INTER-FACULTY MASTER PROGRAM
on NETWORKS and COMPLEXITY
SCHOOL of MATHEMATICS
SCHOOL of BIOLOGY
SCHOOL of GEOLOGY
SCHOOL of ECONOMICS
ARISTOTLE UNIVERSITY of THESSALONIKI

Master Thesis

Title:

A general model of social influence on opinion formation, from the agent's point of view

Ένα γενικό μοντέλο κοινωνικής επιρροής στη διαμόρφωση άποψης, από την οπτική του πράκτορα

Antonios T. Syreloglou

SUPERVISOR: Stefanos Sgardelis, Professor, AUTH

Thessaloniki , October 2020

ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
στα ΔΙΚΤΥΑ και ΠΟΛΥΠΛΟΚΟΤΗΤΑ
ΤΜΗΜΑ ΟΙΚΟΝΟΜΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ
ΤΜΗΜΑ ΒΙΟΛΟΓΙΑΣ
ΤΜΗΜΑ ΓΕΩΛΟΓΙΑΣ
ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Τίτλος Εργασίας

Ένα γενικό μοντέλο κοινωνικής επιρροής στη διαμόρφωση άποψης, από την οπτική του πράκτορα

A general model of social influence on opinion formation, from the agent's point of view

Αντώνιος Τ. Συρέλογλου

ΕΠΙΒΛΕΠΩΝ: Στέφανος Σγαρδέλης, Καθηγητής, Α.Π.Θ.

Εγκρίθηκε από την Τριμελή Εξεταστική Επιτροπή την 23η Οκτωβρίου 2020.

……	……	……
Σ. Σγαρδέλης	Αλεξάνδρα Μπεκιάρη	Νικόλαος Χασάναγας
Καθηγητής Α.Π.Θ.	Επίκουρη Καθηγήτρια Π.Θ.	Ερευνητής Α.Π.Θ.

Θεσσαλονίκη , Οκτώβριος 2020

…….

Αντώνιος Τ. Συρέλογλου
Πτυχιούχος Μαθηματικός Α.Π.Θ.

ABSTRACT

Social influence on opinion formation has attracted a large amount of research interest over the years, being one of the most important social phenomena. In this paper, we propose a general model that describes this process from an agent's point of view, where the agent under consideration forms opinion on a topic. The model is based on a philosophical concept attributed to Leibniz, the alphabet of human thoughts. In particular, we use this concept to theoretically analyze the two fundamental aspects of the phenomenon, namely opinion and influence. This enables us to naturally construct the key parameters of the model. Also, unlike many literature models of opinion formation under social influence, where the process is studied given a social network of interacting agents, our model does not distinguish between influence coming from within a network and influence coming from outside a network. Instead, all sources of social influence are treated the same. In order to provide a framework for the potentially vast number of such sources and make the model more practicable, we formulate an assumption by combining results from research in two fields, namely research on the size and structure of an individual's social network and research on political socialization. The overall approach to constructing the model motivates us to have a brief discussion about opinion and knowledge within a society and to suggest a connection between opinion networks and knowledge networks. Finally, we evaluate the model and the soundness of the assumption using data collected from questionnaires. A number of 32 respondents answered questions concerning the importance of the July 2019 national elections in Greece. The respondents were selected to be young adults of ~24 years of age, so that the parental influence on political topics is still significant and the accordingly designed questions could capture it. Conditions for the representativeness of the sample were not required, since we are not interested in making statistical inferences for an entire population but in the effectiveness of the model in predicting each respondent's opinion. For the purposes of the evaluation, we consider two groups of the respondents. The first group consists of all the respondents, while the second group is formed by removing from the first group those respondents whose influences are not adequately captured by the questions in the questionnaire. We report a much greater effectiveness of the model in predicting opinions in the second group, for which the necessary data is provided.

KEY WORDS

Social Influence, Opinion Formation, Modeling, Alphabet of Human Thoughts, Knowledge

ΠΕΡΙΛΗΨΗ

Η κοινωνική επιρροή στη διαμόρφωση άποψης έχει προσελκύσει μεγάλο ερευνητικό ενδιαφέρον ανά τα χρόνια, καθώς αποτελεί ένα από τα πιο σημαντικά κοινωνικά φαινόμενα. Στο πλαίσιο της παρούσας εργασίας, προτείνουμε ένα γενικό μοντέλο που περιγράφει αυτήν τη διαδικασία από την οπτική ενός πράκτορα, όπου ο υπό εξέταση πράκτορας διαμορφώνει άποψη για ένα θέμα. Το μοντέλο βασίζεται σε μια φιλοσοφική έννοια που αποδίδεται στον Leibniz, το αλφάβητο των ανθρώπινων σκέψεων. Συγκεκριμένα, χρησιμοποιούμε αυτήν την έννοια για να αναλύσουμε θεωρητικά τις δύο θεμελιώδεις πτυχές του φαινομένου, δηλαδή την άποψη και την επιρροή. Αυτό μας επιτρέπει να κατασκευάσουμε φυσικά τις βασικές παραμέτρους του μοντέλου. Επίσης, σε αντίθεση με πολλά βιβλιογραφικά μοντέλα διαμόρφωσης άποψης υπό το καθεστώς κοινωνικής επιρροής, όπου η διαδικασία μελετάται δοθέντος ενός κοινωνικού δικτύου πρακτόρων που αλληλεπιδρούν, το μοντέλο μας δεν ξεχωρίζει την επιρροή που προέρχεται μέσα από ένα δίκτυο από την επιρροή που προέρχεται από έξω. Αντίθετα, όλες οι πηγές κοινωνικής επιρροής αντιμετωπίζονται με κοινό τρόπο. Προκειμένου να παρέχουμε ένα πλαίσιο για τον δυνητικά τεράστιο αριθμό τέτοιων πηγών και να κάνουμε το μοντέλο πιο πρακτικό, διατυπώνουμε μια υπόθεση συνδυάζοντας αποτελέσματα από έρευνες σε δύο τομείς, συγκεκριμένα έρευνες σχετικά με το μέγεθος και τη δομή του κοινωνικού δικτύου ενός ατόμου και έρευνες για την πολιτική κοινωνικοποίηση. Η συνολική προσέγγιση για την κατασκευή του μοντέλου μάς παρακινεί να κάνουμε μια σύντομη συζήτηση σχετικά με τη γνώμη (άποψη) και τη γνώση μέσα σε μια κοινωνία και να προτείνουμε μια σύνδεση μεταξύ των δικτύων γνώμης και γνώσης. Τέλος, αξιολογούμε το μοντέλο και την ορθότητα της υπόθεσης χρησιμοποιώντας δεδομένα που συλλέγονται από ερωτηματολόγια. Ένας αριθμός 32 ερωτηθέντων απάντησαν σε ερωτήσεις σχετικά με τη σημασία των εθνικών εκλογών του Ιουλίου 2019 στην Ελλάδα. Οι ερωτηθέντες που επιλέχθηκαν ήταν νεαροί ενήλικες ηλικίας ~24 ετών, έτσι ώστε να εξακολουθεί να είναι σημαντική η γονική επιρροή σε πολιτικά θέματα και οι κατάλληλα διαμορφωμένες ερωτήσεις να μπορούν να την συλλάβουν. Δεν απαιτήθηκαν προϋποθέσεις για την αντιπροσωπευτικότητα του δείγματος, καθώς δεν ενδιαφερόμαστε για στατιστικά συμπεράσματα που αφορούν σε έναν ολόκληρο πληθυσμό αλλά για την αποτελεσματικότητα του μοντέλου στην πρόβλεψη της άποψης του κάθε ερωτώμενου. Για τους σκοπούς της αξιολόγησης, θεωρούμε δύο ομάδες των ερωτηθέντων. Η πρώτη ομάδα αποτελείται από όλους τους ερωτηθέντες, ενώ η δεύτερη ομάδα σχηματίζεται αφαιρώντας από την πρώτη ομάδα τους ερωτηθέντες για τους οποίους οι επιρροές που δέχονται δεν εντοπίζονται επαρκώς από τις ερωτήσεις του ερωτηματολογίου. Τα ευρήματα δείχνουν πολύ μεγαλύτερη αποτελεσματικότητα του μοντέλου στην πρόβλεψη απόψεων στη δεύτερη ομάδα, για την οποία παρέχονται τα απαραίτητα δεδομένα.

CONTENTS

Chapter ***Page***

1 Introduction……..…….. 9

1.1 Brief literature review…….……9

1.2 The aim of the paper…….……..10

1.3 The added value of the model.……...11

2 The alphabet of human thoughts……...…….12

3 Modeling from the agent's point of view……...16

3.1 Constructing the general model……. 16

3.2 A framework for the agent's social set…….. ……21

3.3 Opinion and knowledge within a group…….27

4 Evaluation of the model ……29

4.1 Method……. …….29

4.2 Results…….…….. 31

4.3 Discussion…….…….33

5 Conclusions……34

6 Appendix……35

7 References……..38

1 Introduction

The process of opinion formation has always been an integral part of humans' behavior in societies. Given a certain topic, the way an individual forms an opinion is inextricably linked with social interactions and, thus, determined by social influence. Discovering the mechanisms of social influence is a key step towards understanding individual and collective opinions being formed in a society.

Let us discuss the concept of opinion at an elemental level. The term opinion is used as a synonym for belief, judgment, view or even idea, where the term idea captures the meaning in a broader sense. An indirect but very helpful way of considering it is through the distinction between opinion and knowledge. Ancient Greek philosopher Plato described knowledge as "*justified, true belief*" or, as stated in [1], "*knowledge requires opinions to be true and moreover justified*". Therefore, knowledge may be viewed as a subset of the set of opinions. Finally, opinions are closely related to behaviors. As given in [2], "*the expression of an opinion represents a behavior*".

Opinion formation takes place in a social environment and is influenced by it. Let us assume that a member of a society holds a specific opinion on a certain topic at a given moment in time. Before the member reaches this state, there is a process of formation that the opinion goes through during which social influence takes place. Such influence may be coming from a potentially large number of sources. For example, a future voter forming an opinion on a political issue may be influenced by: *parents and other family members, friends, colleagues, political figures, scientists, celebrities, tv personalities, newspaper articles, internet posts, favorite authors, fictional characters in books, movies or tv series*[1]*, historical figures, teachers dating back to school or university years, memorable strangers involved in otherwise random everyday events and more.* All these sources – whether they are persons (living or deceased) or non-human entities, real or fictional, close or distant, liked/admired or disliked/detested – have social influence on opinion formation (sources of social influence).

1.1 *Brief literature review*

Many approaches to modeling opinion formation and dynamics (evolution of opinion over time) under social influence have been proposed in the literature. For a survey of the field, the reader may see [3]. Here, we provide a brief presentation of certain models. The classic *French-DeGroot* model [4] describes a mechanism of opinion formation – in particular, consensus formation – among the n agents of a group in discrete time. For an agent i $(1 \leq i \leq n)$, the agent's opinion at time t is denoted by $x_i(t) \in \mathbb{R}$ and the weight given to any other agent j is denoted by a_{ij}. So, agent i's opinion at time $t+1$ is formed as follows

$$x_i(t+1) = a_{i1}x_1(t) + a_{i2}x_2(t) + \cdots + a_{in}x_n(t)\ ,$$

[1] American politician Joe Biden said, in an interview with NBC's "Meet the Press", that the television show "Will & Grace" increased awareness and acceptance of homosexuality in the American public opinion. He said "*I think 'Will and Grace' probably did more to educate the American public than almost anything anybody's ever done so far*". In this case, the characters (Will, Grace or others) or the show itself would be sources of social influence.

which is a weighted average of the agents' opinions at time t.

A generalization of this model is the *Friedkin-Johnsen* model [5], which captures persistent disagreement in addition to consensus; it is one of the few models of opinion formation that has been experimentally validated for groups of small and medium size [6]. The (linear) *Abelson* model [7] is a continuous-time counterpart of the French-DeGroot model and is extended by the *Taylor* model [8], which additionally considers sources of fixed opinions that influence the agents' opinions.

The models mentioned so far are linear and thus, linear methods and tools (such as Markov chains or matrix theory) can be used to promote the analysis. On the other hand, the *bounded confidence* models [9, 10] are non-linear; the agents influence each other only if their opinions are close enough (as defined by a threshold). As a result of the non-linearity aspect and the restrictions imposed (since, for example, Markov chains cannot be applied here), the analysis of these models is heavily based on computer simulations. Finally, in all the previously mentioned models, the agents' opinions are represented by continuous values. An example (of a model) which uses discrete values for the opinions is the *voter* model, independently introduced by Clifford, Sudbury [11] and by Holley, Liggett [12]. In a time-step of the model, if an agent's opinion is to be updated, this is done by adopting the opinion of a randomly chosen neighbor, as defined by the agent's position in the network under consideration.

1.2 *The aim of the paper*

In general, models of opinion formation under social influence involve weights, parameters or thresholds in order to quantify certain aspects that determine the agents' behavior; aspects like interpersonal and other influences, susceptibilities to interpersonal influence, confidence levels or convergence. In this paper, we propose a general model that predicts the agent's opinion on a topic; a model in which the parameters are naturally constructed. Let us explain what we mean by "naturally constructed". The theory we develop, on which the model is based, approaches the process of opinion formation (under social influence) from the point of view of the agent that forms opinion. In that sense, one aim of the theory is to capture the agent's internal mechanisms that determine how social influence affects the opinion being formed. In the context of this theory, the parameters involved are constructed (not just defined) in a way that their meanings emerge naturally. Of course, we do not know the exact mechanisms which take place "inside" an agent and concern opinion formation. As the authors in [13] write: "*The random assignments of threshold account for the lack of knowledge of intrinsic latent tendencies of nodes to adopt neighbor strategies*" (referring to the threshold involved in the *Linear Threshold* model [14], which is a diffusion model for the spread of an idea through a social network). However, given certain assumptions, we believe that it is possible to introduce such a theory, from the agent's point of view.

So, we construct a model of social influence on opinion formation, based on a theory that enables two things. First – as we mentioned – the parameters are naturally constructed. This provides a better understanding of what these parameters contribute to the model and therefore, a possibly more flexible design of simulations. Second – as will be further discussed in the next two paragraphs – a framework is provided for the agent's social set to be captured with no need to explicitly represent it. Let us note that by the agent's social set, we mean the set of all the sources that socially influence the

agent's opinion formation, as in the example of the future voter mentioned above. This will be mainly appreciated when dealing with real data.

Given the literature models we talked about, a model of opinion formation typically considers a social network of interacting agents and, in that way, certain tools derived from graph/network theory are unlocked. Enclosing the process of opinion formation within such a visualized network of agents is an approach that can be interpreted as external, taken from the modeler's point of view. It is an approach that provides a "tangible" network to work with. For the impact of the social network on opinion formation the reader may see [15], while for the (mathematical) study of social networks, we suggest [16].

On the other hand, our model considers not a social network of interacting agents, but an agent's social set as we defined it, which is taken from the agent's point of view. We do not distinguish between influence coming from within a network and influence coming from outside a network. Instead, all sources of social influence are treated the same. Given a topic, we attempt to capture this social set in depth, taking under consideration the potentially large number of sources that socially influence the agent's opinion on the topic (the reader, again, may see the future-voter example). Of course, identifying all these sources is an insurmountable task. That is why the proposed theory provides a framework for the agent's social set to be captured without describing it explicitly; in that way, the model becomes more practicable. In particular, we formulate an assumption by combining results from research in two fields, namely research on size and structure of an individual's social network and research on political socialization.

To summarize, the overall aim of this paper is to establish a novel way of thinking about and approaching the phenomenon of social influence on opinion formation. We are more interested in studying opinion formation in its "free" form and less interested in doing so under controlled conditions (*e.g.* given a social network structure or population of interacting agents). This is reflected in the fact that the model considers the agent's social set (theoretically, all the sources that socially influence the agent's opinion formation).

1.3 *The added value of the model*

There are three remarks to be made as far as the added value of the proposed model is concerned:

1) The theory is inspired by *the alphabet of human thoughts*, a philosophical concept from Gottfried Leibniz. We use a version of this concept as a theoretical base on which we build our model. In particular, we define concepts like simple idea and simple opinion and use them to propose an analysis-decomposition of the fundamental nature of opinion and influence. This analysis-decomposition makes a more detailed description of the phenomenon possible. For example, we are able to distinguish between the magnitude of a source's influence on the agent (how strong the influence is on the agent) and the effect of a source's influence on the agent (whether the influence makes the agent receptive to the source's opinion or rejective of it).

2) Given an agent that forms opinion on a topic, a basic parameter (of the proposed model) represents the extent to which the agent is informed about the topic. By using this parameter, we take into account the amount of total information collected by the agent about the topic, both correct and false information; an agent may be correctly informed or misinformed/disinformed, on the way to forming an opinion.

Unlike the literature models we talked about, which study how opinion evolves over time, we promote a different kind of opinion dynamics based on this basic parameter. In particular, as the extent to which the agent is informed about the topic increases (new information is received by the agent) or decreases (existing information is forgotten by the agent), the agent's opinion evolves.

3) We discuss the concepts of opinion and knowledge from a sociological perspective and how they are perceived within a group of agents (society). We also propose a connection between opinion networks (where opinions are formed and spread) and knowledge networks (where knowledge is communicated).

The paper is organized as follows. In chapter 2, we present Leibniz's concept (the alphabet of human thoughts) and use it in order to construct the theoretical basis of our model. In chapter 3, we describe the model of social influence on opinion formation as it approaches the problem from the agent's point of view and we provide a framework for the model to be applied in a more practical way, under a certain assumption with which we attempt to capture the agent's social set in a way consistent with reality. We also describe the way opinion (networks) and knowledge (networks) are connected. In chapter 4, we perform an evaluation of the model using data collected from questionnaires, concerning opinions on the importance of the July 2019 national elections in Greece and answered by a number of 32 respondents (young adults of ~24 years of age). We report and discuss the results of the evaluation.

2 The alphabet of human thoughts

Gottfried Leibniz [1646-1716] was one of the most important mathematicians and philosophers of the Enlightenment and a true polymath. A lifelong interest and dream of his was to create a universal language ("*characteristic*") that would express all human knowledge and that would be a calculus for reasoning at the same time [17, 18]. As Couturat wrote, Leibniz describes that this language "*would express the composition of concepts by the combination of signs representing their simple elements, such that the correspondence between composite ideas and their symbols would be natural and no longer conventional*" [17].

Let us consider an example to further understand what is meant by natural and conventional. The word man is written as a combination of signs $\{a, m, n\}$, in a conventional way. On the other hand, Leibniz wrote that "*since man is a rational animal, if the number of animal is a, for instance, 2, and the number of rational is r, for instance, 3, the number of man, or h, will be 2x3 or 6*" [19], where the concept man would be expressed as a combination of signs $\{2, 3\}$ representing the simple elements of man (rational and animal). In this natural way, the logical (inter)relations of concepts would be captured and thus, reasoning would be reduced to calculation. Of course, this example (with numbers as signs and the above concepts as simple elements) was just an analogy. For the universal language, Leibniz probably imagined it to involve points and lines (forming geometrical figures) combined with a kind of pictures[2] [20]. For more on

[2] As Leibniz writes [20] "*... just as the ancient Egyptians did, and the Chinese do today. Their pictures, however, are not reduced to a fixed alphabet... with the result that a tremendous strain on the memory is necessary, which is the contrary of what we propose*". Leibniz also mentioned chemical and astronomical signs as examples.

Leibniz's program, his influences, the work of others and the universal language today, the reader may see [19, 21, 22].

So, all complex concepts are expressed through different combinations of simple concepts. Leibniz considered these simple concepts to be (innate) ideas deeply rooted in every human's mind, ultimately simple and universally understood; he called them the alphabet of human thoughts[3]. He wrote [23]: "*Although the number of ideas which can be conceived is infinite, it is possible that the number of those which can be conceived by themselves is very small; because an infinite number of anything can be expressed by combining very few elements... The alphabet of human thoughts is the catalogue of those concepts which can be understood by themselves, and by whose combination all our other ideas are formed*". For the above and more (concerning the alphabet of human thoughts) and for the modern research on the related semantic primitives, the reader can see [24].

To summarize, a complex idea is composed of simple ideas, which are elements of the alphabet of human thoughts. This is all we need for a theoretical basis on which our model will be constructed. We say "theoretical" because a practical alphabet of human thoughts does not exist (at least not yet). And we say "all we need" because certain aspects of the alphabet (as these were suggested by Leibniz), other than the composition of complex ideas, will be of no use in the construction of our model. For example, no specific signs are involved in our theory for the representation of simple ideas. Moreover, the requirement that the number of elements in the alphabet of human thoughts (the number of simple ideas) should be limited is not necessary; being limited provides no added value to our model, but it did to Leibniz's project. Indeed, having a large number of simple ideas (or even infinite) is much more natural and appropriate for our model, as this will be appreciated in the discussion that follows.

Having said all that, the analogy which would be helpful to keep in mind throughout this paper is the one that Leibniz gave and we mentioned above. This analogy incorporates the key aspects of the alphabet of human thoughts according to our needs. Let us be more specific. By the fundamental theorem of arithmetic [25], every natural number greater than 1 can be represented in a unique way as a product of one or more primes (the representation is unique, except for the order of the factors). Therefore, prime numbers, which are infinite, can be viewed as the "*building blocks*" of the natural numbers [26]. For example, it is $1176 = 2^3 \times 3 \times 7^2$. If the number 1176 represented a complex concept c and the primes $2, 3, 7$ represented three distinct simple ideas x, y, z of the alphabet of human thoughts, then concept c would be composed of $\{x, x, x, y, z, z\}$, which is a multiset (and not the set $\{x, y, z\}$, because then information concerning the exact composition would be lost).

We can begin to construct the framework of our model. Let us consider an alphabet of human thoughts, in the way we described it using the analogy with primes; this alphabet is denoted by A. So, A is the set of simple ideas. Since an opinion is an idea (fig.1), we can talk about complex opinions, as we do about complex ideas. A complex opinion c is composed of simple ideas $s_j \in A$, where $j \in \mathbb{N}$; in general, these s_j form a multiset which is denoted by A_c. For example, if $c = 1176$ then we have $A_{1176} = \{2, 2, 2, 3, 7, 7\}$, where $s_1 = s_2 = s_3 = 2$, $s_4 = 3$ and $s_5 = s_6 = 7$ (we should

[3] The idea of an inventory of innate concepts originated in the 17th century. Rene Descartes [1596-1650] suggested that such concepts must be clear (by themselves) and indefinable. Leibniz added a third, crucial property that the simple concepts in the alphabet of human thoughts should satisfy; these concepts should serve as building blocks in the definition of other (complex) concepts.

note again that the expression $c = 1176$ makes sense only within the context of the analogy with primes).

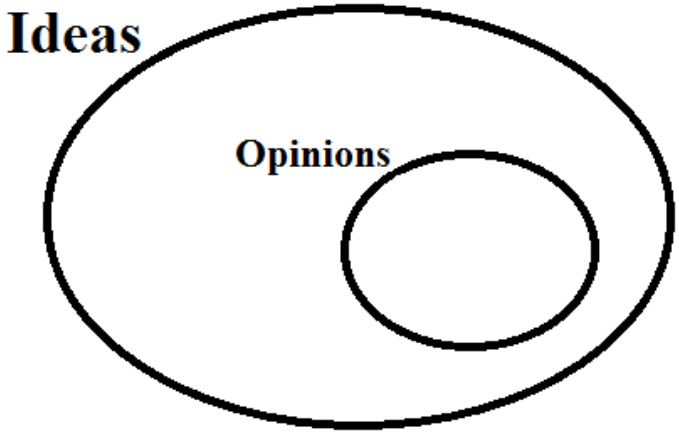

Fig.1. We assume that the set of opinions is contained in the set of ideas. An opinion can be associated with the degree to which an agent or a source agrees or disagrees with it.

Now, let us consider an agent that forms opinion on a topic, under social influence. This process of formation works as follows: the topic itself is viewed as an opinion (the topic is an opinion[4]) and the degree to which the agent agrees or disagrees with the topic-opinion is determined by social influence. As we have mentioned in the introduction, this influence comes from the agent's social set, which consists of all the sources that socially influence the agent's opinion. The complexity of the task prevents us from being able to identify all these sources one by one, given an agent and a topic-opinion[5]. But we can still propose two properties that a source of social influence should satisfy. The first property is obvious:

$\boldsymbol{P_1}$) Given an agent and a topic-opinion, a source of social influence must be within the scope of the agent's knowledge.

The agent needs to be aware of the source in order to be influenced by it. For example, *a book may contain useful information about the topic, but as long as the agent is not aware of the book, no influence can come from it*.

The second property needs more discussion. Influence is achieved through the information that is communicated-transferred from the source to the agent. For example, *a politician communicates certain political views or the self-image of a family person to a voter, an educational cartoon character offers useful information – in a friendly way – to a child watching it, the memory of a deceased grandparent provides valuable advice*

[4] For example, the general topic of climate change can be expressed in the form of an opinion like "Climate change is real" (or any other equivalent statement) and an agent agrees or disagrees with this topic-opinion to a certain degree.

[5] When we refer to sources of social influence, we must always do it given a particular agent and a particular topic-opinion. Let us explain why. Even for the same topic, different agents typically have different social sets and, therefore, different sources of social influence (for example, they may have been raised in different families or they may admire different historical figures *etc.*). Also, an agent forming opinion on different topics may be influenced by different sources (for example, an agent may be influenced by an article in a car magazine when thinking of buying a car, but not when dealing with a family issue). So, the agent and the topic-opinion must be fixed when referring to sources of social influence.

or principles to be followed. Ultimately, all the information received by the agent from a source is grouped together to form an idea – in the agent's mind – about the source. In general, this idea (about the source) is a complex idea and, therefore, is composed of simple ideas from the alphabet of human thoughts.

Given an agent k and a source i, the idea that the agent forms about the source is called the profile of source i according to agent k and is denoted by $p_{i,k}$. Of course, different agents may form different ideas about the same source, since they may receive different information (due to quality[6] of information and/or quantity of information) from the source. When the agent is fixed and there is no ambiguity, the profile of source i according to the agent will be simply denoted by p_i.

We are now ready to state the second property. Let us remember that if c is a complex idea, then A_c is the multiset of its simple ideas. So:

$\boldsymbol{P_{2a}}$) Given an agent, a topic-opinion c and a source i of social influence with profile p_i (according to the agent), we have

$$\left|A_c \cap A_{p_i}\right| > \vartheta(c) \quad \text{, where } \vartheta(c) \in \mathbb{N}\,. \tag{1}$$

It means that the number of common simple ideas between A_c and A_{p_i} must be greater than the threshold $\vartheta(c)$, which is a function of topic-opinion c. In this way, a source i, that influences the agent on topic-opinion c, communicates to the agent information that is sufficiently relevant to the topic-opinion, since we assume that complex ideas with common simple ideas have meanings relevant-similar to each other[7]. For example, *let us consider an agent that seeks advice (forms opinion) on a problem (topic-opinion) as a parent; we expect that this problem and the information contained in a book on how to be a good parent have more simple ideas in common than the problem and a financial article in a newspaper*.

In general, different topics may require different thresholds and that is why the threshold depends on the topic-opinion (function of c). Whatever the case, $\vartheta(c)$ reflects an inevitable element of arbitrariness. Finally, the second property is complete by considering the following:

$\boldsymbol{P_{2b}}$) Given an agent and a topic-opinion c, a source i that satisfies (1) is a source of social influence.

This means that a source (any source) which provides the agent with information sufficiently relevant to the topic, influences the agent on the topic.

---

[6] By saying that they may receive information different due to quality, we mean scenarios like these: the agents are exposed to seemingly the same piece of information from the same source, but one of the agents receives a distorted version of it, or perceives it in a different way as a result of misunderstanding *etc.*

[7] In order to make comparisons, we must consider comparable numbers of simple ideas. For example, let us consider two pairs of complex ideas, c_1 and c_2 & c_3 and c_4, where $\left|A_{c_1}\right| = \left|A_{c_3}\right|$ and $\left|A_{c_2}\right| = \left|A_{c_4}\right|$ (comparable numbers). Then, if $\left|A_{c_1} \cap A_{c_2}\right| > \left|A_{c_3} \cap A_{c_4}\right|$, we can say that c_1 and c_2 are more relevant to each other than c_3 and c_4 are relevant to each other (since the number of common simple ideas between A_{c_1} and A_{c_2} is larger than the number of common simple ideas between A_{c_3} and A_{c_4}).

3 Modeling from the agent's point of view

3.1 *Constructing the general model*

In this chapter, we propose the model of social influence on opinion formation based on the framework we described in the previous chapter. We should start by expanding on something that we have mentioned briefly so far.

We have said that an opinion (topic-opinion) can be associated with the degree to which an agent or a source agrees or disagrees with it. This degree can be (naturally) expressed by a value ranging from -1 to $+1$. A value of $+1$ represents that the agent or the source totally agrees with the topic-opinion, a value of -1 represents that the agent or the source totally disagrees with the topic-opinion, a value of 0 represents that the agent or the source has a completely neutral stance on the topic or has no opinion on the topic. For values between 0 and $+1$, as the value becomes larger, it represents agreement of greater intensity; for values between -1 and 0, as the absolute value becomes larger, it represents disagreement of greater intensity. So, opinions – unlike ideas in general – can be meaningfully associated with the following property: an agent or a source agrees or disagrees with an opinion to a certain degree expressed by a value $v \in [-1, +1]$.

At this point, let us make two important assumptions that we need for the development of our theory. For these assumptions, we define the concept of simple opinion. A simple idea that is also an opinion will be called simple opinion (fig.2). And as we have already said, a complex opinion is composed of simple ideas. Now, the first assumption is that a complex opinion is composed exclusively of simple opinions. In other words:

$\boldsymbol{H_1}$) If c is a complex opinion, then all elements of A_c are simple opinions.

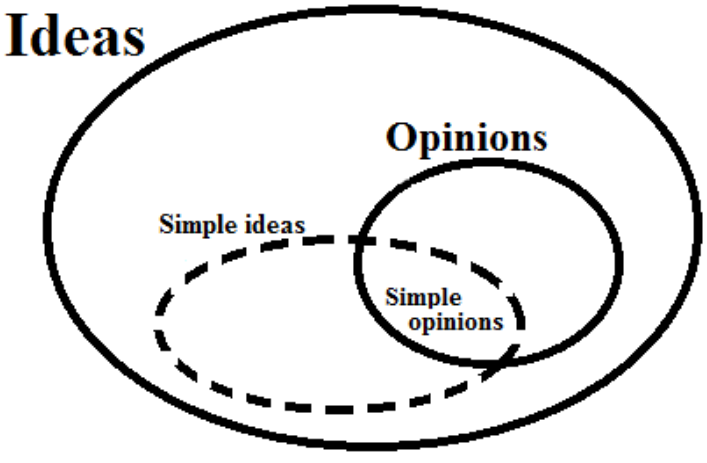

Fig.2. A simple idea that is also an opinion is called simple opinion.

The second assumption is about a distinction between complex and simple opinions, concerning the associated value. A complex opinion can be associated with a value ranging from -1 to $+1$, since such value expresses the degree (to which an agent or a

source agrees or disagrees with the complex opinion). On the other hand, a simple opinion is a simple idea, which means that it is ultimately simple. So, it is natural to assume that given a simple opinion, an agent or a source either totally agrees with it $(+1)$ or totally disagrees with it (-1); there is no room for an intermediate stance, exactly because of the ultimate simplicity of the opinion. Therefore, the assumption can be summarized as follows:

$\boldsymbol{H_2}$) A complex opinion can be associated with a value $v \in [-1, +1]$, whereas a simple opinion can be associated with a value $v' \in \{-1, +1\}$.

Having said all that, we can continue with the process of opinion formation under social influence. Let us consider an agent and a topic-opinion c. Let N be the number of all the sources that socially influence the agent on topic-opinion c, according to properties P_1, P_{2a} and P_{2b}; the set of these sources is called the agent's social set on topic-opinion c, denoted by G. The degree to which the agent agrees or disagrees with topic-opinion c is denoted by λ_0. The value of λ_0 is between -1 and $+1$; indeed, topic-opinion c is generally a complex opinion so, by the H_2 assumption, such value is meaningful. In other words, we could say that λ_0 represents the agent's opinion on topic c.

We also need to represent the sources' opinions on the topic. But we want to capture these opinions from the agent's point of view. After all, influence is achieved through what the agent receives as information from a source. And given the same source, different agents may perceive the source's opinion on a topic in different ways, since they may receive different information from the source (different due to quality and/or quantity of information, just as we described when we defined the profile of a source according to an agent). So, we are interested in a source's opinion the way the agent perceives it, which may be different from the "true" opinion. Taking this under consideration, the degree to which a source i (of social influence on c, $i = 1,2, \dots, N$) agrees or disagrees with topic-opinion c, according to the agent, is denoted by λ_i. Again, λ_i has a value between -1 and $+1$, since topic-opinion c is generally a complex opinion (H_2 assumption). In other words, we could say that λ_i represents source i's opinion on topic c, according to the agent.

We want to highlight the fact that p_i and λ_i represent two different things. Both p_i and λ_i are defined according to the agent, but λ_i depends on the topic under consideration (by definition) while p_i is topic-independent. Moreover, λ_i is expressed by a numerical value while p_i is a complex idea. A source, that is within the scope of the agent's knowledge, always has a profile according to the agent, whether it satisfies (1) or not (so, by P_{2a} and P_{2b}, it is a source of social influence on a certain topic-opinion or not, respectively); on the other hand, λ_i is defined for a source of social influence on a topic-opinion. Finally, any new piece of information received by the agent from a source (or any existing piece of information forgotten) automatically changes-updates the profile (p_i) of the source according to the agent but it does not necessarily change the source's opinion (λ_i) on a topic, according to the agent.

Let us give an example for the last observation of the previous paragraph. *Let us consider a politician that an agent has voted for. The politician (source i) has a certain profile p_i according to the agent; also, the politician totally supports freedom of the press/media (topic-opinion) according to the agent (so, it is $\lambda_i \cong +1$). At some point, the agent learns that the politician's university degree is fake, so p_i is updated. But λ_i probably remains the same. Of course, it could be the case that the agent later learns*

that the politician threatened a journalist in order not to publish the fake-degree story, so p_i is updated once again, but now λ_i probably becomes much less than $+1$ (the politician does not agree with freedom of the press/media that much after all, according to the agent).

So far, we have defined λ_0 and λ_i $(i = 1,2, \dots, N)$. There are two more aspects that must be discussed so that our model can be formulated mathematically. More specifically, given the agent and a source of social influence on topic-opinion c, we provide an answer to how strong the source's influence is on the agent. By answering this question, we are able to make the distinction between a more influential and a less influential source. We also provide an answer to what the effect of the source's influence is on the agent; depending on the source and its influence, the agent may be receptive to the source's opinion or may be rejective of it. In the next paragraphs, we define these aspects and present the complete picture of our model.

Let us start by describing and eventually defining the strength or magnitude of the source's influence on the agent. As we know, property P_{2a} refers to a source i of social influence, given the agent and topic-opinion c, and makes use of the (cardinal) number $\left|A_c \cap A_{p_i}\right|$. This number is denoted by K_i and measures the quantity of information (simple ideas) that is communicated to the agent from source i (that belong to A_{p_i}) and is relevant to the topic-opinion at the same time (belong to A_c). So, for a source i of social influence we have $K_i = \left|A_c \cap A_{p_i}\right|$; for all the sources in G (the agent's social set on topic-opinion c) we have $K_1, K_2, \dots, K_N$. The sum of all these is denoted by K_0, which means

$$K_0 = K_1 + K_2 + \cdots + K_N. \tag{2}$$

In this way, K_0 measures the total amount of information that is received by the agent from all the sources and is relevant to the topic-opinion[8]. It represents the extent to which the agent is informed about the topic (correctly informed and/or misinformed-disinformed). This is an essential parameter in our model.

Now, the magnitude (or strength) of a source i's influence on the agent is denoted by m_i and is defined as

$$m_i = \frac{K_i}{K_0} \quad , \tag{3}$$

where $m_i \in [0,1]$. In other words, m_i is the fraction which represents the contribution of source i to the total amount of information. For example, *for two sources i and j in G (that influence the agent on topic-opinion c) with $K_i > K_j$, we have $m_i > m_j$; indeed, the agent receives a greater amount of (relevant-to- c) information from source i than from source j, therefore the magnitude of source i's influence on the agent is greater than the magnitude of source j's influence. We could say that source i is more influential than source j. For another example, let us consider the magnitude m_i of a source i's influence on the agent; if the agent receives more information (relevant-to- c) from a source other than i, then K_0 becomes larger but K_i remains the same. As a*

[8] One could say that in reality, opinions may also be influenced by pieces of information that are seemingly not very relevant to the topic-opinion under consideration. We assume that our notion of relevance (based on common simple ideas of the alphabet of human thoughts) really captures even these pieces of information, leaving out only the "truly" irrelevant ones.

result, the updated magnitude m_i' of source i's influence on the agent is less than the previous m_i.

The magnitude of a source's influence on the agent (or, equivalently, the characterization of the source as being more or less influential) captures the strength of the influence but not its effect. By effect, we mean that a source may influence the agent to be receptive to its opinion or to be rejective of its opinion. For example, *let us consider the phrase "my enemy's enemy is my friend". This means that "I" (the agent) have an enemy called George (the source) and George considers Paul to be his enemy, so I consider Paul to be my friend. In this case George, being my enemy, influences me to be rejective of his opinion (that Paul is an enemy), therefore I form the opinion that Paul is a friend. So, in order to describe George's influence on me completely, we need to express how strong the influence is (the magnitude) and we also need to express the fact that George is an enemy, which results in the agent being rejective of George's opinion (the effect). Both magnitude and effect are necessary and we can appreciate this in the following way. Let Jim be my friend who considers Paul to be his enemy; Jim, being my friend, influences me to be receptive to his opinion (that Paul is an enemy). But now, Paul is my enemy's enemy and my friend's enemy at the same time. What I consider Paul to be depends on whose influence has the greater magnitude, George's or Jim's*[9]. Having said all that, we define the effect of a source's influence on the agent in the next paragraphs.

As we know, given the agent, topic-opinion c and a source i of social influence, $A_c \cap A_{p_i}$ is the set of simple ideas that belong to A_{p_i} and to A_c at the same time. The fact that these simple ideas belong to A_c makes them simple opinions; indeed, according to assumption H_1, all elements of A_c are simple opinions, since topic-opinion c is a complex opinion. And according to assumption H_2, each of these simple opinions can be associated with a value $+1$ or -1, depending on whether the agent totally agrees $(+1)$ or totally disagrees (-1) with the simple opinion under consideration.

So, we showed that the elements of $A_c \cap A_{p_i}$ are simple opinions. Their natural meaning is still the same; they are the relevant-to- c information that the agent receives from source i. But now, we can consider a sequence of $+1s$ and $-1s$ to represent whether the agent agrees or disagrees with this information (with each of these simple opinions). The sequence is denoted by v_i. The length of the sequence is $K_i = |A_c \cap A_{p_i}|$ (a value $+1$ or -1 for each simple opinion), which can be written as $\ell(v_i) = K_i$. The number of $+1s$ in v_i is denoted by $\ell(v_i^+)$, while the number of $-1s$ in v_i is denoted by $\ell(v_i^-)$. Of course, it is $\ell(v_i) = \ell(v_i^+) + \ell(v_i^-)$. Now, let us consider the difference $\ell(v_i^+) - \ell(v_i^-)$. We have the following cases:

A) If $\ell(v_i^+) - \ell(v_i^-) > 0$, then the quantity of information with which the agent agrees is greater and we say that the agent is receptive to source i's opinion (λ_i).

B) If $\ell(v_i^+) - \ell(v_i^-) < 0$, then the quantity of information with which the agent disagrees is greater and we say that the agent is rejective of source i's opinion (λ_i).

The idea is that by having source i's opinion (λ_i) multiplied by a positive difference (the agent is receptive to λ_i), the sign of λ_i remains the same; on the other

[9] As we will see later, what I consider Paul to be is a weighted average (with the magnitudes as weights). In that sense, a possible outcome could be *e.g.* I consider Paul to be closer to being my friend than to being my enemy, since the magnitude of George's influence is greater than that of Jim's influence.

hand, by having source i's opinion (λ_i) multiplied by a negative difference (the agent is rejective of λ_i), the sign of λ_i is reversed. For case A, if the difference becomes larger, then the agent is more receptive to source i's opinion. For case B, if the absolute value of the difference becomes larger, then the agent is more rejective of source i's opinion. Finally, we are ready to present the definition. The effect of a source i's influence on the agent is denoted by e_i and is defined as

$$e_i = \frac{\ell(v_i^+) - \ell(v_i^-)}{\ell(v_i)} = \frac{\ell(v_i^+) - \ell(v_i^-)}{K_i} \quad , \tag{4}$$

where $e_i \in [-1, +1]$. The difference is divided by $\ell(v_i) = K_i$ so that the value of e_i is between -1 and $+1$.

Let us combine all the above, which are all the necessary parts of our model. Given an agent, a topic-opinion c and the agent's social set G (on topic-opinion c) which contains N sources of social influence, the mathematical formula of our model is the following:

$$\lambda_0 = \sum_{i=1}^{N} m_i \, e_i \lambda_i \quad . \tag{5}$$

The agent's opinion on topic c, or the degree to which the agent agrees or disagrees with topic-opinion c, is the weighted average of products $e_i\lambda_i$ $(i = 1,2, \dots, N)$, with the magnitudes (m_i) as weights. The agent's opinion formation is influenced by source i's opinion according to the effect and the magnitude of the influence, for all the sources in the agent's social set.

Given (3) and (4), formula (5) is written as:

$$\lambda_0 = \sum_{i=1}^{N} m_i \, e_i \lambda_i = \sum_{i=1}^{N} \frac{K_i}{K_0} \cdot \frac{\ell(v_i^+) - \ell(v_i^-)}{K_i} \cdot \lambda_i = \sum_{i=1}^{N} \frac{\ell(v_i^+) - \ell(v_i^-)}{K_0} \cdot \lambda_i$$

so, we have

$$\lambda_0 = \frac{1}{K_0} \sum_{i=1}^{N} \left(\ell(v_i^+) - \ell(v_i^-)\right) \cdot \lambda_i \quad . \tag{6}$$

We should note that, unlike the literature models of opinion formation we talked about, there is no time parameter in our model. In that sense, opinion dynamics is not about evolution of opinion over time; it is about evolution of opinion over parameter K_0, which represents the extent to which the agent is informed about the topic ((6)). So, given a K_0, the agent's opinion λ_0 is calculated. *As K_0 increases (new information is received by the agent) or decreases (existing information is forgotten by the agent), the agent's opinion evolves. In models like the Friedkin-Johnsen model, where a recursive definition is used for the influence process, the group members' initial opinions must be known for the opinion changes to be calculated. Our model is designed to treat such opinions as any other opinion; an "initial" opinion is calculated according to its corresponding K_0.*

Finally, let us remember the example "my enemy's enemy is my friend" and represent it using our model. *The topic-opinion is "Paul is a friend". George considers Paul to be his enemy, which means that George totally disagrees with the topic-opinion and* $\lambda_{Geo} = -1$. *The fact that George is my enemy is expressed by* $e_{Geo} = -1$ *(simplified). Jim considers Paul to be his enemy, so Jim also totally disagrees with the topic-opinion and* $\lambda_{Jim} = -1$. *The fact that Jim is my friend is expressed by* $e_{Jim} = +1$ *(simplified). Let* $m_{Geo} = 0.6$ *and* $m_{Jim} = 0.4$, *which means that the magnitude of George's influence is greater than that of Jim's influence*. By using formula (5), we have:

$$\begin{aligned}\lambda_0 = m_{Geo}\, e_{Geo}\, \lambda_{Geo} + m_{Jim}\, e_{Jim}\, \lambda_{Jim} &= 0.6 * (-1) * (-1) + 0.4 * (+1) * (-1) \\ &= 0.6 - 0.4 \\ &= +0.2 \quad .\end{aligned}$$

This is the degree to which I agree with the topic-opinion. The interpretation of the outcome is that I consider Paul to be closer to being my friend than to being my enemy.

At this point, let us note that the notation we have used so far concerns a fixed agent and a fixed topic-opinion. In order to avoid ambiguity when, for example, there are two or more agents or an agent forms opinion on two or more topics, we consider the following notations, just as we defined the profile $p_{i,k}$ of source i according to agent k. Given an agent k, a topic-opinion c and a source i of social influence, we have:

$$G_{k,c}\ ,\ \lambda_{0,k,c}\ ,\ \lambda_{i,k,c}\ ,\ K_{i,k,c}\ ,\ K_{0,k,c}\ ,\ v_{i,k,c}\ ,\ m_{i,k,c}\ ,\ e_{i,k,c}\ .$$

3.2 *A framework for the agent's social set*

Model (5) requires knowledge of the agent's social set G. But in real applications, the task of identifying all the sources in G seems to be unachievable; any source that satisfies properties P_1 and P_2 is a source of social influence. Nevertheless, we attempt to capture the agent's social set as adequately as possible without explicitly describing it, in order to make the model more practicable.

Let us consider a person that is influenced on a topic c by N sources, each communicating relevant-to-c information to the person. Obviously, the amount of information that is communicated ($K_i,\ i = 1,2,\dots,N$) varies from source to source, depending on the level of the agent's exposure to the source (relevant-to-c exposure). For example, *given a common political topic of general nature, some sources of social influence may have much larger* K_i *(e.g. a child's parents, a person's spouse etc.), other sources may have smaller* K_i *(e.g. friends, colleagues etc.) and other sources may have much smaller* K_i *(e.g. books, political figures, sports idols, newspaper articles, celebrities, movies etc.).* We believe that this diversity can be captured, since there is an underlying pattern in a person's exposure to sources.

Research on the size and structure of an individual's social network (the people with whom social relationships are maintained) has suggested that social relationships can be clustered into groups of escalating sizes, where a group of larger number of relationships (people) contains less intense relationships [27, 28]. In other words, an individual maintains strong social relationships with a small number of people. Given a

group which contains relationships of a certain intensity, the size of the group (the number of people with whom such relationships can be maintained) is limited by time, cognitive and other constraints [27, 29, 30, 31].

The research mentioned in the previous paragraph is about an agent's social network, containing the people with whom the agent has social relationships. In our theory, as we know, we consider not an agent's social network, but an agent's social set which includes all sources (not just people and not just in a social relationship) that socially influence the agent on a topic. But it is natural to assume that certain observations made about the agent's social network are true for the agent's social set as well. In particular, sources of social influence can be clustered into groups of different sizes, where a group containing sources that communicate smaller amounts of information (smaller K_i) is a group of greater size (larger number of sources). So, an agent receives large amounts of information (large K_i) from a small number of sources. Again, several constraints – most notably time and cognitive constraints – are the reasons for this; an agent can only manage fewer sources with large K_i, but many more sources with small K_i. For example, *given a common political topic of general nature, a person has typically received large amounts of information from his/her parents* [32]. *Such amounts can be matched by other sources (e.g. the person's spouse later in life or a lifelong close friend or a favorite author after extensive reading of the author's relevant books) but only in a limited number, due to the constraints. On the other hand, the person has been (briefly) exposed to many sources of social influence over the years, each of which communicated a small amount of information (numerous newspaper articles, books, movies, songs, several acquaintances, historical figures etc.).*

So, sources of social influence can be clustered into groups according to the above. The group which contains sources with large K_i is typically structured and static; indeed, a new source to be included in (or an existing one to be removed from) the group is difficult to happen, since communicating a large amount of information to the agent (or letting the existing large amount of information to be forgotten by the agent) requires time and effort. On the other hand, the group which contains sources with small K_i is dynamic (characterized by constant changes in sources), since a small amount of information is easy to be communicated or forgotten. In addition to being dynamic, the potentially vast number of sources and their diversity in nature makes the group unstructured. Obviously, the problem is found in the latter group. However, we believe that this problem can be dealt with to a certain degree, by proposing a connection between the structured-static and the unstructured-dynamic group. The discussion in the next paragraphs is about this connection.

As we know, the sources that socially influence the agent on a topic satisfy properties P_1 and P_2. The agent can have control over both properties and therefore, control over sources. First, let us discuss control over property P_1. There are two ways in which a source of social influence gets to be within the scope of the agent's knowledge. One way is that the source is externally imposed on the agent and the agent has no control over this, *e.g. a child does not usually get to choose classmates or a person hired by a company typically has no say in who his/her colleagues are going to be* (a similar discussion, in the context of political socialization, can be found in [33]). But the other way is that the agent can actively search and discover sources in the social environment, *e.g. a person chooses to read a greater number of newspaper articles and to watch tv news in different channels for multifaceted information, or chooses to participate in activities where exposure to certain sources (otherwise not encountered)*

is achieved. Now, let us discuss control over property P_2, which concerns the externally imposed sources as well. The agent can receive a greater amount of (relevant-to- c) information from a source by choosing to learn more from the source. For example, *a person's grandfather may have a* $K_i = |A_c \cap A_{p_i}|$ *less than* $\vartheta(c)$ *(not a source of social influence on* c*, where* c *is a political topic), but the person starts discussing politics with the grandfather more frequently and learns more from him and eventually* K_i *becomes greater than* $\vartheta(c)$ *(and the grandfather now influences the person on* c*).*

It is clear that such control, as described above, depends on the agent's interest in the topic. Indeed, an agent with greater interest in the topic actively searches for and learns from sources with greater intensity[10]. For the role of political interest and attentiveness in receiving political messages, the reader may see [34]. It is also clear that the agent's interest in a topic is stimulated by sources which influence[11] the agent, *e.g. parents, friends and teachers play an important role in developing adolescents' societal interest* [35]. So, sources of social influence stimulate the agent's interest in a topic and this interest is directed at sources of social influence again (since the agent searches for and learns from sources).

Let us provide evidence from research on political socialization and youth civic development. One of the findings in [36] suggests that youth having political discussions more frequently with their parents report attending to national news (via print and broadcast media) more frequently. Studying intergenerational transmission, the authors in [37] examined children aged 18 (in 1965) and their parents and found that children raised in highly politicized families were more likely to reject their parents' opinion regarding school integration (than were children in less politicized families). More specifically, these children were more likely to support school integration, given that their parents did not commonly support it. Jennings, Stoker and Bowers provided the following conclusion: "*On the one hand, then, having a politicized family environment typically encourages the child to learn from the parent and to adopt the parent's views. On the other hand, it also leaves the child more attuned to outside political influences*". In 1965, such outside political influences were provided by the civil rights movement. In [33], the study suggests that young adults coming from more politicized homes are more likely to deviate from their parents' partisan preferences, since these adults are more likely to be exposed to other political influences (new social contexts and political events).

All this evidence supports our earlier observation, that sources of social influence stimulate the agent's interest in a topic and this interest is directed at sources of social influence again. *Indeed, parents (sources) stimulate their child's interest in politics (through political discussions or by providing a politicized environment) and this interest is expressed by attending to national news or being exposed to other political influences, so it is directed at sources from the media/civil rights movement/new social contexts/political events respectively*. In the context of the evidence we presented, we

[10] A necessary condition for a person to discover sources and learn from them is access to these sources. A person with no access to sources, due to socioeconomic reasons, cannot learn from them even if he/she is interested in the topic.

[11] These sources are indeed sources of social influence on the topic since, for the agent's interest in the topic to be stimulated by them, the agent needs to be aware of them (P_1) and they need to communicate to the agent relevant-to-topic information in a sufficiently large amount (P_2). So, for a source to stimulate the agent's interest, it is necessary for the source to influence the agent. But it is not sufficient; interest is really an opinion and for example, the agent may not feel interested in the topic as a result of other sources' influence or of being rejective of the source's attempt to stimulate.

just described an interest-based connection between parents and sources from the above categories. Now, let us remember the earlier discussion about clustering of sources into groups and let us consider the familiar scenario which involves two groups, the structured-static group which contains a small number of sources with large K_i each and the unstructured-dynamic group which contains a large number of sources with small K_i each. *According to what we have said about common political topics of general nature, parents typically belong to the static group while sources from the media/civil rights movement/new social contexts/political events typically belong to the dynamic group*. So actually, we established a connection between sources in the static group and sources in the dynamic group.

Finally, let us clarify the exact nature of this connection. *In the cases of the evidence, we see that having political discussions more frequently leads to attending to news more frequently and that being raised in a more politicized environment leads to being more exposed to other political influences. We observe that "more frequently" leads to "more frequently" and that "more politicized" leads to "more exposed". This can be interpreted in the following way: the intensity of the influence generated by parents determines the intensity of the influence generated by the sources from the media/civil rights movement/new social contexts/political events*. We believe that this pattern is not limited to parents and the above sources and is not limited to political topics, but it is characteristic of the connection between the entire static group and the entire dynamic group for any topic. So, let the sources that influence a person on a topic be clustered into two groups, a static and a dynamic. We assume that the following is true: the intensity of the influence generated by the static group determines the intensity of the influence generated by the dynamic group.

Let us express our assumption in mathematical terms. We consider an agent, a topic-opinion c and the agent's social set G (on topic-opinion c) which contains N sources of social influence. Let these sources be clustered into two groups, a static and a dynamic. The static group is denoted by X and the number of sources in it by N_X, while the dynamic group is denoted by Z and the number of sources in it by N_Z ($N_X < N_Z$, $N_X + N_Z = N$). Now, the intensity of the influence generated by a source i is the product $m_i|e_i|$, which involves the magnitude and the effect of the source's influence. Note that we consider the absolute value of the effect, since we want to measure the intensity of the influence without regard to its sign. Then, the intensity of the influence generated by a group is the sum of the intensities of all the influences generated by the sources in the group. For example, *let X (static group) contain only the parents and let $m_{father} = m_{mother}$ and $e_{father} = +1$, $e_{mother} = -1$. The intensity of the influence generated by X is*

$$\begin{aligned} m_{father}|e_{father}| + m_{mother}|e_{mother}| &= m_{father}|+1| + m_{mother}|-1| \\ &= m_{father} + m_{mother} \\ &= 2 * m_{mother} \quad , \end{aligned}$$

which is greater than 0. *If we did not consider absolute values, the influence generated by X would have zero intensity. This would be clearly false, since influences are not only present (generated by parents), they have extreme effects as well ($+1$ and -1).* So by considering absolute values, all influences contribute to the total intensity.

Based on the entire reasoning that we developed in the previous paragraphs, the assumption is expressed as follows:

$$\sum_{i=1}^{N_X} m_i|e_i| = \sum_{i=N_X+1}^{N} m_i|e_i| \quad , \tag{7}$$

where a source i for $i = 1,2,\dots,N_X$ is considered to belong to static group X, while a source i for $i = N_X+1, N_X+2, \dots, N$ is considered to belong to dynamic group Z. So, the intensity of the influence generated by the static group determines the intensity of the influence generated by the dynamic group. More specifically, the intensity generated by the static group results in an equal amount of intensity being generated by the dynamic group. The way we expressed the assumption allows of the interpretation in the opposite direction as well, meaning that the intensity generated by the dynamic group results in an equal amount of intensity being generated by the static group. The opposite direction is possible since, for example, *a child attending to news may result in the child having political discussions with the parents*. For a similar discussion, the reader may see [36].

Equality (7) is written as:

$$\sum_{i=1}^{N_X} \frac{K_i}{K_0} \cdot \left|\frac{\ell(v_i^+) - \ell(v_i^-)}{K_i}\right| = \sum_{i=N_X+1}^{N} \frac{K_i}{K_0} \cdot \left|\frac{\ell(v_i^+) - \ell(v_i^-)}{K_i}\right|$$

therefore

$$\frac{1}{K_0} \sum_{i=1}^{N_X} |\ell(v_i^+) - \ell(v_i^-)| = \frac{1}{K_0} \sum_{i=N_X+1}^{N} |\ell(v_i^+) - \ell(v_i^-)|$$

and hence

$$\sum_{i=1}^{N_X} |\ell(v_i^+) - \ell(v_i^-)| = \sum_{i=N_X+1}^{N} |\ell(v_i^+) - \ell(v_i^-)| \quad . \tag{8}$$

Now, let the average of the quantities $|\ell(v_i^+) - \ell(v_i^-)|$ in static group X be denoted by e_X and let the average of the quantities $|\ell(v_i^+) - \ell(v_i^-)|$ in dynamic group Z be denoted by e_Z. So, we have:

$$e_X = \frac{1}{N_X} \sum_{i=1}^{N_X} |\ell(v_i^+) - \ell(v_i^-)| \tag{9}$$

and

$$e_Z = \frac{1}{N_Z} \sum_{i=N_X+1}^{N} |\ell(v_i^+) - \ell(v_i^-)| \quad . \tag{10}$$

Given (9) and (10), equality (8) is written as:

$$N_X e_X = N_Z e_Z \quad . \tag{11}$$

If the number N_X of sources in the static group and the average e_X are known, then equation (11) provides information about the number N_Z of sources in the dynamic group (given the average e_Z) or about the average e_Z (given N_Z). Given the structured and limited (in number of sources) nature of the static group, information about the static group (like N_X and e_X) is more effectively obtained than information about the dynamic group; equation (11) provides a way to calculate the latter using the former information.

Finally, the framework that we propose for the agent's social set is based on the assumption we have described. So far, the assumption involved two groups, a static one and a dynamic one. But for many topics, sources of social influence follow the typical social structure of "family-friends-acquaintances". Taking this structure under consideration, we consider a possible clustering of sources into three groups, two structured-static groups (X and Y, representing "family" and "friends" respectively) and one unstructured-dynamic group (Z, representing "acquaintances"). The "family" group contains a small number of sources with very large K_i each, the "friends" group contains a larger number of sources with smaller but still large K_i each and the "acquaintances" group contains a much larger number of sources with much smaller K_i each. Of course, these groups may not literally represent members of the family or actual friends and acquaintances. For example, *parents typically belong to the "family" group, but another source with large enough K_i may belong to the group as well, e.g. a person's closest friend. The "friends" group may contain not only friends but other sources with similar K_i as well, e.g. a teacher or a religious text that the person reads frequently. Similarly for the "acquaintances" group, in which a source only needs to have a small K_i*. We believe that a clustering of sources into more than three groups not only has reduced practicability, it is also more forced and less natural.

We collect all the above discussion in the following assumption. We mention that the average K_i in groups X, Y, Z is denoted by K_X, K_Y and K_Z respectively. So, we have:

H_3) Given an agent, a topic-opinion c and the agent's social set G (on topic c) which contains N sources of social influence, these sources can be clustered into:

a) two groups, the static group X and the dynamic group Z, where $N_X < N_Z$, $N_X + N_Z = N$, $K_X > K_Z$ and

$$N_X e_X = N_Z e_Z \tag{11}$$

or

b) three groups, the static groups X and Y and the dynamic group Z, where $N_X < N_Y < N_Z$, $N_X + N_Y + N_Z = N$, $K_X > K_Y > K_Z$ and

$$N_X e_X + N_Y e_Y = N_Z e_Z \quad . \tag{12}$$

Let us consider two topics, c_1 and c_2, on which the agent forms opinion under social influence. We say that c_1 is more general than c_2 (or that c_2 is more specific than c_1), if

$|A_{c_1}| > |A_{c_2}|$. Indeed, a more general topic is expected to cover a wider range of ideas, therefore it is composed of a greater number of simple ideas. For example, *a topic concerning the role of playing in child development is more general than a topic about the expected sales of item x produced by company z*. Given the fact that a general topic is composed of a large number of simple ideas, forming opinion on this topic is likely to involve more sources of social influence with diverse K_i, probably resulting in a clustering of sources into three groups. On the other hand, forming opinion on a specific topic (composed of a small number of simple ideas) is likely to involve sources with similar K_i (with low standard deviation), meaning that there is probably no need for a clustering at all ("clustering" of all sources into one group).

3.3 *Opinion and knowledge within a group*

As we mentioned in the introduction, Plato (in Theaetetus) described knowledge as justified true belief. This view sparked a widespread debate in epistemology about whether the three conditions (justified, true, belief) are necessary and sufficient conditions for knowledge [38, 39, 40, 41, 42, 43] (even Plato argued against such a way of defining knowledge). In this chapter, we consider this definition as the starting point for our discussion and we approach knowledge from a sociological perspective. We also discuss the way opinion networks and knowledge networks are connected, based on this approach.

Knowledge is justified, true belief (*JTB*). However, when considering a group of agents (society), there are cases in which a sufficiently justified (perceived as such) yet false belief is considered to be true, therefore knowledge. For example, *fake news involves such cases, where the justification – always incorrect – may mislead people and make them believe that such news is knowledge.* For another example, *a formerly accepted scientific theory, later proved wrong, used to be considered knowledge*. On the other hand, when the topic is about actual knowledge (justified true belief), there are cases in which this true belief is considered to be false, therefore not knowledge. For example, *people who believe that the Earth is flat do not consider Earth's sphericity to be knowledge*.

Let us summarize the above as follows. For a belief to be knowledge, it must be true and justified (*JTB*). For a belief to be considered knowledge by an agent, it must be perceived as true by the agent (even if it is false). From a sociological perspective, the emphasis is on "considered". Similarly, there is a difference between the phrase "a topic is a matter of opinion/knowledge" and the phrase "a topic is considered a matter of opinion/knowledge". In the next paragraphs, we discuss the latter.

Let us have a group of agents (society). We present two informal, self-evident definitions.

Concept 1: Topic considered a matter of knowledge within the group.
A topic is considered a matter of knowledge within the group (society) if opinions on the topic are considered true or false.

Concept 2: Topic considered a matter of opinion within the group.
A topic is considered a matter of opinion within the group (society) if opinions on the topic are not considered true or false.

We are interested in these two concepts because concept 1 is the item of knowledge networks, while concept 2 is the item of opinion networks. A possible connection between the two concepts lies in an observation that we have already mentioned. When a topic is considered a matter of opinion within a society, a member of the society is typically influenced on the topic by sources that are different from another member's sources. For example*, the topic "is basketball your favorite sport to watch?" is considered a matter of opinion, since opinions on this are not considered right-true or wrong-false (according to the informal definition). At the same time, a person may be influenced on this by his/her family members, his/her friends, sports idols of his/her era/childhood/country, newspaper articles promoting sports etc., while another person is influenced by sources that most likely are very different*. In other words, a matter of opinion is associated with agents' social sets that are different.

Having said that, let us consider a spectrum with one end being "social sets completely different from each other" and the other end being "social sets equal to each other" (fig.3). Given a group (society) of M agents forming opinion on a topic c, the agents' social sets are $G_{1,c}$, $G_{2,c}$, ... , $G_{M,c}$. The label "social sets completely different from each other" is written as $G_{k,c} \cap G_{l,c} = \emptyset$, whenever $k \neq l$. The label "social sets equal to each other" is written $G_{1,c} = G_{2,c} = \cdots = G_{M,c}$. So, when a topic is considered a matter of opinion within the group, we are closer to the left end (as seen in fig. 3). As we move towards the right end, topics are associated with social sets that are more and more similar to each other; in this region of the spectrum, we may find topics that are considered matters of knowledge within the group. For example, *let us have a work-related technical topic that concerns a group of people working in an office. These people are influenced on this topic by mostly (if not entirely) common sources, like each other, shared emails/documents/phone calls etc. (no parents or favorite authors). And the view on the topic is that it involves right or wrong decisions, so it is considered a matter of knowledge.* Another example is *education-related topics which are considered knowledge, with common books, common educational tv programs etc. for sources*.

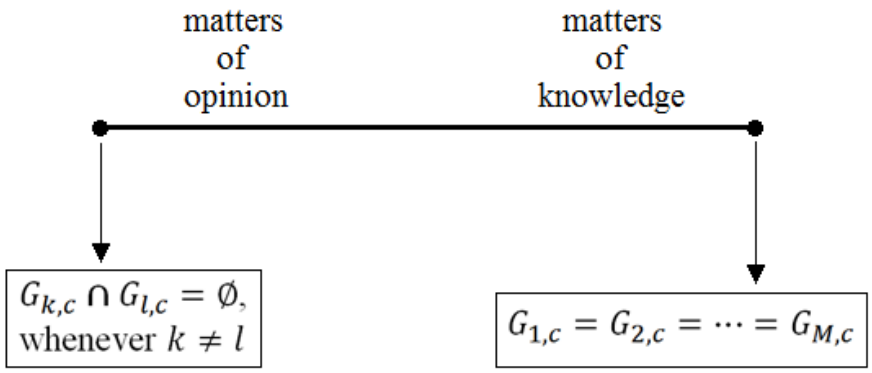

Fig.3. When a topic is considered a matter of opinion within the group, we are closer to the left end. When a topic is considered a matter of knowledge within the group, we are closer to the right end.

4 Evaluation of the model

4.1 *Method*

We use the model of social influence on opinion formation to process data collected from questionnaires (see Appendix), concerning attitudes towards elections and in particular, opinions on the importance of the July 2019 national elections in Greece. The respondents (32 in number) were chosen to be young adults of ~24 years of age. By this age group, we wanted to achieve two things, namely that the elections of July 2019 were not their first elections (so that the "first-time-enthusiasm" would be avoided) and that they were still significantly influenced by their parents on political topics. They answered a total of 13 questions, 8 of which concerned personal and familial information about politics in general and the elections in particular and the rest 5 were knowledge questions on political events that took place not after the election day and its results. The way we assign values to these answers is presented below, with the first value in the parenthesis corresponding to the first answer, the second value to the second answer *etc*.

Question 1 provides a value for the respondent's opinion (λ_0) on the importance of the 2019 national elections (values 0.75, 0.5, 0.25, 0, -0.25). Questions 2 and 3 provide values for the father's and mother's opinion, λ_1 and λ_2, respectively (values 0.75, 0.5, 0.25, 0, -0.25 and 0). Questions 4 and 5 provide values for K_1 and K_2 respectively; indeed, the frequency of political discussions with the father and the mother is viewed as an index for the quantities of information communicated from them (values 1, 2, 3, 4). Questions 6 and 7 provide values for e_1 and e_2 respectively (values 0.8, 0.4, 0, -0.4, -0.8). Question 8 is used to assess the respondent's compatibility with the approach taken in this application. Questions 9-13 provide values for the part of K_0 that is attributed to sources in the dynamic group. These are questions that cover different aspects of the conventional political space in Greece, so an agent knowing the answers to these questions is interpreted as being informed about the topic from possibly different sources in the dynamic group. Each right answer to questions 9-13 is interpreted as a source in the dynamic group with $K_i = 1$ (it is kept small, so that it is consistent with the dynamic group). These sources can be regarded as compressed expressions of broader political forces/spaces containing many sources. Note that we prefer the term right instead of correct. Since we are interested in the extent to which the agent is informed about the topic (correctly informed and/or misinformed-disinformed), the former term is more appropriate than the latter. Of course, the only right answer in questions 9-12 is the correct answer, but in question 13 we accept two right answers[12].

Now, let us make two simplification assumptions. First, we expect that people may agree or disagree on politics, but whatever the case they agree with each other on the importance of elections. Questions 6 and 7, which provide values for e_1 and e_2, ask for the level of agreement or disagreement with parents on politics in general. So whatever the case may be, it is interpreted as agreement with parents on the importance of elections, at the same level as on politics. Therefore, we consider the absolute values of e_1 and e_2 as the correct values; similarly for the sources in the dynamic group.

[12] We accept the correct and one more close to it. The quantitative difference between these two answers and the technical nature involved makes it easy to be misunderstood, so these details do not affect the general picture that even this particular incorrect answer shows that the agent is informed.

Second, we said that the sources in the dynamic group represent broader political forces and spaces, which are mainly sources that are characterized by a strong political identity (for example, political parties, politicians *etc.*). In that sense, the opinion of such a source must be close to +1, otherwise it would be contradictory and inconsistent with its strong political identity. Again, there are members in a society that are opposed to elections, but they must be vocal in expressing such an opinion, which we believe is far from the typical case. These two simplification assumptions can be summarized in the following way: *1)* the quantity $\ell(v_i^+) - \ell(v_i^-)$, for $i = 1,2,\dots,N$, is taken in its absolute value, since e_i is taken in its absolute value in (4) and *2)* we have $\lambda_i = +1$, for $i = 3,\dots,N$ (for sources in the dynamic group).

So for each agent-respondent, we collect λ_0, λ_i (for $i = 1,2,\dots,N$), K_1, K_2 and the absolute values of e_1 and e_2. By assumption H_3, we have a clustering of sources into two groups and we also assume that the static group for each agent contains only two sources, the agent's father and mother[13]. So, $N_X = 2$ and N_Z is the number of right answers. Also, from equation (2), adding K_1, K_2 and the number of right answers together gives us an index for K_0. We work as follows:

- We calculate $\ell(v_i^+) - \ell(v_i^-)$, for $i = 1,2$, from (4).
- We calculate e_X from (9).
- We calculate e_Z from (11).

So, we can calculate the predicted λ_0 from (6):

$$predicted\ \lambda_0 = \frac{1}{K_0} \sum_{i=1}^{N} \left(\ell(v_i^+) - \ell(v_i^-)\right) \cdot \lambda_i$$

which, according to the assumption of absolute values, is written

$$predicted\ \lambda_0 = \frac{1}{K_0} \sum_{i=1}^{N} |\ell(v_i^+) - \ell(v_i^-)| \cdot \lambda_i \quad . \qquad (13)$$

The part of the sum that concerns the sources in the dynamic group, namely

$$\sum_{i=3}^{N} |\ell(v_i^+) - \ell(v_i^-)| \cdot \lambda_i \quad ,$$

is written (for $\lambda_i = +1$, for $i = 3,\dots,N$) as

[13] Given the discussion we had earlier on the role of parents in political socialization and due to the young age of respondents, we believe that the main influence from sources in the static group is that of parents. Of course, influence coming from friends and other sources is always present and perhaps a clustering of sources into three groups is preferable under certain circumstances. But in our analysis here, we consider only parents as sources of social influence in the static group and we discuss the cases where such an analysis is not sufficient.

$$\sum_{i=3}^{N} |\ell(v_i^+) - \ell(v_i^-)| = N_Z \, e_Z$$

according to (10). All variables in (13) are now known, therefore the predicted λ_0 is calculated.

4.2 *Results*

We present all the values obtained from the questionnaires and the values for the predicted λ_0 in table 1. We also report results about the effectiveness of the model in predicting the respondents' opinions, for two groups of respondents. The first group consists of all the respondents (32 in number). The second group is formed by removing from the first group those respondents that are incompatible with the use of the model in this application. If our theory and model are built on solid foundation, then we expect a greater effectiveness of the model for the second group. In the next paragraph, we explain which respondents are removed in order to form the second group and why.

Let us remember that question 8 is used to assess the respondent's compatibility with the approach taken in this application. In particular, the question asks whether the respondent's political opinions can be placed in one region of the political spectrum (left, centre, right). A respondent who answers that his/her political opinions are clearly placed in one region or belong to more than one regions is compatible with the approach of our analysis, since we assumed that influence comes from the conventional political

Table 1. Values obtained from the questionnaires and values of the predicted λ_0

	λ0	λ1	λ2	K1	K2	e1	e2	right answers	predicted λ0
1	0,75	0,75	0,5	3	1	0,4	0,8	1	0,66
2	-0,25	0,5	0,5	4	4	0,4	0,4	2	0,48
3	0,75	0,75	0,5	4	2	0,8	0,8	2	1
4	0,5	0,5	0,5	3	3	-0,8	-0,8	2	0,9
5	0,75	0,75	0,75	2	2	0,4	0,4	1	0,56
6	0,5	0,5	0,5	2	2	-0,4	-0,4	0	0,6
7	0,25	0,5	0,75	1	2	0	0,4	3	0,233333333
8	0,5	0,75	0,5	2	1	0,8	0,8	4	0,571428571
9	0,75	0,5	0,5	2	1	-0,8	-0,4	4	0,428571429
10	0,25	0,25	0,25	2	1	0,4	-0,4	2	0,3
11	0,5	0,5	0,5	1	1	0	0	1	0
12	0,5	0	0	4	2	0	-0,4	3	0,088888889
13	0,25	0,5	0,5	3	2	0,4	0,4	2	0,428571429
14	0,25	0	0	1	1	0	0	2	0
15	0,5	0	0,75	4	4	0,8	0,8	4	0,733333333
16	0,25	0,5	0,5	2	2	0,4	0	2	0,2
17	0,75	0,25	0,5	2	4	0	0,4	3	0,266666667
18	0,75	0,5	0,5	1	1	-0,4	0	1	0,2
19	0,5	0,5	0,5	1	1	0	0	0	0
20	0,5	0,5	0,5	2	2	-0,4	-0,4	2	0,4
21	0,5	0,5	0,5	2	1	0,4	0,8	3	0,4
22	0,5	0,75	0,75	2	1	0,4	-0,4	3	0,35
23	0,75	0,5	0,25	1	1	0,4	0,4	5	0,157142857
24	0,5	0,5	0,25	2	1	0,4	0	2	0,24
25	0,25	0,5	0,5	1	1	0,8	0,4	2	0,45
26	0,75	0,75	0,75	2	2	0,4	0,4	4	0,35
27	0,75	0,75	0,75	2	2	0,4	0,4	2	0,466666667
28	0,25	0,5	0,5	3	3	0,4	0,4	2	0,45
29	0,75	0,75	0,75	3	3	0,4	0,4	4	0,42
30	-0,25	0	0,75	1	1	0	0	3	0
31	0,5	0	0,75	3	3	0,8	0,4	3	0,5
32	0,75	0,75	0,75	1	2	0,4	0,8	3	0,583333333

regions. On the other hand, a respondent who answers that his/her political opinions do not belong to any region is an agent that may be influenced by other unconventional sources. At the same time, questions 9-13 are designed in order to capture exposure to sources only from the conventional regions, so other regions are not examined. By removing these incompatible values (with index numbers 7, 11, 12, 13, 15, 17, 19 and 22) from the dataset, we expect to get a clearer picture. We also remove three more values (with index numbers 2, 18 and 23), which are also incompatible but for different reasons. Respondent 2 reports the extreme $\lambda_0 = -0.25$, while his/her parents' opinions are $\lambda_1 = \lambda_2 = 0.5$ and the frequency of political discussions with them is very often ($K_1 = K_2 = 4$). The conclusion is that he/she is heavily influenced by a friend or other source that we do not include in the analysis. Respondent 18 reports a very strong positive opinion ($\lambda_0 = +0.75$), but at the same time political discussions with family are rare ($K_1 = K_2 = 1$) and the respondent has only one right answer; the influence boosting the respondent's very strong opinion probably comes from sources other than the parents or the dynamic group as we captured it. Finally, respondent 23 is the only respondent that has all five answers right. This means that a clustering of sources into three groups would probably be preferable, since one static group alone (just with parents, with low values for K_1, K_2, e_1 and e_2) does not generate enough intensity of influence to justify the 5-right-answers measurement, according to assumption H_3. There should be another static group as well, possibly containing friends or teachers. Or equivalently, the intensity of the influence generated by the dynamic group is not sufficiently matched by the intensity of the influence generated by one static group, therefore another static group is probably missing.

Finally, let us explain how the values of λ_0 are treated. Typically, λ_0 is a variable with continuous values between -1 and 1. In the questionnaire, however, Question 1 is posed in such a way that it quantifies an ordinal variable. So, λ_0 can be treated (perhaps more appropriately) as an ordinal variable as well. Depending on the type of variable, the statistical tools for the evaluation of the model are different. We present the results about the effectiveness of the model in predicting the respondents' opinions, for both types of variable, in table 2.

Table 2. Measures of association between the observed λ_0 and the predicted λ_0, for the first and the second group of respondents

			First group (32 respondents)	Second group (21 respondents)
Observed λ_0 as continuous variable	Linear regression	Slope	0.3109	0.6560^{**}
		Intercept	0.3636^{***}	0.2038
	Pearson correlation		0.2864	0.6049^{**}
	RMSE		0.3190948	0.2246345
Observed λ_0 as ordinal variable	Polyserial correlation		0.3223771	0.6530366

Note: The p-value is calculated for the linear regression coefficients (slope and intercept) and the Pearson correlation coefficient, ** < .01 *** < .001.

4.3 *Discussion*

Let us analyze the results reported in table 2. Predicted λ_0 is a continuous variable. Observed λ_0, on the other hand, can be treated as a continuous variable with a limited number of possible values, namely 0.75, 0.5, 0.25, 0 and -0.25, or as an ordinal variable with ordered categories as values, namely "very important", "just important", "just an obligation", "indifferent" and "against". Considering two continuous variables enables us to examine the association between the observed-predicted pairs in the following three ways: by regressing the observed values on the predicted values, by calculating the Pearson correlation and by calculating the root mean square error (RMSE). However, considering one continuous and one ordinal variable, an appropriate statistical tool is polyserial correlation.

In order to evaluate the model for its predictive accuracy, the desired results from the linear regression should be a value close to 1 for the slope and a value close to 0 for the intercept (so that the regression line is close to the 1: 1 line), combined with a value close to 1 for the Pearson correlation coefficient of the observed λ_0 and the predicted λ_0. For RMSE, a small value is desired. And as with the Pearson correlation, we want a value close to 1 for the polyserial correlation. However, the goal is not to report high levels of predictive accuracy. High levels of accuracy would require more information (a larger number of questions in the questionnaires). Instead, one of our objectives was to describe opinion formation under social influence with as few questions as possible. Also, the research involves an element of arbitrariness as far as the K_i are concerned. The theoretical values of the various K_i are not known (this would require knowing the value of the corresponding $|A_c \cap A_{p_i}|$), so we assign relative values to the K_i according to the rules of the theory, but still an element of arbitrariness cannot be avoided.

Therefore, we do not expect high levels of predictive accuracy, but we still want to assess whether the model is able to capture and interpret the processes of influence. In other words, we want to assess the soundness of the theory (and the assumption included) with which the model is constructed. The way to do this is by comparing the predictive effectiveness of the model for the first group of respondents with the predictive effectiveness of the model for the second group. As we know, the second group is formed by removing the incompatible cases from the first group. The incompatible cases are those cases for which the additional necessary information is not available, because the questionnaire did not include the additional necessary questions. In other words, the theory-model should be able to capture and predict these cases with more confidence, if additional information was available. For example, questions 9-13 provide information for sources in the dynamic group, but only for those from the conventional political regions. A respondent whose political opinions do not belong to any region is possibly influenced by sources from unconventional political regions and so, necessary information for the model must be missing. Also, when the data shows that another static group must be present, then an unsuitable version of assumption H_3 is used (we always use (11) given the available information, but in these cases (12) is the appropriate choice). For all these reasons, we distinguish between compatible and incompatible cases.

We expect a greater effectiveness of the model in predicting opinions in the second group. A much greater effectiveness would mean that missing information is important or in other words, that the model produces significantly improved predictions when the necessary data is provided.

We see that all measures show a much greater effectiveness of the model in predicting opinions in the second group. The value for the slope is 0.6560 and it is statistically significant, while the value for the intercept is not statistically significant. The same measures for the first group give a value of 0.3109 for the slope and a statistically significant value of 0.3636 for the intercept. The Pearson correlation coefficients are 0.6049 (statistically significant) and 0.2864 (not statistically significant), for the second and first group respectively. Similarly improved results are reported for the polyserial correlation. Finally, RMSE gives a value (error) of 0.2246345 for the second group, which is smaller than the value of 0.3190948 for the first group, exactly as desired.

As an overall conclusion, we could say that parents are crucially important when it comes to social influence (and especially for political topics) and often are key members of the static group. However, a more detailed collection of information, including friends, teachers, hobbies *etc.*, is necessary in order to provide the model and the equations in assumption H_3 with all the input needed and therefore, all cases to be captured and predicted.

5 Conclusions

In this paper, we presented a novel way of describing social influence on opinion formation. The theory is based on the alphabet of human thoughts, a philosophical concept that dates back to Descartes and Leibniz years (with influences even dating back to earlier years). The alphabet of human thoughts is the reason behind the novelty of our approach, but it is a major limitation at the same time. It has been characterized as "absurdly optimistic" [44]. On the other hand, the great mathematician Kurt Gödel believed that it was feasible [45]. Whatever the case may be, a practical yet universal version of it remains elusive.

The version of the alphabet of human thoughts that we consider in this paper is of paramount importance in theoretically constructing the model. This version of the alphabet theoretically involves an infinite number of simple ideas $(s_1, s_2, s_3, ...)$. The semantic contents of the simple ideas may not be known, but we can still use them in simulations. In particular, we can consider a large collection of distinct simple ideas $(s_1, s_2, \dots, s_T$ where T is as large as required for the simulation under consideration). In that way, the various social situations and contexts (studied in simulations) could be flexibly represented by appropriately constructing the various A_{p_i} and the A_c, using these simple ideas[14]. For example, we could specify the extent to which the agent is informed about the topic (K_0) or how general we want a topic to be by assigning an accordingly large number of simple ideas (let us remember the definition of a general topic c, involving $|A_c|$, that was mentioned in the end of chapter 3.2 after assumption H_3). So, simple ideas could be used as sociological "particles" in simulations and the corresponding parameters could be deterministically calculated. In applications with real data, however, other approaches must be followed; such approaches should probably be more application-specific, as we see in the application of this paper.

[14] For a simplified simulation design, we could construct the various A_{p_i} and the A_c as sets (instead of multisets that we consider in our theory).

Furthermore, more evaluations must be performed to test the validity of the assumption H_3 and of the theory in general. Evaluations on larger scales should involve sources of social influence in numbers as large as possible and high levels of predictive accuracy should be the goal.

6 Appendix

We present the questionnaire:

Question 1 — Multiple Choice
Σχετικά με τις τελευταίες εθνικές εκλογές (7 Ιουλίου 2019), τι ισχύει από τα παρακάτω για εσάς?
Θεώρησα ότι είναι πολύ σημαντικές εκλογές
Θεώρησα ότι είναι απλά σημαντικές εκλογές
Θεώρησα ότι είναι απλά μια υποχρέωση του κάθε πολίτη
Μου ήταν αδιάφορες ως εκλογές
Είμαι γενικά αντίθετος με τον θεσμό των εκλογών

Question 2 — Multiple Choice
Σχετικά με τις τελευταίες εθνικές εκλογές (7 Ιουλίου 2019), τι ισχύει από τα παρακάτω για τον πατέρα σας?
Ο πατέρας μου θεώρησε ότι είναι πολύ σημαντικές εκλογές
Ο πατέρας μου θεώρησε ότι είναι απλά σημαντικές εκλογές
Ο πατέρας μου θεώρησε ότι είναι απλά μια υποχρέωση του κάθε πολίτη
Του ήταν αδιάφορες ως εκλογές
Ο πατέρας μου είναι γενικά αντίθετος με τον θεσμό των εκλογών
Δεν ξέρω την άποψή του

Question 3 — Multiple Choice
Σχετικά με τις τελευταίες εθνικές εκλογές (7 Ιουλίου 2019), τι ισχύει από τα παρακάτω για τη μητέρα σας?
Η μητέρα μου θεώρησε ότι είναι πολύ σημαντικές εκλογές
Η μητέρα μου θεώρησε ότι είναι απλά σημαντικές εκλογές
Η μητέρα μου θεώρησε ότι είναι απλά μια υποχρέωση του κάθε πολίτη
Της ήταν αδιάφορες ως εκλογές
Η μητέρα μου είναι γενικά αντίθετη με τον θεσμό των εκλογών
Δεν ξέρω την άποψή της

Question 4 — Rating Scale

Πόσο συχνά συζητάτε για πολιτικά θέματα με τον πατέρα σας?

Σπάνια (1)

Κάποιες φορές (2)

Συχνά (3)

Πολύ συχνά (4)

Question 5 — Rating Scale

Πόσο συχνά συζητάτε για πολιτικά θέματα με τη μητέρα σας?

Σπάνια (1)

Κάποιες φορές (2)

Συχνά (3)

Πολύ συχνά (4)

Question 6 — Rating Scale

Πόσο συμφωνείτε/διαφωνείτε στα πολιτικά θέματα με τον πατέρα σας?

Συμφωνώ στα περισσότερα με τον πατέρα μου (1)

Συμφωνώ σε περισσότερα από όσα διαφωνώ με τον πατέρα μου (2)

Ούτε συμφωνώ ούτε διαφωνώ με τον πατέρα μου (3)

Διαφωνώ σε περισσότερα από όσα συμφωνώ με τον πατέρα μου (4)

Διαφωνώ στα περισσότερα με τον πατέρα μου (5)

Question 7 — Rating Scale

Πόσο συμφωνείτε/διαφωνείτε στα πολιτικά θέματα με τη μητέρα σας?

Συμφωνώ στα περισσότερα με τη μητέρα μου (1)

Συμφωνώ σε περισσότερα από όσα διαφωνώ με τη μητέρα μου (2)

Ούτε συμφωνώ ούτε διαφωνώ με τη μητέρα μου (3)

Διαφωνώ σε περισσότερα από όσα συμφωνώ με τη μητέρα μου (4)

Διαφωνώ στα περισσότερα με τη μητέρα μου (5)

Question 8 — Multiple Choice

Ας θεωρήσουμε τον κλασικό διαχωρισμό των πολιτικών απόψεων σε αριστερές-κεντρώες-δεξιές. Τι ισχύει για εσας?

Οι πολιτικές μου απόψεις ανήκουν ξεκάθαρα σε έναν από τους τρεις χώρους

Οι πολιτικές μου απόψεις ανήκουν σε παραπάνω από έναν χώρους

Οι πολιτικές μου απόψεις δεν ανήκουν πουθενά

Question 9 — Multiple Choice

Πόσα κόμματα βρίσκονται σήμερα στη Βουλή?

4

5

6

7

Question 10 — Multiple Choice

Ποιοι είναι οι δύο αντιπρόεδροι της Νέας Δημοκρατίας?

Κωστής Χατζηδάκης, Παναγιώτης Πικραμμένος

Άδωνις Γεωργιάδης, Παναγιώτης Πικραμμένος

Κωστής Χατζηδάκης, Άδωνις Γεωργιάδης

Κανένα δίδυμο από τα παραπάνω

Δε γνωρίζω να πω με σιγουριά

Question 11 — Multiple Choice

Ποια από τις παρακάτω εφημερίδες είναι κομματικό όργανο του ΣΥΡΙΖΑ?

Το Documento

Ο Ριζοσπάστης

Η Εφημερίδα των Συντακτών

Καμμία εφημερίδα από τις παραπάνω

Δε γνωρίζω να πω με σιγουριά

Question 12 — Multiple Choice

Ποιος ήταν υπουργός εξωτερικών όταν υπεγράφη η συμφωνία των Πρεσπών?

Νίκος Δένδιας

Νίκος Κοτζιάς

Γιώργος Κατρούγκαλος

Κανένας από τους παραπάνω

Δε γνωρίζω να πω με σιγουριά

Question 13 — Multiple Choice

Στα πόσα ευρώ έχει διαμορφωθεί σήμερα ο κατώτατος μισθός?

585

650

710

Κανένα από τα παραπάνω

Δε γνωρίζω να πω με σιγουριά

7 References

[1] Ioannidis E., *Modeling Knowledge Networks*, PhD Thesis, Aristotle University of Thessaloniki, School of Mathematics, 2018.

[2] Helbing D., Opinion Formation Models, *Quantitative Sociodynamics*, Theory and Decision Library (Series B: Mathematical and Statistical Methods), vol 31., Springer, Dordrecht, 1995.

[3] Sobkowicz P., "Modelling Opinion Formation with Physics Tools: Call for Closer Link with Reality", *Journal of Artificial Societies and Social Simulation* 12(1)11, 2009.

[4] DeGroot M. H., "Reaching a Consensus", *Journal of the American Statistical Association*, 69, pp. 118-121, 1974.

[5] Friedkin N. E. and Johnsen E. C., "Social influence and Opinions", *J. Mathematical Sociology*, 15(3-4), pp. 193–205, 1990.

[6] Friedkin N., Johnsen E., *Social Influence Network Theory*, Cambridge Univ. Press, New York, 2011.

[7] Abelson R., Mathematical models of the distribution of attitudes under controversy, N. Frederiksen, H. Gulliksen (Eds.), *Contributions to Mathematical Psychology*, Holt, Rinehart & Winston, Inc, New York, pp. 142-160, 1964.

[8] Taylor M., "Towards a mathematical theory of influence and attitude change", *Human Relations* 21 (2), 121-139, 1968.

[9] Hegselmann R. and Krause U., "Opinion Dynamics and Bounded Confidence Models, Analysis, and Simulation", *Journal of Artificial Societies and Social Simulation*, vol. 5, pp. 1-33, 2002.

[10] Deffuant G., Neau D., Amblard F. and Weisbuch G., "Mixing Beliefs among Interacting Agents", *Advances in Complex Systems*, vol. 3, pp. 87-98, 2000.

[11] Clifford P. and Sudbury A., "A Model for Spatial Conflict", *Biometrika*, vol. 60, pp. 581-588, 1973.

[12] Holley R. and Liggett T.M., "Ergodic Theorems for Weakly Interacting Systems and the Voter Model", *Annals of Probability*, vol. 3, pp., 643-663, 1975.

[13] Kasthurirathna D., Harre M., Piraveenan M., "Optimising influence in social networks using bounded rationality models", *Soc. Netw. Anal. Min.*, 6:54, 2016.

[14] Kempe D., Kleinberg J. and Tardos E., "Maximizing the spread of influence through a social network", Proceedings of the ninth ACM SIGKDD International

Conference on Knowledge Discovery and Data Mining, ACM, pp. 137–146, 2003.

[15] Acemoglu D. and Ozdaglar A., "Opinion Dynamics and Learning in Social Networks", *Dynamic Games and Applications*, vol. 1, pp. 3-49, 2011.

[16] Wasserman S. and Faust K., *Social Network Analysis-Methods and Applications* (Structural Analysis in the Social Sciences), Cambridge Univ. Press, 1994.

[17] Couturat L., *La Logique de Leibniz*, chpts. 3,4, Paris: Felix Alcan, 1901.

[18] Rescher N., "Review of On the Project of a Universal Character by Jonathan Cohen", *Journal of Symbolic Logic* 19: 133, 1954.

[19] Woltzenlogel Paleo B., *Leibniz's Characteristica Universalis and Calculus Ratiocinator Today*, 2016.

[20] Parkinson G. H. R., ed. and trans., *Leibniz: Logical Papers*, Oxford Univ. Press., 1966.

[21] Loemker L., ed. and trans., *Leibniz: Philosophical Papers and Letters*, 165–66, 192–95, 221–28, 248–50, and 654–66, Synthese Historical Library, Dordrecht: D. Reidel, 1969.

[22] Rutherford D., "Philosophy and language" in Jolley N., ed., *The Cambridge Companion to Leibniz*, Cambridge Univ. Press., 1995.

[23] Couturat L., *Opuscules et Fragments Inédits de Leibniz*, 1903.

[24] Geiger R. A. & Rudzka-Ostyn, B. (Eds.), *Conceptualizations and Mental Processing in Language*, Berlin, Boston: De Gruyter Mouton, 2011.

[25] Hardy G. H. and Wright E. M., *An Introduction to the Theory of Numbers*, 5th ed. Oxford, England: Clarendon Press, pp. 3 and 21, 1979.

[26] du Sautoy M., *The Music of the Primes: Searching to Solve the Greatest Mystery in Mathematics*, Harper Collins. p. 23, 2003.

[27] Dunbar RIM, "The social brain hypothesis", *Evol Anthr* 6(5):178–190, 1998.

[28] Hill R. A. & Dunbar R. I. M., "Social network size in humans", *Hum. Nature* 14, 53–72, 2003.

[29] Bernard HR, Killworth PD, "On the social structure of an ocean-going research vessel and other important things", *Soc Sci Res* 2(2):145–184, 1973.

[30] Sutcliffe A, Dunbar RIM, Binder J, Arrow H, "Relationships and the social brain: Integrating psychological and evolutionary perspectives", *Br J Psychol* 103(2):149–168, 2012.

[31] Roberts SGB, Dunbar RIM, "Communication in social networks: Effects of kinship, network size, and emotional closeness", *Pers Relatsh* 18:439–452, 2011.

[32] Quintelier E., "Intergenerational transmission of political participation intention", *Acta Polit* 50, 279–296, 2015.

[33] Dinas E., "Why Does the Apple Fall Far from the Tree? How Early Political Socialization Prompts Parent-Child Dissimilarity", *British Journal of Political Science*, vol. 44, no. 4, pp. 827–852, 2014.

[34] Zaller J., *The Nature and Origins of Mass Opinion*, Cambridge, United Kingdom: Cambridge University Press, 1992.

[35] Wanders F.H.K., Dijkstra A. B., Maslowski R., van der Veen I. & Amnå E., "The Role of Teachers, Parents, and Friends in Developing Adolescents' Societal Interest", *Scandinavian Journal of Educational Research*, 2020.

[36] McIntosh H., Hart D., Youniss J., "The Influence of Family Political Discussion on Youth Civic Development: Which Parent Qualities Matter?", *Political Science and Politics*, July, 2007.

[37] Jennings K., Stoker L., Bowers J., "Politics across generations: family transmission re-examined", *The Journal of Politics*, 71, 782–799, 2009.

[38] Gettier E. L., "Is Justified True Belief Knowledge?", *Analysis*, 23 (6): 121–123, 1963.

[39] Scheffler I., *Conditions of Knowledge: An Introduction to Epistemology and Education*, Chicago: Scott, Foresman, 1965.

[40] Boghossian P., *Fear of Knowledge: Against relativism and constructivism*, Oxford, UK: Clarendon Press, Chapter 7, p 95-101, 2007.

[41] Kirkham R. L., "Does the Gettier Problem Rest on a Mistake?", *Mind*, 93 (372): 501–513, 1984.

[42] Pollock J. L., Cruz J., *Contemporary Theories of Knowledge*, (2nd ed.), Rowman & Littlefield Publishers, 1999.

[43] Plantinga A., *Warrant: The Current Debate*, Oxford University Press, 1992.

[44] Parkinson G. H. R., Mary Morris, trans., *Leibniz: Philosophical Writings*, London: J M Dent, 1973.

[45] Dawson, Jr., J. W., *Logical Dilemmas: The Life and Work of Kurt Gödel*, Wellesley MA: A. K. Peters, 1997.