high values of k_i (i.e. a sigmoid function) the influence-response function will exhibit two thresholds, an upward and a downward threshold.[19]

Knowing the shape of the production function, which in principle can depend on the particular context, is therefore relevant since the predictions for the collective dynamics will critically depend on this feature. In many examples of public goods proposed in the literature the production function is concave, corresponding to the classical free-riding phenomenon in which incentives to contribute decrease with the number of contributors. There is also relevant work, however, that deals with non-concave production functions. For example, Winter (2005) studies a situation where individuals have the option of reducing the probabilities of failure of a joint project by investing towards their decisions. In such a scenario, the production function is convex and thus the incentives to contribute increase with the number of

[19] Notice that one of the two, or both, thresholds might be a degenerate threshold which would imply that a non-monotonic production function may lead to an upward threshold, a downward threshold, or no threshold at all (i.e. where one action is always prefered to the other).

contributors. Note that public good games constitute a class of games from which all three forms of externalities may easily arise: positive, negative and non-monotonic. As a consequence public good games which are superficially very similar (i.e. with slightly different shaped production functions) can result in qualitatively distinct collective behavior. Conversely, applications which at face value seem quite different, such as technology adoption and some kinds of public good games, can lead to similar qualitative aggregate outcomes. In the next section we extend our framework to include implicit externalities, showing that even "games" derived from entirely distinct origins, can be (in some circumstances) treated in a unified way.

2.2 Implicit externalities

Implicit externalities arise indirectly as a result of inferences that individuals make about the information regarding the decision that is held by others. Unlike explicitly externalities, we assume that the utility of choosing one action is independent of the number of individuals choosing the same action. Nevertheless, since individuals are uncertain about the utility corresponding to each action, the information about the actions taken by other individuals in the population is used to reduce this uncertainty and infer the optimal choice.

For example, the number of people in a restaurant may be taken as a signal of the quality of the food. Also, information regarding which books (or CDs) are best-sellers might constitute an important influence on purchase decisions precisely because of this signaling effect. Using the behavior of others as a screening device is enhanced when the market is complex and sufficient knowledge is needed to make an optimal judgement, such as in the case of unfamiliar products with multiple brands. Notice that we assume that individuals are uncertain about the benefits corresponding to the purchase of the goods, which implies that these are goods whose characteristics can only be well known after use.

There are many ways one can describe a model with uncertainty; here we take a very simple but standard approach to the problem. For concreteness, assume that there are only two states of the world relevant for the decision (e.g., high and low quality of a new technology) and individuals are uncertain about which is the real state. An individual's utility depends on his action (adopt or not the new technology) and the state of the world. Formally, for all $i \in N$, let $A = \{0, 1\}$ be the set of possible actions where if $a_i = 0$ agent i does not adopt (e.g. does not purchase the product), and if $a_i = 1$ agent i does adopt (e.g. purchases the product). Let $W = \{w_0, w_1\}$ be the possible states of the world. Then, $u_i : W \times A \to \mathbb{R}_+$ is the utility function that maps both the action of individual i and the state of the world to i's utility.

Initially, we assume that all players are choosing action 0 and thus obtain a utility of

$u_i(w_0, 0) = u_i(w_1, 0)$ which is independent of the state of the world. However, if an individual decides to adopt, her utility would be $u_i(w_1, 1)$ or $u_i(w_0, 1)$ in state of the world w_1 and w_0 respectively. We assume $u_i(w_1, 1) \geq u_i(w_0, 1)$ as well as a symmetry condition, which implies that the difference between choosing the optimal and suboptimal action in each state of the world is the same, i.e.,

$$u_i(w_0, 0) - u_i(w_0, 1) = u_i(w_1, 1) - u_i(w_1, 0). \tag{1}$$

We also assume that individuals know their utilities conditional on the state of the world but not which state pertains. Under these conditions, their best response is to choose the action a_i^* that maximizes their expected utilities, that is,

$$a_i^* \in \arg\max_{a_i \in A} \sum_{w \in W} P_i(w \mid \widehat{a}_{-i}) u_i(w, a_i)$$

where $P_i(w \mid \widehat{a}_{-i})$ is the updated belief that the state of the world is w, given the action chosen by all other individuals in the population. Note that if each individual only observes a subset of the population then $P_i(w \mid \widehat{a}_{-i})$ only depends on the components of $\widehat{a}_{-i}$, which correspond to i's neighbors.

The beliefs $P_i(w \mid \widehat{a}_{-i})$ can be generated in many different ways, each of which relies on a different set of assumptions. For concreteness, we present here a simple updating rule to illustrate how implicit externalities can also give rise to influence-response functions that only depend on the number of individuals choosing each action. Let us assume that the only possible transition is from 0 to 1 (i.e. once an individual adopts this is an irreversible decision). Before making a decision, an individual receives a signal about the state of the world s, which is either 1 or 0. If the state of the world is w_0, the signal is 0 with probability $p > \frac{1}{2}$; whereas if the state of the world is w_1, the signal is 1 with probability p. Therefore, p measures the accuracy of the signal: the higher the value of p, the more informative the signal becomes. Furthermore, p need not be the same across individuals; it can be obtained from a distribution (with $f(p)$ and $F(p)$ as pdf and cdf respectively) where $\overline{p}$ is the average accuracy. [20]

Consider first the case where an individual, say $i \in N$, *only* observes her own signal before making a decision. Assume also that, *a priori* each state of the world is equiprobable. That

[20] Note that our framing is very similar to Bickchandani et al. (1998), with the important exception that we treat observations as independent (i.e. non sequential). Also Young (2005) and Jensen (1982) posit the consequences for the shape of the adoption curve in a different context, where agents directly observe the realized payoffs of the two competing technologies. Another related model is Bala and Goyal (1998). There are several reasons why the model proposed by Bala and Goyal (1998) is significantly more complicated. First, in their model, given a state of the world, the utility obtained from each action is randomly determined. Second, individuals observe the utilities as well as the choices of neighbors. Third, and more importantly, individuals use past experience both of their own and their neighbors to decide what to choose in each period.

is, $\Pr(w_1) = \Pr(w_0) = \frac{1}{2}$. Then if she receives a signal 1, using Bayes rule, her beliefs about the state of the world are updated as follows

$$\Pr(w_1 \mid s_i = 1) = \frac{\Pr(s_i = 1 \mid w_1)\Pr(w_1)}{\Pr(s_i = 1 \mid w_1)\Pr(w_1) + \Pr(s_i = 1 \mid w_0)\Pr(w_0)}$$

where $\Pr(w_1 \mid s_i = 1)$ is the probability assigned to state w_1, after receiving a signal of 1 and $\Pr(s_i = 1 \mid w_r)$ is the probability of receiving a signal 1, given that the state of the world is w_r, where $r \in \{0, 1\}$.

Note that

$$\Pr(w_1 \mid s_i = 1) = \frac{p_i(1/2)}{p_i(1/2) + (1 - p_i)(1/2)} = p_i.$$

By symmetry, we can show that

$$\Pr(w_0 \mid s_i = 0) = \frac{p_i(1/2)}{p_i(1/2) + (1 - p_i)(1/2)} = p_i$$

where $\Pr(w_0 \mid s_i = 0)$ is the probability assigned to state w_0, after receiving a signal of 0.

If $s_i = 1$, choosing 1 has a higher expected utility than choosing 0. Specifically

$$p_i u_i(w_1, 1) + (1 - p_i)u_i(w_0, 1) > p_i u_i(w_1, 0) + (1 - p_i)u_i(w_0, 0)$$

or

$$p_i(u_i(w_1, 1) - u_i(w_1, 0)) > (1 - p_i)(u_i(w_0, 1) - u_i(w_0, 0))$$

since $p_i > \frac{1}{2}$ and condition (3) holds.

Analogously, if $s_i = 0$ then action 0 is optimal since

$$p_i u_i(w_0, 0) + (1 - p_i)u_i(w_1, 0) > p_i u_i(w_0, 1) + (1 - p_i)u_i(w_1, 1).$$

Consider now a more general situation. Assume that in addition to receiving a private signal, individual i observes the action taken by n_i other individuals before making his decision. More specifically, assume that k_i of them are choosing 1 and $n_i - k_i$ are choosing 0. Under these circumstances, she must (i) try to infer from their actions their signals; and (ii) update his beliefs given his signal and the inferred signals of others. To address part (i) we assume that an individual thinks that other individuals make their decisions independently of each other, therefore, they only look at their own signal. This simplifying assumption is a consequence of the independent effect assumption considered throughout the paper and implies, in effect, that the actions of others reveal their signals in a trivial way (i.e., action 1 if signal 1 and action 0 if signal 0). This behavior is consistent with the idea of pluralistic ignorance from social-psychology (C. Bicchieri and Y. Fukui, 1999): people pay attention to others, but do not think that others are also paying attention to others.

For part (ii), we consider that an individual applies Bayes rule to update her beliefs after observing her own signal and all other signals.[21] Since she does not know the signal's accuracy of others, we assume that she takes as an approximation the average $\overline{p}$; however, she does know the accuracy of her own signal p_i. Therefore, if $s_i = 1$

$$\Pr(w_1 \mid (k_i, n_i - k_i)) = \frac{\Pr((k_i, n_i - k_i) \mid w_1) \Pr(w_1)}{\Pr((k_i, n_i - k_i) \mid w_1) \Pr(w_1) + \Pr((k_i, n_i - k_i) \mid w_0) \Pr(w_0)} \tag{2}$$

where $\Pr(w_1 \mid (k_i, n_i - k_i))$ is the probability assigned to state w_1, given that k_i neighbors are choosing action 1 and $n_i - k_i$ are choosing action 0.

Equation (2) can also be written as

$$\Pr(w_1 \mid (k_i, n_i - k_i)) = \frac{\overline{p}^{k_i}(1-\overline{p})^{n_i-k_i} p_i}{\overline{p}^{k_i}(1-\overline{p})^{n_i-k_i} p_i + (1-\overline{p})^{k_i}(\overline{p})^{n_i-k_i}(1-p_i)}.$$

Given that the utilities are symmetric, action 1 is preferred to action 0 if and only if $\Pr(w_1 \mid (k_i, n_i - k_i)) > \frac{1}{2}$. Then action 1 is chosen (after receiving a signal of 1) if and only if

$$1 + (\frac{\overline{p}}{1-\overline{p}})^{n_i - 2k_i}(\frac{1-p_i}{p_i}) < 2$$

or equivalently

$$(\frac{\overline{p}}{1-\overline{p}})^{n_i - 2k_i} < \frac{p_i}{1-p_i}$$

which implies

$$\frac{n_i}{2} - \frac{1}{2}\frac{\ln(p_i/(1-p_i))}{\ln(\overline{p}/(1-\overline{p}))} \leq k_i.$$

Assuming now that $s_i = 0$, then

$$\Pr(w_1 \mid (k_i, n_i - k_i)) = \frac{\overline{p}^{k_i}(1-\overline{p})^{n_i-k_i}(1-p_i)}{\overline{p}^{k_i}(1-\overline{p})^{n_i-k_i}(1-p_i) + (1-\overline{p})^{k_i}(\overline{p})^{n_i-k_i} p_i};$$

thus, action 1 is chosen if and only if

$$\frac{n_i}{2} + \frac{1}{2}\frac{\ln(p_i/(1-p_i))}{\ln(\overline{p}/(1-\overline{p}))} \leq k_i.$$

Note that, if the signal's accuracy is equal to the average $\overline{p}$, the thresholds are $\frac{n_i-1}{2}$ and $\frac{n_i+1}{2}$ after a signal of 1 and 0 respectively. A consequence of this observation, illustrated in Figure 5, is that if $p_i > \overline{p}$, the two thresholds spread out; whereas if $p_i < \overline{p}$, the thresholds become closer.

Insert Figure 5 about here

[21] Applying Bayes rule and pluralistic ignorance at the same time might seem like a logical contradiction. Really, however, we are assuming bounded rationality, and Bayes rule is just a well defined way of saying that people do, in fact, make inferences from behavior of others. As we will show, Bayes rule leads to simple heuristics that seem plausible, thus ultimately it is irrelevant whether individuals actually apply Bayes rule or use some other procedure (Gigerenzer, Todd and Group, 1999).

To conclude, the influence-response function that arises from our simple inference procedure is:

$$r(k_i) = \begin{cases} 1 \text{ if } k_i^U \leq k_i \\ 0 \text{ otherwise} \end{cases}$$

where

$$k_i^U = \begin{cases} \frac{n_i}{2} - \frac{1}{2}\frac{\ln(p_i/(1-p_i))}{\ln(\overline{p}/(1-\overline{p}))} \text{ if } s_i = 1 \\ \frac{n_i}{2} + \frac{1}{2}\frac{\ln(p_i/(1-p_i))}{\ln(\overline{p}/(1-\overline{p}))} \text{ if } s_i = 0 \end{cases}$$

There is, of course, nothing unique about this procedure, and one could propose other, equally valid procedures for deriving social influence from the observation of other people's actions. Furthermore, while only upward thresholds arise in this model, one could also think of ways to generalize it to account for other forms of externalities (e.g. assuming that individuals have opposite preferences, and that this is common knowledge). The important point is that although implicit externalities are based on a different set of assumptions than explicit externalities, the consequence for the functional form of the decision rule can be qualitatively indistinguishable. For example, if we compare the model of implicit externalities proposed here with the example of public goods with a convex production function introduced in Section 2.1.3, we find that the behavior of individuals in both scenarios is characterized by an upward threshold which will ultimately give rise to very similar qualitative collective outcomes.

To conclude with the first part of the paper, we reiterate that our aims are two-fold: (1) To provide a unified framework for studying social influence, and show how a wide range of micro-mechanisms can be characterized by simple influence-response functions. (2) To study how individual response functions aggregate and produce collective dynamics. Up to this point, we have focused on the first question; in what follows we address the second.

3 Collective dynamics

Although useful for the purposes of deriving influence-response functions, in reality our distinction between explicit and implicit externalities is somewhat artificial, for two reasons. First, many real decision making situations may very well embody both kinds of externalities, which in practice may be difficult or even impossible to disentangle. For example, when an individual is observed to change her behavior to conform to some group norm, did she do so because (a) she privately disagrees with the norm, but wants to avoid sanctioning from other group members; (b) the benefits of group coordination are so great that the precise norms on which they are coordinated are irrelevant; (c) her beliefs have shifted unconsciously through repeated exposure to group beliefs; or (d) she has made a conscious decision to alter her beliefs on the assumption that "everyone else can't be wrong"? In this hypothetical scenario, the first two explanations represent explicit externalities, and the latter two represent implicit externalities, where in each case the underlying psychology and cost-benefit analysis (if there is one) may be quite different. In truth, however, probably all these effects are in play simultaneously, and thus it may be ultimately unhelpful to draw distinctions between them.

A second potential problem with separating explicit and implicit externalities is that in many strategic situations, even in relatively simple two-player games, the players may have considerable uncertainty regarding the strategies being employed by their opponents, or even the game they are playing. Thus in order to optimize their expected utilities over time, players must engage in what is called "strategic learning" (Young, 2004), which involves both explicit and implicit externalities. As with the scenario above, it may be impossible in practice to determine how much of the social influence players feel can be attributed to learning from others, and how much is inherent to the structure of the underlying game itself.

In both these scenarios, however, it may still be possible to elicit an individual's influence-response function empirically, either from data or controlled experiments. We are not aware of any explicit attempts to map influence-response functions; however the seminal experiments of S. E. Asch (1953) and the variations that followed (R.Bond and P.B. Smith, 1996) suggest that such an exercise is possible, at least under some circumstances. Recent work by J. Leskovec et al. (2005) also suggests that influence-response functions can be reconstructed from empirical data, derived from online recommender networks.

Assuming that influence-response functions can be constructed in the absence of an unambiguous theoretical framework, we now show that regardless of their origin, it is precisely the functional form of the influence-response function that critically determines the aggregate behavior. This step represents an important advantage of our approach which in effect allows us to compute the equilibria, given only the influence-response functions (i.e. without knowing how they were obtained in the first place).

For tractability reasons, to describe the collective outcomes, we will assume *global* and *anonymous* interactions in the population. Specifically, this implies that $w_{ij} = w$ for all $i, j \in N$, where for simplicity we normalize $w = 1$. Although, these assumptions are crucial for our results to hold it is clear that, in reality, individuals have only limited information about the behavior of others and it is precisely the intricate structure of the social network what leads to a great amount of unpredictability in the collective outcomes (Watts, 2002). The study of the collective dynamics accomplished here, however, helps us understand the phenomenon of social influence in its simplest form and constitutes a starting point that we believe will be useful for guiding further research.

As an approximation, useful for deriving the collective outcomes, consider a continuous population of individuals. We therefore redefine the influence-response function ($r_i : [0,1] \to [0,1]$) as a function that maps the fraction of individuals choosing action 1 in the population I, with the probability of choosing action 1, $r_i(I)$. Although in all the examples described in the first part of the paper the population is finite, the results obtained with this continuous approximation provides predictions that are appropriate to explain the qualitative collective outcomes of a finite but large population. In what follows, we embed the model in a dynamic framework of strategy revision, and analyze how individual-level decision processes aggregate and produce collective outcomes.

3.1 Discrete dynamics (synchronized revisions)

The approach followed in this section is closely related to work of Granovetter (1978, 1986 and 1983) in which he analyzes how aggregate outcomes depend on the distribution of preferences in the population (i.e. thresholds). Our approach, however, is conceivable more general in connecting primitive assumptions about the behavior of individuals to different equilibrium outcomes as well as providing a systematic and formal overview of the problem of social influence. In addition, later in the paper, we depart from Granovetter's approach by considering asynchronous updating dynamics as well as stochastic best responses.[22]

Consider the following simple dynamics. In each period, all individuals synchronously revise their strategy and choose a myopic-best response. That is, an individual chooses the action that maximizes his utility (or expected utility), given the action of all other individuals in the previous period. This dynamics is consistent with a bounded rationality view of the world in that agents choose a best response to the strategy profile of the previous state, without anticipating that all other agents will also revise (and probably change) their strategy next period. Based on this assumption, we can describe how the fraction of individuals choosing action 1, denoted by $I(t)$ hereafter, evolves over time as follows:

$$I(t+1) = H(I(t)) \tag{3}$$

where $H : [0,1] \to [0,1]$ only depends on the influence-response functions of the individuals in the population as described shortly.

[22] Michael Chwe (1999) also presents an extension of Granovetter's work where, in contrast with our approach, completely rational individuals use the information about the thresholds of their closest neighbors to infer their behavior (although individuals care about the behavior of the whole population). The main result (obtained through simulations of small networks) is that network position is much more important in influencing people with low thresholds than people with high thresholds.

Let $r_q(I)$ denote the influence-response function of individuals with a unique threshold $q \in [0,1]$ in the positive and negative externality case, and a vector of two thresholds $q = (q^u, q^d) \in [0,1] \times [0,1]$ in the non-monotonic externality case. Assume that q is distributed in the population according to the pdf $f(q)$ and let $I_q(t)$ be the the fraction of individuals choosing action 1 at time t with a threshold in the interval $[q, q+\delta]$ (where $\delta \sim 0$ in the case of positive and negative externalities, and $\delta = (\delta_1, \delta_2) \sim (0,0)$ in the non-monotonic externality case). Also let $I(t)$ be the overall fraction of individuals choosing action 1 at time t. For every value of q, we can approximate the dynamics by[23]

$$I_q(t+1) = r_q(I(t))$$

where

$$I(t) = \int_0^1 f(q) I_q(t) dq$$

Therefore, we obtain that

$$I(t+1) = r(I(t)) \tag{4}$$

where we define

$$r(I(t)) = \int_0^1 f(q) r_q(I(t)) dq$$

as the average influence-response function. Then, following the notation introduced above, $r(I(t)) = H(I(t))$.

Consider the following cases:

(a) **Positive externalities**: Each individual's influence-response function is characterized by an upward threshold q_i^u, where $i \in N$, which is distributed in the population with $f(q^u)$ and $F(q^u)$ as pdf and cdf respectively. Notice that $F(q^u)$ is the fraction of individuals with thresholds lower or equal to q^u. It is straightforward to show that

$$r(I(t)) = \int_0^1 f(q^u) r_{q^u}(I(t)) dq^u = \int_0^{I(t)} f(q^u) dq^u = F(I(t))$$

Hence, the fixed points of $F(q^u)$ correspond with the stationary states of the dynamics. Furthermore, the shape of $F(q^u)$ determines the stability of the fixed points in interesting ways. Consider first the case where the population is homogenous and therefore all individuals have the same threshold, say I^u. Then $r(I)$ is a step function, characterized by an upward threshold I^u. As illustrated in Figure 5.a, the stationary states of the dynamics are $I^1 = 1$ and $I^0 = 0$. These states correspond to the states where all individuals are choosing the same action, i.e. $a^0 = (0, 0, ...0)$ and $a^1 = (1, 1, ..., 1)$; the basin of attraction of each

[23] Obviously, this equation is an exact representation of the dynamics if $\delta \to 0$.

stationary state is given by I^u. In other words, if the initial fraction of individuals choosing action 1 is below I^u, the dynamics in the next period reaches I^0 which is a sable state; otherwise, it reaches I^1, which is also stable.

An example with positive externalities where these results apply is in the context of the provision of a public good with a convex production function. In equilibrium either everybody contributes to the public good or nobody does. Notice that the (efficient) equilibrium where all individuals contribute is selected only if, on account to exogenous actions not explained in our model, a fraction larger than I^u of individuals contribute in the initial state (for example because they are induced to do so). Obviously, the likelihood of such an event is related to the value of I^u, which in turn depends on the parameters of the micro-mechanical details of the process such as benefits and costs of contributing.

Consider now as a more general case where the upward thresholds follow a symmetric beta distribution (i.e., $B(\mu, \mu)$) with cdf $F_\mu(q^u)$, where $\mu > 0$. It is worth noting that, if $\mu > 1$ then the states I^0 and I^1, are stable, whereas if $\mu < 1$ (bimodal distribution), the unique stable state is such that some individuals are choosing 1 and others are choosing 0. This distinction might have important consequences for various phenomena; for example, assuming that an infinitesimally small fraction of individuals initially chooses 1, would this action spread to a larger fraction of the population or, would it vanish? The answer is that action 1 will spread if and only if

$$F'_\mu(0) > 1$$

which corresponds with the condition $\mu < 1$, as illustrated in Figure 5.

Insert Figure 5 about here

In other words, heterogeneity in the population allows for a cascades to occur: initially only a small fraction of individuals adopt action 1 (those that clearly prefer 1 to 0); but due to the positive-externality effects other individuals adopt 1 as well; and so on. In fact, small differences in the distribution of thresholds might lead to radically different equilibrium outcomes. For example, in the beta distribution of thresholds presented above, the cases where μ is slightly above and below 1. These results also highlight the importance of knowing the full distribution of thresholds instead of simply the average, since two distributions with the same average threshold can have very different aggregate outcomes.[24]

Returning to the example of public goods with convex production functions, we find that, unlike in the homogeneous case, heterogenous populations may be able to sustain contribution in equilibrium even in cases where only a small fraction of the population contributes initially. Whether or not this result is possible, as well as whether or not the corresponding

[24] Granovetter (1978) emphasizes precisely this point.

cascade causes everyone, or just a fraction to contribute depends critically on the specific distribution of thresholds.

(b) **Negative externalities**: Each individual's influence-response function is characterized by a downward threshold q_i^d, for every $i \in N$, distributed in the population with $f(q^d)$ and $F(q^d)$ as pdf and cdf respectively. In this case, it is straightforward to show that

$$r(I(t)) = \int_0^1 f(q^d) r_{q^d}(I(t)) dq^d = \int_{I(t)}^1 f(q^d) dq^d = 1 - F(I(t))$$

Consider first the case where all individuals have the same threshold, say I^d. Then, the influence-response function $r(I)$ is step function characterized by the downward threshold I^d. As illustrated in Figure 6.b, $H(I)$ has no fixed points and the dynamics gets trapped in a cycle that alternates between the states I^0 and I^1. For instance, if the initial condition is below I^d. In the next period all individuals choose action 1 (I^1); then in the following period all individuals simultaneously choose 0; and so on. Consider, for example, that individuals have to decide whether to open or not a grocery store. Obviously, the higher the number of already existing grocery stores, the lower the benefits derived from opening a new one. Assuming an homogeneous population and the simultaneous updating dynamics described above, we conclude that there is no stationary states but instead the dynamics cycles around the two extreme cases; all individuals open a store or nobody does. This is clearly not a good prediction of what occurs in reality. What seems to be driving the result is the extreme assumption that all individuals simultaneously revise their strategy every period. If, on the contrary (as studied in Section 3.2) only a few individuals update every period, the results provide more reasonable predictions.

Assume now the more general case where the thresholds are distributed in the population according to a continuous cdf $F(q^d)$. In such a case, given that $r(0) = 1$, $r(1) = 0$ and $r(I)$ is decreasing, there exist a unique stationary state of the dynamics I^*, where $I^0 < I^* < I^1$, the stability of which will depend on the slope of $F(q^d)$ evaluated at the stationary state I^*. Specifically, if $F'(I^*) < -1$ the dynamics behaves aperiodically, whereas if $F'(I^*) > -1$, the dynamics converges to I^*. Finally, if $F'(I^*) = -1$, the dynamics cycles around the value I^*. As an illustration see Figure 6. Note that, also in this case, the distribution of thresholds has some qualitative impact on the results; however, unlike the positive externality case, the long-run behavior of the dynamics does not depend on initial conditions.

As a consequence of this general result, public goods games with concave production functions, therefore, never converge to a situation where all individuals contribute. This finding contrasts with the result obtained for public goods with convex production functions; thus highlighting our earlier point that superficially similar applications might lead to very distinct collective behavior.

Insert Figure 6 about here

(c) **Non-monotonic externalities**: In this last case we assume that each individual's influence-response function is characterized by two thresholds, an upward threshold q_i^u and a downward threshold q_i^d, where $q_i^u \leq q_i^d$. Let $f(q^u, q^d)$ be the joint density function. Then

$$r(I) = \int_0^1 \int_0^1 f(q^u, q^d) r_{(q^u,q^d)}(I(t)) dq^u dq^d = \int_0^I (\int_I^1 f(q^u, q^d) dq^d) dq^u$$

and therefore,

$$r(I) = \int_0^I (\int_0^1 f(q^u, q^d) dq^d - \int_0^I f(q^u, q^d) dq^d) dq^u$$

or equivalently

$$\begin{aligned} r(I) &= \int_0^I f_u(q^u) dq^u - \int_0^I (\int_0^I f(q^u, q^d) dq^d) dq^u = \\ & F_u(I) - \int_0^I (\int_0^I f(q^u, q^d) dq^u) dq^d \\ &= F_u(I) - \int_0^I (\int_0^I f(q^u, q^d) dq^u - \int_I^1 f(q^u, q^d) dq^u) dq^d \end{aligned}$$

where $f_h(q^h)$ and $F_h(q^h)$ for $h \in \{u, d\}$, represent the marginal density and cumulative distribution functions, respectively of $f(q^u, q^d)$.

Therefore,

$$r(I) = F_u(I) - \int_0^I f_d(q^d) dq^d + \int_0^I (\int_I^1 f(q^u, q^d) dq^u) dq^d.$$

Given that $q^u < q^d$, then

$$\int_0^I (\int_I^1 f(q^u, q^d) dq^u) dq^d = 0$$

and therefore

$$r(I) = F_u(I) - F_d(I).$$

In this case, the stable states depend critically on the shape of $F_u(I) - F_d(I)$. A general conclusion is that, as long as there is a positive fraction of individuals with a non-degenerate downward threshold, I^1 is never stationary. We focus first on the case where all individuals have the same two thresholds I^u (an upward threshold) and I^d (a downward threshold).[25] In such a case and as illustrated in Figure 7.a, given any initial condition, the dynamics

[25] This is obviously just a particular example of a non-monotonic externality, however, in this paper we focus on this case since most of the insights obtained here can easily be extended to the more general form.

converges to the state where all individuals are choosing action 0 (I^0). For example, if the initial condition is either below I^u or above I^d, all individuals have incentives to switch to action 0 next period. If, by contrast, the initial condition is above I^u and below I^d, it takes two periods to reach the state I^0; all individuals switch to action 1 in the first period, but then switch to action 0 in the second period. To illustrate this case with an example, consider the context of public goods with an s-shaped production function; then, independent of initial conditions, the dynamics will rapidly converge to a situation where none of the individuals in the population contribute. As with the case of negative externalities, this result is driven by the synchronous nature of the updating; and more reasonable results prevail once this assumption is violated (see Section 3.2).

Also, Figure 7.b and 7.c show that for a wide range of distribution functions there are two stationary states, I^0 and I^*, where $0 < I^* < 1$. The stability of I^* depends on the slope of $F_u(I) - F_d(I)$ at I^*. This case constitutes a mixture between positive and negative externalities: as in the positive externality case, the long run behavior of the dynamics may depend on initial conditions; but like negative externalities, the dynamics might exhibit aperiodic behavior.

As explained earlier, this type of non-monotonic externality seems common in the case of fashionable products. In fact, as the results seem to point out, precisely in fashion one might observe some sort of cyclic or chaotic phenomena, where styles that were popular in the past are reinvented and become popular again. Apart from fashion, non-monotonic externalities also may arise in the context of public good games with an s-shaped production function, thus raising again the importance of knowing the shape of the production function to better predict the collective outcomes.

Insert Figure 7 about here

3.2 Continuous Dynamics (asynchronized revisions)

In the previous section, we have considered a discrete dynamics where in every period *all* individuals choose the action which is a best response to what the remaining individuals did in the last period. In this section, however, we consider a dynamics where only a small fraction of individuals revise every period. There are three main reasons why this asynchronized version of the dynamics seems more appropriate for most contexts. First, although in some situations, actors' decisions will be forcibly synchronized, in general that will not be the case, thus no synchronous mechanism exists. Second, although we have focused on analyzing how individuals behave with respect to a simple binary decision, in reality, individuals have to face multiple decisions each of which might reflect different aspects of their lives (e.g. which car to buy, political party to vote, TV series to watch, etc.).

Therefore, it is reasonable to assume that due to inertia as well as time constraints individuals do not revise (and optimize) all decisions every period. Third, if we assume that only a small fraction of individuals revise every period, the action profile does not vary significantly from one period to another; thus, the action profile at a certain time is a good approximation of the action profiles in the near future, which makes a myopic-best response a reasonable behavioral assumption.

For tractability reasons, instead of analyzing a discrete dynamics incorporating this feature, we study its continuous counterpart; that is, in every period t, individuals revise their strategy at a rate $\lambda > 0$ and, as before, if an individual revises her action, she chooses a myopic-best response. The dynamics describing the evolution over time of the fraction of individuals choosing action 1 is therefore:

$$\frac{dI(t)}{dt} = \widetilde{H}(I(t))$$

where $\widetilde{H} : [0,1] \to [0,1]$ depends on the influence-response functions of the individuals in the population as we will show below.

To compute the stationary states of the dynamics we use standard mean-field theory. Specifically we assume that the transitions from action 0 to action 1 and vice versa take place at the average rate. We consider a population characterized by a distribution of thresholds, noting that the homogeneous case simply corresponds to a particular instance of this more general framework. Recall from the previous section that $I_q(t)$ represents the fraction of individuals choosing action 1 at time t with a threshold in the interval $[q, q+\delta]$ (where again $\delta \sim 0$ in the case of positive and negative externalities and $\delta = (\delta_1, \delta_2) \sim (0,0)$ in the non-monotonic externality case). Hence, the mean-field dynamical equation for each $I_q(t)$ can be be approximated by:

$$\frac{dI_q(t)}{dt} = -I_q(t) rate_{I(t),q}(1 \to 0) + (1 - I_q(t)) rate_{I(t),q}(0 \to 1) \tag{5}$$

where $rate_{I(t),q}(1 \to 0) = \lambda(1 - r_q(I(t)))$ and $rate_{I(t),q}(0 \to 1) = \lambda r_q(I(t))$. Noting that $r_q(I(t))$ is the influence-response function of individuals with threshold q.[26] Substituting the expressions for the rates in equation (5), we find that

$$\frac{dI_q(t)}{dt} = \lambda(r_q(I(t)) - I_q(t)). \tag{6}$$

As the overall fraction of agents choosing 1 in the population is

$$I(t) = \int_0^1 f(q) I_q(t) dq \tag{7}$$

[26] Again, this is an exact approximation of the dynamics if $\delta \to 0$.

then it follows that

$$\frac{dI(t)}{dt} = \int_0^1 f(q) \frac{dI_q(t)}{dt} dq. \tag{8}$$

Now, replacing equation (6) in equation (8), we find that

$$\frac{dI(t)}{dt} = \int_0^1 f(q)\lambda(r_q(I(t)) - I_q(t))dq$$

or equivalently,

$$\frac{dI(t)}{dt} = \lambda(r(I(t)) - I(t)) \tag{9}$$

where, we define

$$r(I(t)) = \int_0^1 f(q) r_q(I(t))dq$$

again as the average influence-response function. Thus, following the notation introduced above, we obtain that $\widetilde{H}(I(t)) = \lambda(r(I(t)) - I(t))$.

After imposing the stationary condition ($\frac{dI(t)}{dt} = 0$), we find that

$$I = r(I) \tag{10}$$

which coincides with the condition obtained in the discrete dynamics. In other words, the stationary states of the discrete and continuous dynamics coincide. However, as we will show next, the stability properties of the stationary states are typically different. Let us again consider positive, negative and non-monotonous externalities separately.

(a) **Positive externalities**: The influence-response function of an individual is characterized by an upward threshold. These thresholds are distributed in the population according to the pdf $f(q^u)$. Note that, as well as in the discrete dynamics case it can easily be shown that $r(I(t)) = F(I(t))$ where $F(q^u)$ is the cdf of thresholds. Therefore the dynamical equation is

$$\frac{dI(t)}{dt} = \lambda(F(I(t)) - I(t)),$$

where again stationary states coincide with the fixed points of the threshold distribution function. In addition, for positive externalities, not only the set of stationary states, but also their stability properties coincide with those obtained in the discrete dynamics. The reason why this holds is that in the positive externality case, the discrete dynamics only exhibited equilibrium behavior (because $F'(I) \geq 0$ for all $I \in [0, 1]$). The rates of convergence to the stable states, however, now depend on the updating rate λ. To illustrate, consider the distribution of thresholds with cdf $F(q^u)$ represented in Figure 8. Then, as shown in Figure 9.a the dynamics converge to state I^0 or state I^1 depending on whether the initial fraction of agents choosing action 1 is above or below the interior solution for $F(I^*) = I^*$.

Introduce Figures 8 and 9 about here

(b) **Negative externalities**: The influence-response function of an individual is again characterized by a downward threshold, where these thresholds are distributed according to the pdf $f(q^d)$. As in the discrete case, $r(I(t)) = 1 - F(I(t))$, where $F(q^d)$ is the cdf of thresholds, and the dynamical equation is

$$\frac{dI(t)}{dt} = \lambda(1 - F(I(t)) - I(t)).$$

As illustrated in Figure 9.b, there exists a unique stationary state which is also stable, where some individuals choose 1 and others choose 0. This result holds for all distribution of thresholds and differs with the result obtained for the discrete dynamics. For example, in the homogenous population case, the continuous dynamics converge to an interior state I^* (i.e. $0 < I^* < 1$), whereas the discrete dynamics cycled from state I^0 to state I^1 and vice versa.

(c) **Non-monotonic externalities**: As before, individuals are characterized by having two thresholds, an upward threshold and a downward threshold q^u and q^d respectively, where these thresholds can vary across individuals in the population. Assume $f(q^u, q^d)$ is the pdf of the distribution of thresholds in the population. Then, as already shown for the discrete dynamics, the aggregate influence-response function $r(I(t)) = F_u(I(t)) - F_d(I(t))$, where $F_u(q^u)$ and $F_d(q^d)$ are the cdfs of the marginal distributions. Hence, the dynamical equation is

$$\frac{dI(t)}{dt} = \lambda(F_u(I(t)) - F_d(I(t)) - I(t))$$

where once again the resulting behavior will depend critically on the properties of the distribution function. It is straightforward to see that, for certain threshold distributions we obtain the dynamics illustrated in Figure 9.c where there exist two stable states I_0^* and I^* such that $0 < I^* < 1$, again in contrast with the more complicated dynamics of the synchronous case.

To conclude, the analysis of the continuous dynamics (as compared with the discrete version) highlights the importance of the aggregation mechanisms as well as the distribution of preferences. In particular, aggregate outcomes cannot be determined by a simple counting of preferences (or the average preference) since this might generate wrong results. Furthermore, when individuals are subject to social influence, one cannot infer individual dispositions from long run collective behavior, but must also consider the intermediate states of the dynamics that led to it.

4 Extensions

4.1 Stochastic influence-response function

In the previous sections we have studied situations where the influence response is a deterministic function of the social signal I, i.e. $r_i(I) \in \{0,1\}$, for every $i \in N$. Nevertheless, one of the advantages of our approach is that we can generalize the model to account for stochastic influence-response functions i.e. where for every $i \in N$, $r_i(I) \in [0,1]$. In order to determine the collective outcomes for this generalization, one must simply substitute in the dynamical equations obtained in Section 3 the deterministic influence-response functions by their stochastic counterparts. There are many reasons why an individual may behave stochastically (or appear to do so): (a) uncertainty about the social signal (b) errors in the computation of the best-response (c) uncertainty about the game that is being played (d) uncertainty about who the individual is going to play the game with. In all these hypothetical scenarios, the influence-response function cannot be described as a deterministic function of the social signal I. To clarify this point we formalize next two of the cases described above.

First, consider case (a) where individuals are uncertain about the value of I. Moreover, each individual receives a random signal which is normally distributed with mean I and variance σ. Therefore, the signal is denoted as $\widetilde{I}$ and is distributed as $N(I,\sigma)$. After receiving the signal, individuals choose a (deterministic) myopic-best response. Consider, for example, the case with positive externalities in which the influence-response function is characterized by an upward threshold q; that is,

$$r(\widetilde{I}) = \begin{cases} 1 \text{ if } \widetilde{I} \geq q \\ 0 \text{ if } \widetilde{I} < q. \end{cases} \tag{11}$$

Given that $\widetilde{I}$ is a random variable, we can rewrite equation (11) as follows:

$$r(I) = \Pr(\widetilde{I} \geq q) = \Pr(N(I,\sigma) \geq q)$$

which is stochastic.

Assume that not only the threshold $q \in [0,1]$, but also the variance $\sigma \in [0,L]$ of the signal (where L denotes the maximum variance) is distributed in the population with a density function $f(\sigma,q)$. Then, the average influence-response function necessary to obtain the collective outcomes is simply:

$$r(I) = \int_0^M \Big(\int_0^1 \Pr(N(I,\sigma) \geq q) f(q,\sigma) dq \Big) d\sigma.$$

An alternative way to motivate stochastic decisions would be to consider a situation where individuals make errors when computing their best responses (case (b) mentioned above). Assume, however, that the propensity to play an action is exponentially related to its utility[27]; that is,

$$r(I) = \frac{e^{\beta u(1,I)}}{e^{\beta u(1,I)} + e^{\beta u(0,I)}},$$

where $\beta > 0$ measures the degree of randomness in the choice of i. In particular, if $\beta \to +\infty$ this behavioral rule coincides with the standard (deterministic) myopic best-response; whereas if $\beta = 0$ both actions are chosen with equal probability, independently of I. Also note that $u(a, I)$ is the utility function associated with choosing action $a \in \{0, 1\}$ when a fraction of I individuals are choosing action 1. Consider a heterogeneous population where different individuals are characterized by different values of β. More specifically, let $f(\beta)$ be the pdf of the distribution of β's in the population. The aggregate influence-response function would correspond to

$$r(I) = \int_0^{+\infty} \frac{e^{\beta U(1,I)}}{e^{\beta U(1,I)} + e^{\beta U(0,I)}} f(\beta) d\beta.$$

4.2 Irreversible dynamics

The dynamics studied in the body of the paper allows for transitions to go in both directions. That is, an individual choosing action 1 might be willing to choose action 0 and vice versa. There are some situations, however, where this reversibility in actions is not reasonable; for example, in the implicit externality model presented in Section 2.2 only transitions from action 0 to action 1 are possible. For concreteness, let us focus on the continuous dynamics presented before in the paper. The dynamical equation of this variation of the model is

$$\frac{dI_q(t)}{dt} = (1 - I_q(t))\lambda r_q(I(t))$$

which in turn implies that

$$\frac{dI(t)}{dt} = \int_0^1 f(q)(1 - I_q(t))\lambda r_q(I(t)) dq,$$

where $f(q)$ is the pdf of the distribution of thresholds in the population.

After imposing the stationary condition, $\frac{dI(t)}{dt} = 0$, we obtain the equation

$$\int_0^1 f(q) r_q(I(t)) dq = \int_0^1 f(q) I_q(t) r_q(I(t)) dq. \tag{12}$$

[27] This stochastic best response rule was proposed by Young (1998).

There are again three possible cases according to the form of the externality: (i) positive externalities (ii) negative externalities (iii) non-monotonic externalities, where we consider in detail only case (i). Note that equation (12) can be written as

$$\int_0^I f(q)dq = \int_0^I f(q)I_q dq,$$

which holds if and only if

$$I_q = 1 \text{ for all } q \leq I. \tag{13}$$

In what follows we show that state I is sustained in equilibrium if and only if $F(I) \leq I$ where $F(q)$ is the cdf of the distribution of thresholds in the population. First we shall prove that if I is such that $F(I) > I$, then I cannot be an equilibrium state. To do so, consider $\widetilde{I}$ such that $F(\widetilde{I}) > \widetilde{I}$; then

$$\int_0^{\widetilde{I}} f(q)dq = \int_0^1 f(q)\widetilde{I}_q dq,$$

which in turn implies that there exists a value of $\widetilde{q} \leq \widetilde{I}$, such that $\widetilde{I}_{\widetilde{q}} < 1$. Notice, however, that equation (13) is not satisfied and thus $\widetilde{I}$ is not an equilibrium state.
To complete the proof, take $\widetilde{I}$ such that $F(\widetilde{I}) \leq \widetilde{I}$; we must now show that $\widetilde{I}$ is then an equilibrium state. To do so, we must find $\{\widetilde{I}_q\}_{q\in[0,1]}$ such that $\widetilde{I} = \int f(q)\widetilde{I}_q dq$ and such that condition (13) holds. By assumption, we know that

$$\int_0^{\widetilde{I}} f(q)dq \leq \widetilde{I},$$

which allows us to construct $\widetilde{I}_q$ in a way that $\widetilde{I}_q = 1$ for all $q \leq \widetilde{I}$.

Following similar steps, we can prove that for case (ii) a fraction I of individuals choosing action 1 is stationary if and only if $1 - F(I) \leq I$.

5 Final remarks

The main goal of this paper was to understand how individual-level decision processes can be classified as well as aggregated to produce collective outcomes. We have focused on situations where individuals make simple binary decisions (adopt or not a certain action) with externalities; that is, where the decision taken by an individual depends on the decision taken by others. We have briefly reviewed a number of literatures in which the social, psychological, and economic origins of this phenomenon have been explained, as well as a wide variety of models of collective behavior in the presence of externalities. It is often the case, however, that when theoretically grounded models are proposed, they are difficult to generalize; in particular it is unclear how one model relates to another. In this paper, we attempt to overcome this problem by presenting a unified and reasonable general framework to encompass externalities that may vary in origin (explicit or implicit externalities) and form (positive, negative and non-monotonic externalities). We therefore treat in a similar way a number of applications, which at face value seem quite different (e.g. public good games and technology adoption, among others).

The condition of independent effects (or similarly pluralistic ignorance) allows us to describe the behavior of individuals as simple influence-response functions: a one-dimensional (i.e. scalar) function of the social signal. In the most general version of the model, the social signal corresponds to a weighted sum of the number of individuals adopting the action, whereas in the simplest scenario (where there is global and anonymous interactions) the social signal is simply the overall fraction of adopters in the population. In this latter case, individuals are characterized by thresholds that depend only on the total number of others making a decision. In some applications (public goods with convex production functions, technology adoption, etc.) individuals are characterized by upward thresholds (i.e. a certain fraction of individuals must adopt before a given individual does so), while in others (e.g. public goods with concave production functions) they are characterized by downward thresholds (i.e. individuals adopts only if the fraction of individuals that have already adopted is below their threshold). Finally, we have also studied more complex scenarios where individuals are characterized by two thresholds (an upward and a downward threshold) to account for applications such as public good games with s-shaped production functions.

To compute collective outcomes one needs to know the full distribution of thresholds in the population as well as a dynamic model of how these individual preferences interact and aggregate. Both ingredients are key to describe collective behavior and one cannot lead to reasonable predictions without the other. In fact, given a dynamic model, populations with similar average thresholds (or preferences) may generate very different results precisely because of differences in their distribution of thresholds (e.g. a unimodal versus a bimodal distribution of upward thresholds). Also, two different dynamic models (e.g. synchronous versus asynchronous updating dynamics) can give rise to different outcomes even when considering the same population (i.e. distribution of thresholds).

The main finding of this paper (obtained both with the synchronous and asynchronous dynamics) is a qualitative one: on the one hand applications which superficially are very similar, such as public goods games with slightly different shaped production functions, can give rise to qualitatively distinct collective behavior; but on the other hand, very different applications such as some kinds of public good games and certain kinds of social learning problems might generate similar results. Another relevant aspect of our framework, however, is more of a practical one: it allows for significant heterogeneity across individuals, where heterogeneity is typically hard to incorporate in standard economic models, in our framework heterogenous populations can be modelled just as easily as homogenous populations.

For simplicity, we have analyzed a situation with global and anonymous interactions among the individuals in a population. We believe that the formal and systematic analysis of this benchmark case (i.e., where we abstract from the network structure) is a worthwhile enterprise that helps us understand the phenomenon of social influence in its simplest form and constitutes a starting point for further research. Given the interdisciplinary character of the phenomenon of social influence or more generally contagion processes, there is a wide literature encompassing the fields of economics, sociology, statistical physics and epidemiology trying to introduce networks into the analysis. For tractability reasons, the literature in economics has concentrated on the study of structured networks (i.e. networks with some recurrent pattern). Although some general results have been obtained (e.g. Morris, 2000 and Young, 1998, among others), these results are difficult to apply in practice. For instance Morris (2000) characterizes the threshold for the degree of risk dominance of a certain action in a coordination game, that guarantees its contagion in the population. This condition, although extremely general would be difficult to compute unless the network has some recurrent pattern. As a natural complement to this approach there is also significant work where, also for tractability reasons, the network is considered as random (see for example, Watts, 2002; López-Pintado, 2005; Pastor-Satorrás and Vespignani, 2001, among others). A disadvantage of most of these models (both the structured and random approach) is that, although they have satisfactorily answered the question of whether the diffusion of a certain behavior will take place or not, the extent (or size) of the diffusion is virtually ignored.[28]

A generalization of the network metaphor would allow not only for heterogenous actors, but heterogenous classes of actors; for example, media figures, organizations and even institutions, all of which could operate at a different level and interfere with the standard peer to peer influence in intriguing ways. One could for instance assume that the initial fraction of adopters in the population (an exogenous feature of our model) is actually related to mass media advertisement efforts, which then determines critically whether the product spreads or disappears. A more elaborated (and probably more realistic) model would be to consider a hierarchical structure of social influence, where different layers in the hierarchy would correspond to interactions of different entities (e.g. the bottom layer would be the standard social influence, whereas the top layers would correspond to interactions across institutions).

To conclude, one clear advantage of linking the behavior of individuals with the benefits and costs associated with each decision (as described in Section 2) is that it allows us to compare

[28] An exception is Watts (2002) which describes (through simulations) the distribution of possible cascade sizes.

the efficiency of different equilibrium outcomes. If for instance, the collective outcomes predict multiple pareto-ranked equilibria, a plausible policy implication would be to alter the incentives of a small subset of individuals (for example reducing their costs of adopting the desired behavior) to trigger the system towards an efficient outcome (Peterhansl and Watts, 2005).

6 Appendix

Proof of the Propostion 1: Consider the extreme case where $R_i(\widehat{a}_{-i}) = R(\widehat{a}'_{-i})$ if and only if $\widehat{a}_{-i} = \widehat{a}'_{-i}$. Then, we must show that the vector of weights $\{w_{ij}\}_{j\in N}$ are such that also

$$\sum_{j\in N\backslash\{i\}} w_{ij}a_j = \sum_{j\in N\backslash\{i\}} w_{ij}a'_j$$

if and only if $\widehat{a}_{-i} = \widehat{a}'_{-i}$. It is straightforward to see that this condition is analogous to the condition that for all $j \in N\backslash\{i\}$, w_{ij} cannot be written as a combination of sums and minuses of $\{w_{ik}\}_{k\neq j}$. Let us show that we can actually find a vector of weights satisfying this property. Let w be an even number and $w_{ij} = w^{j+1}$. For the sake of contradiction assume that there exists $j^* \in N\backslash\{i\}$ such that

$$w^{j^*+1} = \sum_{k\in N_+} w^{k+1} - \sum_{h\in N_-} w^{h+1} \tag{14}$$

where $N_+ \subset N$, $N_- \subset N$ and for simplicity of notation assume that we have already simplified the equation so that $N_+ \cap N_- = \varnothing$. Then

$$w^{j^*+1} = w^{m+1}(\sum_{k\in N_+} w^{k-m} - \sum_{h\in N_-} w^{h-m})$$

where $m = min\{N_+ \cup N_-\}$. Assume that $m \in N_+$[29], then

$$w^{j^*+1} = w^{m+1}(1 + \sum_{k\in N_+\backslash\{m\}} w^{k-m} - \sum_{h\in N_-} w^{h-m}). \tag{15}$$

By construction $j^* \neq m$. Thus, if $j^* > m$ then

$$w^{j^*-m} = (1 + \sum_{k\in N_+\backslash\{m\}} w^{k-m} - \sum_{h\in N_-} w^{h-m}) \tag{16}$$

Equation (16) leads to a contradiction since the left hand side is an even number whereas the right hand side is an odd number.

Furthermore, if $j^* < m$ then

$$1 = w^{m-j^*}(1 + \sum_{k\in N_+\backslash\{m\}} w^{k-m} - \sum_{h\in N_-} w^{h-m})$$

[29] A very similar argument would follow if instead $m \in N_-$.

which is also a contradiction since the left hand side of this equation is an odd number, whereas the right hand side is even.[30]

References

[1] Anderson, R. M., May, R. M. 1991. Infectious Diseases of Humans. Oxford University Press, Oxford.

[2] Arthur, W. B., Lane, D. 1993. Information Contagion. Structural Change and Economic Dynamics. 4, 81-104.

[3] Asch, S. E. 1953. Effects of Group Pressure Upon the Modification and Distortion of Judgments. D. Cartwright and A. Zander, Group Dynamics: Research and Theory. Row, Peterson and Co. 151-62.

[4] Bala, V., Goyal, S. 1998. Learning from Neighbors. Review of Economic Studies. 65, 595-621.

[5] Banerjee, A. V. 1992. A Simple Model of Herd Behavior. The Quarterly Journal of Economics. 107, 797-817.

[6] Bernheim, B. D. 1994. A Theory of Conformity. Journal of Political Economy. 102, 841-77.

[7] Bicchieri, C., Fukui, Y. 1999. The Great Illusion: Ignorance, Informational Cascades and the Persistence of Unpopular Norms. Business Ethics Quarterly. 9, 27-55.

[8] Bikhchandani, S., Hirshleifer, D., Welch, I. 1998. Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades. The Journal of Economic Perspectives. 12, 151-70.

[9] ____. 1992. A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades. Journal of Political Economy. 100, 992-1026.

[10] Blume, L., Durlauf, S. N. 2003. Equilibrium Concepts for Social Interaction Models. International Game Theory Review. 5, 193-209.

[11] Bond, R., Smith, P. B. 1996. Culture and Conformity: A Meta-Analysis of Studies Using Asch's (1952b, 1956) Line Judgment Task. Psychological Bulletin. 119, 111-37.

[12] Bramoullé, Y. 2005. Anti-coordination and Social Interactions. Games and Economic Behavior, forthcoming.

[30] Note that, if we wanted to impose that the weights must be between 0 and 1, then we would simply have to construct the weights as $w_{ij} = \frac{w^{j+1}}{w^{n+1}}$ and the same proof would hold.

[13] Bramoullé, Y. López-Pintado, D., Goyal, S., Vega-Redondo, F.2004. Network Formation in Anti-coordination Games. International Journal of Game Theory. 33, 1-19.

[14] Bramoullé, Y., Kranton, R. 2005. Strategic Experimentation in Networks. Mimeo, Université Laval.

[15] Brock, W. A., Durlauf, S. N. 2001. Discrete Choice with Social Interactions. Review of Economic Studies. 68, 235-60.

[16] Cialdini, R. B. 2001. Influence: Science and Practice. Allyn and Bacon, Needham Heights, MA.

[17] Chwe, Michael. 1999. Structure and Strategy in Collective Action. The American Journal of Sociology, 105, 128-56.

[18] Dodds, P. S., Watts, D. J. 2005. A Generalized Model of Social and Biological Contagion. Journal of Theoretical Biology, 232, 587-604.

[19] ____. Universal Behavior in a Generalized Model of Contagion. 2004. Physical Review Letters. 92, 218-221.

[20] Ellison, G. 1993. Learning, Local Interaction and Coordination. Econometrica, 61, 1047-71.

[21] Ferguson, N. M., Keeling, M. J., Edmunds, W. J., Gani, R., Grenfell, B. T., Anderson R., Leach, S. 2003. Planning for Smallpox Outbreaks. Nature, , 425, 681-85.

[22] Festinger, L., Schachter, S., Back, K. 1950. Social Pressures in Informal Groups. New York: Harper.

[23] Gigerenzer, G., Todd, Peter M., and ABC Research Group. 1999. Simple Heuristics That Make Us Smart. Oxford University Press, New York.

[24] Glaeser, E. L., Sacerdote, B., Scheinkman, J. A. 1996. Crime and Social Interaction. Quarterly Journal of Economics,111, 507-48.

[25] Glance, N. S., Huberman, B. A. 1993. The Outbreak of Cooperation. Journal of Mathematical Sociology, 17, 281-302.

[26] Goldstein, D. G., Gigerenzer, G. 2002. Models of Ecological Rationality: The Recognition Heuristic. Psychological Review, 109:75-90.

[27] Goyal, S., Vega-Redondo, F. 2005. Network Formation and Social Coordination. Games and Economic Behavior. 50, 178-207.

[28] Granovetter, M. S. 1978. Threshold Models of Collective Behavior. American Journal of Sociology. 83, 1420-43.

[29] Granovetter, M. S., Soong, R. 1986. Threshold Models of Interpersonal Effects in Consumer Demand. Journal of Economic Behavior & Organization. 7, 83-99.

[30] ____ .1983. Threshold Models of Diffusion and Collective Behavior. Journal of Mathematical Sociology. 9, 165-79.

[31] Heal, G., Kunreuther, H. 2003. You Only Die Once: Managing Discrete Interdependent Risks. NBER Working Paper w9885. National Bureau of Economic Research.

[32] Jackson, M., Watts, A. 2002. On the Formation of Interaction Networks in Social Coordination Games. Games and Economic Behavior, 41, 265-91.

[33] Jensen, R. 1982. Adoption and Diffusion of an Innovation of Uncertain Profitability. Journal of Economic Theory. 21, 182-93.

[34] Kahan, D. M. 1997. Social Influence, Social Meaning, and Deterrance. Virginia Law Review. 83, 349-95.

[35] Kandori, M., Mailath, J., Rob, R. 1993. Learing, Mutation, and Long Run Equilibria in Games. Econometrica. 61, 29-56.

[36] Katz, M. L., Shapiro, C. 1985. Network Externalities, Competition, and Compatibility. American Economic Review, 75, 424-40.

[37] Kim, H., Bearman, P. S. 1997. The Structure and Dynamics of Movement Participation. American Sociological Review. 62, 70-93.

[38] Kuran, T. 1991. Now out of Never: The Element of Surprise in the East European Revolution of 1989. World Politics. 44, 7-48.

[39] Leibenstein, H. Bandwagon, Snob, and Veblen. 1950. Effects in the Theory of Consumers' Demand. Quarterly Journal of Economics. 64, 183-207.

[40] Leskovec, J., Adamic, L. A., Huberman, B. 2005. The Dynamics of Viral Marketing. Technical Report HP Labs, Hewlett-Packard.

[41] Liebowitz, S. J., Margolis, S. E. 1998. Network Effects and Externalities. The New Palgrave's Dictionary of Economics and the Law. MacMillan, New York, pp. 671-75.

[42] ____. 1994. Network Externality: An Uncommon Tragedy. Journal of Economic Perspectives. 8, 133-50.

[43] Lohmann, S. 1994. The Dynamics of Informational Cascades: The Monday Demonstrations in Leipzig, East Germany, 1989-91. World Politics, 47, 42-101.

[44] López-Pintado. 2004. Diffusion in Complex Social Networks. IVIE Working Paper.

[45] ____. 2005. Contagion and Coordination in Random Networks. Mimeo. Columbia University.

[46] Milgram, S. 1969. Obedience to Authority. Harper and Row. New York.

[47] Milgrom, P. and Roberts, J. 1990. Rationalizability, Learning, and Equilibrium in Games with Strategic Complementarities. Econometrica. 58, 1255-1277.

[48] Morris, S. 2000. Contagion. Review of Economic Studies. 67, 57-78.

[49] Oliver, P. E. 1980. Rewards and Punishments as Selective Incentives for Collective action: Thoretical Investigations. American Journal of Sociology. 85, 1356-75.

[50] ____. 2001. Whatever Happened to Critical Mass Theory? A Retrospective and Assessment. Sociological Theory. 19, 292-311.

[51] Oliver, P. E. and Marwell, G. A. 1985. Theory of the Critical Mass. I. Interdependence, Group Heterogeneity, and the Production of Collective Action. American Journal of Sociology, 91: 522-56.

[52] Ostrom, E., Burger, Joanna, F., Christopher B., Norgaard, R. B. and Policansky, D. 1999. Revisiting the Commons: Local Lessons, Global Challenges. Science. 284, 278-82.

[53] Pastor-Satorrás, R., Vespignani, A. 2001. Epidemic Spreading in Scale-free Networks. Physical Review Letters. 86, 3200-3203.

[54] Peterhansel, A. and Watts, D. J. 2005. A Social Planning Approach to Coordination Failure. Mimeo. Columbia University.

[55] Rae, J. 1905. The Sociological Theory of Capital. The Macmillan Company, London.

[56] Schelling, T. C. 1971. Dynamic Models of Segregation. Journal of Mathematical Sociology, 1, 143-86.

[57] ____. 1973. Hockey Helmets, Concealed Weapons, and Daylight Saving: A Study of Binary Choices with Externalities. Journal of Conflict Resolution, 17, 381-428.

[58] Shiller, R. J. 2000. Irrational Exuberance. Princeton University Press, Princeton, N.J.

[59] Simmel, G. 1957. Fashion. American Journal of Sociology, 62, 541-58.

[60] Strang, D. and Soule, S. A. 1998. Diffusion in Organizations and Social Movements: From Hybrid Corn to Poison Pills. Annual Review of Sociology. 24, 265-90.

[61] Valente, T. W. 1995. Network Models of the Diffusion of Innovations. Hampton Press, Cresskill, NJ.

[62] Veblen, T. 1912. The Theory of the Leisure Class: An Economic Study of Institutions. The Macmillan Company, New York.

[63] Watts, D. J. 2002. A Simple Model of Information Cascades on Random Networks. Proceedings of the National Academy of Science. 99, 5766-71.

[64] Peterhansel, A., Watts, D. J. 2005. A Social Planning Approach to Coordination Failure. Mimeo. Columbia University, .

[65] Winter, E. Incentives and Discrimination. American Economic Review, forthcoming.

[66] Young, H. P., Burke, M. A. 2001. Competition and Custom in Economic Contracts: A Case Study of Illinois Agriculture. American Economic Review. 91, 559-73.

[67] Young, H. P. 1993. The Evolution of Conventions. Econometrica. 61, 57-84.

[68] Young, H. Peyton. 1998. Individual Strategy and Social Structure: An Evolutionary Theory of Institutions. Princeton University Press. Princeton, N.J.

[69] _ _ _ _. 2004. Strategic Learning and its Limits. Arne Ryde Memorial Lectures. Oxford University Press, Oxford.

[70] _ _ _ _. 2005. The Spread of Innovations through Social Learning. Mimeo. John Hopkins University and Oxford University

[71] _ _ _ _. 2005. The Spread of Innovations within Groups. Mimeo. John Hopkins University and Oxford University.

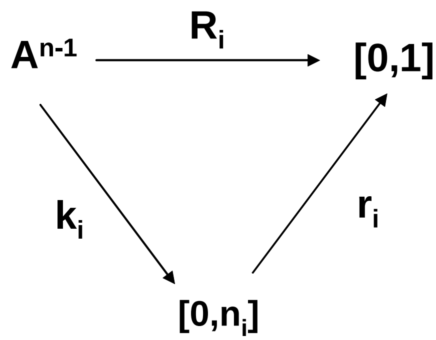

Figure 1: Representation of the function R_i as a composition of the functions k_i and r_i.

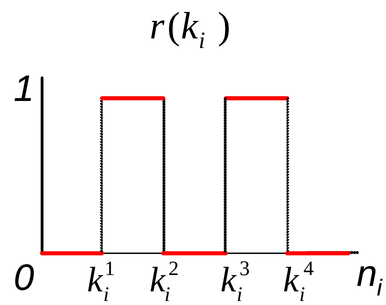

Figure 2: An influence-response function $r_i(k_i)$ with four thresholds.

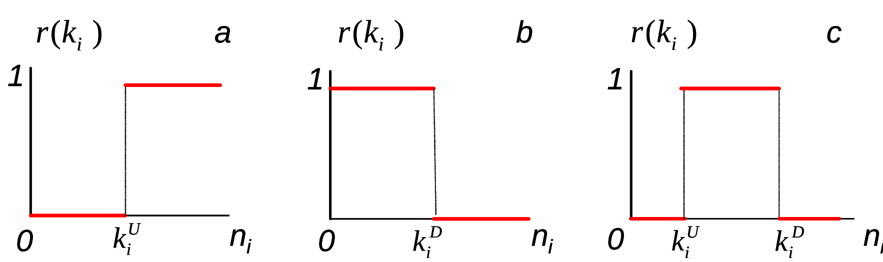

Figure 3: Different kinds of influence-response functions: (a) upward threshold influence-response function (b) downward threshold influence-response function, and (c) upward-downward threshold influence-response function.

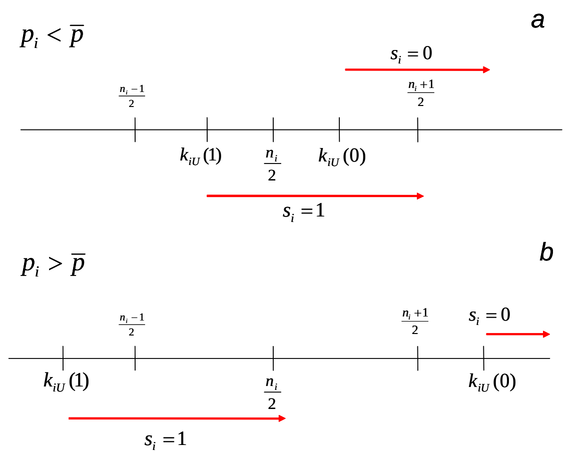

Figure 4: The upward threshold when the signal is either 1 ($k_{iU}(1)$) or 0 ($k_{iU}(0)$) and the average accuracy is (a) above individual i's accuracy $p_i < \overline{p}$ or (b) below individual i's accuracy $p_i > \overline{p}$.

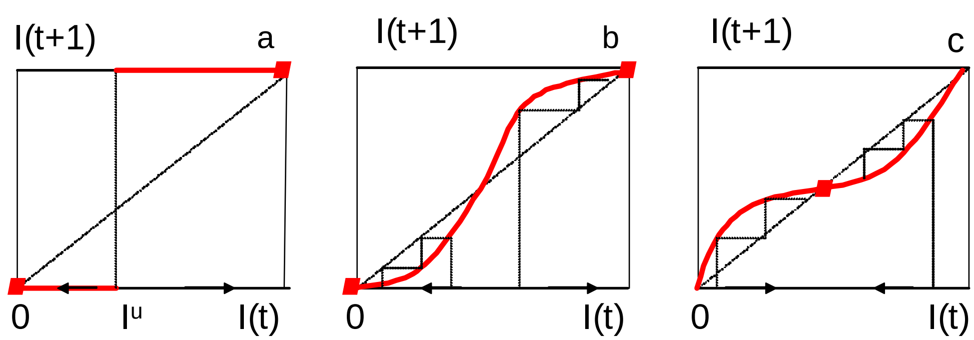

Figure 5: $I(t+1) = F(I(t))$ where $F(I(t))$ is the cdf of (a) a degenerated distribution of thresholds at I^u (b) a symmetric Beta distribution of thresholds $B(\mu, \mu)$ where $\mu > 1$ (c) a symmetric Beta distribution of thresholds $B(\mu, \mu)$ where $\mu < 1$.

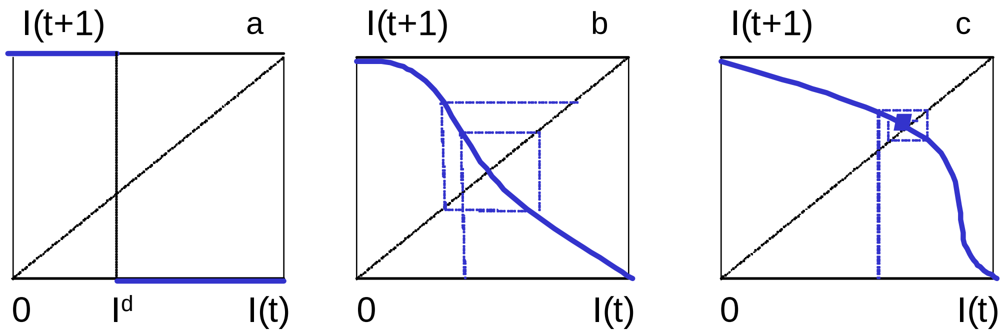

Figure 6: $I(t+1) = F(I(t))$ where $F(q^d)$ is the cdf of the distribution of downward thresholds in the population. In figure (a) the distribution is degenerated at I^d and therefore the dynamics cycles between I^0 and I^1. In figure (b) the dynamics is cahotic because $F'(I^*) < -1$, where I^* is the stationary state, i.e. $F(I^*) = I^*$. In figure (c) the dynamics converges to I^* because $F'(I^*) > -1$.

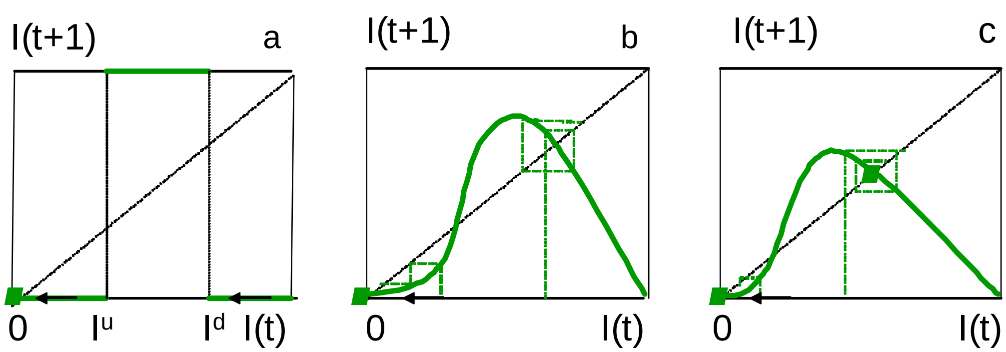

Figure 7: $I(t+1) = F_u(I(t)) - F_d(I(t))$ where $F_u(q^u)$ and $F_d(q^d)$ are the marginal distributions of the upward and downward threshold distributions in the population. In figure (a) the distribution of thresholds is degenerated and thus all individuals in the population have two thresholds I^u and I^d. Figure (b) represents the case where there are two stationary states I^0 and I^* such that $0 < I^* < 1$. The state I^0 is stable whereas I^* is unstable since $F'_u(0) - F'_d(0) < 1$ and $F'_u(I^*) - F'_d(I^*) < -1$. Figure (b) represents the case where there are two stationary states I^0 and I^* such that $0 < I^* < 1$. Both states are stable since $F'_u(0) - F'_d(0) < 1$ and $F'_u(I^*) - F'_d(I^*) > -1$.

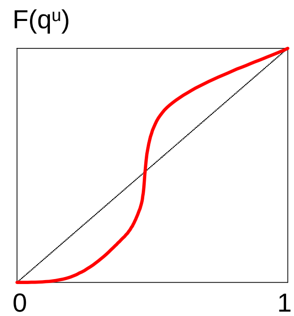

Figure 8: Distribution of upward thresholds in the population $F(q^u)$.

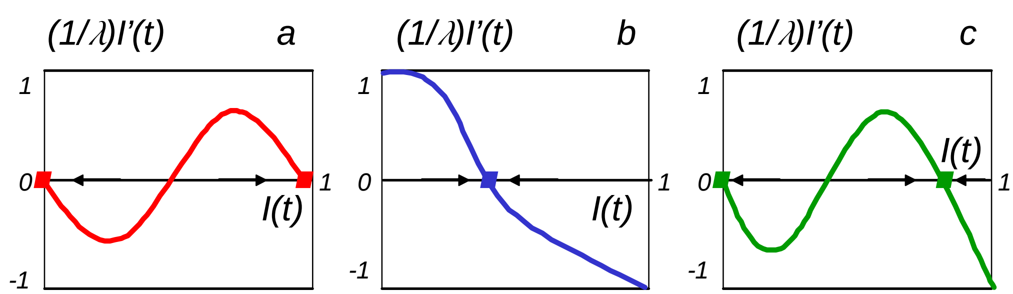

Figure 9: $I'(t)$ as a function of $I(t)$. Figure (a) represents a situation where the thresholds are upward thresholds and distributed in the population according to a cdf function with the same qualitative features as $F(q^u)$ (see figure 10). Figure (b) represents a situation where the thresholds are downward thresholds. Figure (c) represents a situation where individuals have two thresholds, an upward and a downward threshold. In all figures the arrows indicate the direction of the dynamics when $I(t)$ lies in the corresponding regions.