In search of \textit{Homo Sociologicus}

\textbf{Dissertation Proposal}

written by

Xue, Yun Qi

under the supervision of \textbf{Prof. Rohit Parikh}, and submitted to the Graduate Faculty in Computer Science in partial fulfillment of the requirements for the degree of

\textbf{Ph.D. in Computer Science}

at the \textit{The City University of New York}.

Date of the exam: 04/09/2015

Members of the Thesis Committee:

Prof. Shweta Jain
Prof. Delaram Kahrobaei
Abstract

The subject of my thesis is to build an epistemic logic system that is able to show the spreading of knowledge and beliefs in a social network that contains multiple subgroups. Epistemic logic is the study of logical systems that express mathematical properties of knowledge and belief. In recent years, there have been increasing number of new epistemic logic systems that are focused on community properties such as knowledge and belief adoption among friends.

In addition to previous work, this study will allow multiple languages. It enables a sub-culture within a network of agents. It also shows how knowledge and belief changes within a subject. For example, when a group of people is using a particular sub-language to adjust their beliefs about politics, their beliefs about weather may very well be intact.

My dissertation includes three major sections, expert influence, distributed common knowledge, and simulations. Some preliminary work from the first two sections has been presented to two conferences.
Contents

1 Introduction ............................................ 3

2 Rational Choice Theory .................................. 6
  2.1 The Theory ........................................... 6
  2.1.1 Preferences ...................................... 6
  2.1.2 Utility ........................................... 7
  2.1.3 Choice ............................................ 8
  2.1.4 Expected Utility ................................ 10
  2.2 Empirical Results from Experimental Economics ...... 13
    2.2.1 Public Goods Game ................................ 13
    2.2.2 Ultimatum Game ................................ 16

3 Social Outliers - Autism ................................ 17
  3.0.3 Introduction .................................... 17
  3.0.4 Theory of Mind .................................. 18
  3.0.5 The Affective Foundation Theory .................. 19
  3.0.6 Mirror Neurons .................................. 19
  3.0.7 Discussion ...................................... 20

4 Signaling ............................................... 22
  4.0.8 Sender-Receiver Model ............................ 22
  4.0.9 The Lewis-Skyrms Signaling Game ................. 23
  4.0.10 Information in ‘Signals’ ......................... 24

5 Social Software: From Social Procedure to Homo Sociologicus 26
  5.1 Conclusion and Discussion .......................... 27
    5.1.1 The Jigsaw Puzzle: in Search of Homo Sociologicus .. 27
    5.1.2 Possible Future Direction: Defining and Modeling Social Signals? .................................. 28

6 Knowledge and Epistemic Logic ........................ 29
  6.1 Knowledge .......................................... 29
  6.2 Formal Frameworks ................................ 29
  6.3 Friends’ Influence .................................. 30
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.4   Language Splitting</td>
<td>31</td>
</tr>
<tr>
<td>6.4.1 P Axioms</td>
<td>32</td>
</tr>
<tr>
<td>7     Network Science</td>
<td>34</td>
</tr>
<tr>
<td>7.1   Networks</td>
<td>34</td>
</tr>
<tr>
<td>7.2   Basic Concepts</td>
<td>36</td>
</tr>
<tr>
<td>7.3   The Scale-Free Property</td>
<td>38</td>
</tr>
<tr>
<td>8     Proposed Work</td>
<td>41</td>
</tr>
<tr>
<td>8.1   Introduction</td>
<td>41</td>
</tr>
<tr>
<td>8.2   Expert Influence</td>
<td>43</td>
</tr>
<tr>
<td>8.3   Distributed Common Knowledge</td>
<td>44</td>
</tr>
<tr>
<td>8.4   Simulations of Knowledge Influence</td>
<td>45</td>
</tr>
<tr>
<td>9     Time-line</td>
<td>46</td>
</tr>
<tr>
<td>Bibliography</td>
<td>46</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

What is a social man (homo sociologicus)?

With this question in mind, we started the journey in empirical studies such as experimental economics. Later it extended to theoretical and philosophical areas on the related topics of homo sociologicus (as contrasted with homo economicus).

The idea of homo economicus was originated by the prominent 19th-century scholar, John Stuart Mill [61]. In his work in 1836 [41], “...does not treat the whole of man’s nature as modified by the social state, nor of the whole conduct of man in society...”. Such a self-interested idea of man is also associated with the founding father of modern economics, Adam Smith, in his famous work ‘The Wealth of Nations [79]’. He wrote: “It is not from the benevolence of the butcher, the brewer, or the baker that we expect our dinner, but from their regard to their own interest”. Later in 1881, Edgeworth even endorsed the idea as “the first principle of Economics” [17]. The idea of homo economicus has been a major influence on economic theories and models ever since [75].

However, Sen points out the absurdity of such idea in his widely cited paper, “Rational Fools: A Critique of the Behavioral Foundations of Economic Theory” in 1977 [75]. He gives an amusing example to illustrate the problem of Homo Economicus assumption: “Where is the railway station?” he asks me. “There,” I say, pointing at the post office, “and would you please post this letter for me on the way?” “Yes,” he says, determined to open the envelope and check whether it contains something valuable.

While Sen agrees with Edgeworth’s first principle in a strictly defined context, he also points out that we cannot forgo the complex psychological issues beneath choices and calls for “actually testing”, which was indeed done by experimental economists years later. With the flourishing of behavioral and experimental economics, researchers are able to bring a vast amount of data on the table to show that in many situations people do not just act on self-interest [63] [21]. The most notable publication, which is cited 1238 times to this date since 2001, is “In search of homo economicus: behavioral experiments in 15 small-scale

1Social human and Economic human respectively
societies” [28]. Their results not only show the same kind of deviation from the predications based on the homo economicus assumption, but also show that economic decisions are closely related to social/group activities. Around the same period, a group of economists started calling for defining social preferences [20] [19].

It is clear that the homo economicus assumption works only within a tightly restricted economic context. But what about homo sociologicus? The term was introduced by German sociologist Ralf Dahrendorf in 1958 [13]. Here we will borrow this term and examine what it means to be a social man within economic theories and framework. There is already a branch of studies in decision theory focusing on social procedures and decision making in groups [60] [70]. However, what we want to focus is on the individual level of decision making with the assumption of a social man. Therefore, in this paper we will review areas that can help us define a rational social man.

Any topics related to society are almost guaranteed to be interdisciplinary studies, so is this one. We will first look at some related empirical results from experimental economics, namely from Public Goods Game (PGG) and Ultimatum Game (UG). Both games have social elements and both empirical results deviate from the theoretical predictions. Then we will review the fundamentals of classic rational choice theory, which serve as our theoretical foundation in economics. Inevitably, when we study real human decision making, we come across bounded rationality [78], which assumes limited computational power of real people and limited information accessibility. So what are the consequences of bounded rationality in a social man? We would like to know how rationality is bounded socially as in the results found by Bowles at el in [28]. We want to find out whether there are social signals that help decision makers make social-related decisions. Therefore we will dig deeper to review literature for signals. Another important and related area is the psychology of reasoning, which gives firm understanding of rationality. In this area, we will also review the literature on one particular kind of subject, autistic people, who are considered to have relatively low social intelligence. However, some recent publications [25][2][67] show that autistic people can reason about social situations, but not in the form of what normal people possess.

In view of these non-computer-science-related fields, we have a broad understanding of current research on the matter of social interactions. Therefore, we will put these literature reviews in one chapter as the background.

We can summarize a few issues from these studies. Firstly, how to capture the multi-cultural aspects of our social interactions? Here the multi-culture can be from different countries, or from different sub-groups within a society. We will see in the experimental studies of some social games such as PGG and UG that the results are largely dependent on culture, unlike the theoretical predictions from classical rational choice theory. Secondly, on which level of social...
interactions can we computer scientists, more specifically logicians help modeling? Through the reviews of signaling and autism, it is not hard to discover that although autistic people send unusual social signals, in many cases they can still acquire social knowledge.

Therefore, we conclude that we can build a system that captures the knowledge dynamics in social interaction with technical capacity of accommodating multiple cultures. Hence, in chapter 3, we further study dynamic epistemic logic including its latest development on friends’ influence in a network based community. In additionally, we review some technical work from Parikh [53] that allows multiple languages in belief changes. We intend to simulate our new framework and compare the simulation with real network observed. So in chapter 4 we will briefly review some network science. Chapter 5 is proposed research directions. We conclude this proposal with the time-line of the dissertation.
Chapter 2

Rational Choice Theory

2.1 The Theory

This part of review is mostly following the related topics in *Lecture Notes in Microeconomic Theory* by Ariel Rubinstein [65].

2.1.1 Preferences

Basic Definitions

We define that a variety of options\(^1\) as a finite set \(X\). We then further define a binary relation \(\succeq\) that is a collection of ordered pairs of elements from \(X\). For example, when \(x, y \in X\), we can have \((x, y) \in \succeq\), which can also be denoted as \(x \succeq y\). It means option \(x\) is seen at least as good as option \(y\).

This binary relation also includes two additional definitions: one is symmetric, *indifferent*: \(\sim\), and the other is asymmetric *strictly better*: \(\succ\). For two elements \(x, y \in X\), \(x \sim y \iff [(x \succeq y) \text{ and } (y \succeq x)]\). While \(x \succ y \iff [(x \succeq y) \text{ but not } (y \succeq x)]\).

Axioms

*Completeness*: For any two options \(x, y \in X\), \(x \succeq y\) or \(y \succeq x\).

This axiom says that a decision maker can always choose between two options. Although considering the definition of the binary relation \(\succeq\) and its related definitions, it is clear that the completeness axiom allows three situations for any two options. First, when \(x \succ y\), it means that the decision maker always chooses option \(x\) over \(y\). Similarly, the second situation is when \(y \succ x\), the decision maker chooses \(y\). The last situation is when \(x \sim y\), which means the decision maker chooses \(x\) and \(y\) at random\(^2\).

\(^1\)In this review paper, we use ‘options’ interchangeably with ‘alternatives’.

\(^2\)As how Gilboa put it in his book ‘Rational Choice’
Transitivity: For any three options \( x, y, z \in X \), if \( (x \succeq y \text{ and } y \succeq z) \), then \( x \succeq z \).

This axiom prevents cyclical preferences in both individual and group decision making situations.

2.1.2 Utility

When we compare two options \( x, y \in X \), we often say that we prefer one of them. For example, 'I prefer higher grades'. This can be written in following form:

\[
x \succeq y \text{ if } V(x) \geq V(y)
\]

Here \( V : X \rightarrow \mathbb{R} \) is a function that assigns a real number to each element in \( X \). In our example of grades, it has a clear numerical representation. But we also want to represent sets of options that do not have clear numerical evaluation. For example, Jill prefers Thai food over Japanese food because she likes spicy food, i.e. the spicier food has a higher value to her. Ideally we want to define a function \( U : X \rightarrow \mathbb{R} \) that represents the binary relation \( \succeq \) if for any two elements \( x, y \in X, x \succeq y \Leftrightarrow U(x) \geq U(y) \). We call this function a utility function.

To show the existence of such utility function, we will look into two situations, namely when \( X \) is finite, and When \( X \) is continuous [62].

Finite Space

When the set \( X \) is finite, a utility function that represents \( \succeq \) relationship always exists. To prove that, we can first show that any subset of \( X \) has a minimal element\(^3\) through induction on the size of subsets given \( X \) is complete and transitive. Then we can prove following proposition.

Proposition 1. For a finite set \( X \), the binary relation \( \succeq \subseteq X \times X \) that is complete and transitive, has a utility representation with natural numbers.

Proof. Given a finite set \( X \), we can define \( X_1 \) as a subset of \( X \) and contains all the minimal elements of \( X \). Then we further define \( X_2 \) as a subset of \( X - X_1 \) with all the minimal elements in \( X - X_1 \). We can keep construct such minimal subsets till we have \( X = X_1 \cup X_2 \cup X_3 \ldots X_k \), and \( k \leq |X| \). We then define \( U(x) = k \) if \( x \in X_k \). Further more, when \( a \succeq b \) and \( a, b \in X \), we know \( U(a) > U(b) \) and \( a \not\in X_1 \cup X_2 \cup X_3 \ldots X_{U(b)} \). When \( a \succ b \), \( U(a) = U(b) \).

Continuous Space

Often in economics the set \( X \) is set to be an infinite subset of a Euclidean space, \( \mathbb{R}^n \). We want to show that often there is a utility representation in such case too. Continuity requires that for two options \( a, b \in X \), if \( a \) preferred over \( b \), then the “neighboring” elements around \( a \) should be preferred over the “neighboring” elements around \( b \). To formalize this, we will start with some definitions.

\(^3\)Or minimal elements, if they are equivalent.
Definition 2. Let \( a \) be an element in \( X \). For the set of all points that have distance less than \( r \) \((r > 0)\) from \( a \), we call it a ball around \( a \), and denote it as \( \text{Ball}(a, r) \).

Definition 3 (Continuous a). We call a preference relationship \( \succeq \) on \( X \) continuous if whenever \( a \succ b \), there are balls \( B_a \) and \( B_b \) such that \( x \succ y \) for all \( x \in B_a \) and \( y \in B_b \).

Definition 4 (Continuous b). We call a preference relationship \( \succeq \) on \( X \) continuous when the set \( \{ (x, y) | x \succeq y \} \subseteq X \times X \) is a closed set. In other words, for all \( n \) and \( a_n, b_n \in X \) such that \( a_n \to a \) and \( b_n \to b \), we have \( a \succeq b \).

A preference relation \( \succeq \) on \( X \) satisfies ‘Continuous a’ if and only if it satisfies ‘Continuous b’.

Debreu [15] proved a famous theorem that pushes a step further.

Theorem 5 (Debreu’s Theorem). For a continuous preference relationship \( \succeq \) on \( X \), there exists a continuous utility function \( U(x) : X \to \mathbb{R} \).

2.1.3 Choice

Both utility functions and preference relations are just mental attitude of a person towards a set of options. However, they do not explain how that person actually make a choice in a real life situation. A person can think of a preference ordering for different options, but it does not necessarily mean that he would make a choice accordingly in a real decision problem. Therefore we need to introduce the definition of choice function. Before that we will first look at some preliminary definitions.

We will look at the set of possible alternatives \( X \) again, in which any non-empty subset \( A \) of \( X \) could be a choice problem. Any member \( x \in A \) is a choice. In some situations, the decision maker considers relevant choice problems. We pair the collection of choice problems, \( D \subseteq X \), with \( X \), and call \( (X, D) \) a context. When a context is given, for each particular problem \( A \subseteq D \), a choice function \( C(A) \) outputs a unique element from \( A \) that is the choice of the problem.

Since this chapter is about rational choice theory, we shall discuss what kind of behaviors are considered rational within this theory. Roughly speaking, we consider a decision maker rational when he has a preference relation \( \succeq \) on the set of alternatives \( X \), and facing a choice problem \( A \subseteq D \), he chooses an optimal element in \( A \). In other words, we call a choice function \( C \) rationalizable when \( C(A) = C\succeq(A) \) for any \( A \) in the domain of \( C \).

Next, we will further review an important condition for rationalizable choice functions: condition \( \alpha \) by Sen [74], which is also referred to as Chernoff’s condition [12].

Condition \( \alpha \)

Given two problems \( A \) and \( B \), both in context \( D \), we say that a choice function \( C \) satisfies condition \( \alpha \), if \( A \supseteq B \) and \( C(B) \in A \), then \( C(A) = C(B) \). We define
Definition of Choice Functions

Let $C_\succeq$ be a choice function that always outputs a single most preferred element in $X$ given a preference relation $\succeq$. $C_\succeq$ satisfies condition $\alpha$.

Sometimes it is also called "independence of irrelevant alternatives" and was first introduced by Arrow [7] and Nash [43]. A different way to state it would be: if an alternative $x \in B$ chosen is an element of $A$ of $B$, then $x$ must be chosen from $A$.

Dutch Book Arguments

Dutch Book Argument states that anyone who does not try to maximize a preference relation will not survive. In economics, a decision maker can be Dutch-booked if he or she has intransitive preferences. For example, given three alternatives: A, B, and C, the decision maker, say Tom, has the following preference: $A \succ B$, $B \succ C$, but $C \succ A$. Then someone can take advantage of his by first selling $A$ to Tom for $B + \epsilon$; then selling $B$ to Tom for $C + \epsilon$; then selling $C$ to Tom for $A + \epsilon$. At the end, Tom has paid $3\epsilon$ with nothing in return.

Notes on ‘Alternatives’

When we talked about Dutch Book Argument, we revealed an irrational choosing behavior that preserves intransitive preferences on a set of alternatives. In some situations, the violation of rationality is due to inaccurate or changing specification of alternatives. Now we are going to review a famous dinner example from Luce and Raiffa 1957 [39]. In a restaurant, a customer chooses chicken from the menu with only steak tartare and chicken. At the same time, he chooses steak tartare from the menu with steak tartare, chicken and frog legs. It looks like this customer violates the condition $\alpha$, hence we could consider him irrational with his choices. However, it is possible that he realized the fact that second menu with the frog legs indicates a high level of cooking skills. Making a steak tartare also requires high level of cooking skills. Following such reasoning, you may consider that this customer is actually not that irrational, but actually smart. Rubinstein adds this paragraph to remind us that sometimes the same set of alternatives can have different meaning.

We also should realize the particular reasoning of this customer also implies another condition: he is a new customer. That means even though he has access to his own preferences, he does not have full information about the choices that he could make in this particular restaurant.

In Sen’s 93 paper [76], he also discusses the problem of internal consistency.

Choice Functions

We will continue the definition of choice functions. So far, the functions we discussed have only one solution to every choice problem. It is certainly possible that given a preference relation and a choice problem, there are more than one optimal solutions. Therefore we will explain a further fine-grained definition: choice correspondence.
Given a choice problem $A$, $C(A)$ is a non-empty subset of $A$. It is obvious that the decision maker has to select one element from $C(A)$. Basically $C(A)$ is set of equivalent optimal choices he could select from.

The weak Axiom of Revealed Preference

Samuelson originated the revealed preference approach in 1938 [68]. It was also proposed by Houthakker [32] and by von Neumann and Morgenstern [84]. Arrow adapted it to set-valued choice functions in [6]. Sen provides a systematic treatment of the axiomatic structure of the theory of revealed preference in [77].

**Definition 6.** For any $x, y \in X$, we say $x$ is indirectly revealed preferred to $y$, denoted as $x P^* y$, if only if there is a sequence $z^i$, $i = 0, ..., n$, and $z^1 = x$ and $z^n = y$, such that for all $i$, $z^i P z^{i+1}$.

**Definition 7.** For any $x, y \in X$, we say $x$ is indirectly revealed preferred to $y$ in the wide sense, denoted as $x Wy$, if only if there is a sequence $z^i$, $i = 0, ..., n$, and $z^1 = x$ and $z^n = y$, such that for all $i$, $z^i R z^{i+1}$.

For all $x, y \in X$, we have the following axioms:

1. **Weak Axiom of Revealed Preference (WARP):** If $x \sim P y$, then not $y R x$.

2. **Strong Axiom of Revealed Preference (SARP):** If $x P^* y$, then not $y R x$.

3. **Strong Congruence Axiom (SCA):** If $x W y$, then for any non-empty subset $B$ in choice problem $A$ such that $y \in C(B)$ and $x \in B$, $x$ must also belong to $C(B)$.

4. **Weak Congruence Axiom (WCA):** If $x R y$, then for any $B$ in $A$ such that $y \in C(B)$ and $x \in B$, $x$ must also belong to $C(B)$.

After showing the equivalence of all four axioms mentioned above in [77], Sen raises two important questions: (1) Are the rationality axioms to be used only after stipulating them to be true? (2) Are there reasons to expect that some of the rationality axioms will tend to be satisfied in choices over “budget sets” but not for other choices?

2.1.4 Expected Utility

So far in our description of rational choice theory, we assumed that the decision maker has the information about available options and outcome of the choice.

---

4 Instead of following the discussion from Rubinstein’s book, we will review the concepts from the original paper from Sen on ‘Choice Functions and Revealed Preference’ [77].

5 $x P y$ is equivalent to our earlier notation: $x \succ y$. Sen defines $x P y$ as $x$ is chosen while $y$ is available but rejected.

6 $x R y$ is equivalent to our earlier notation: $x \succeq y$. 

---

13
However in reality, we often do not have exact information about the outcome, and face risks or uncertainties. That is the relationship between actions and outcome is not deterministic.

This aspect is especially crucial for this review paper, since ultimately we want to define choices made in a social environment that consists of many people. Then uncertainty is inevitable. Expected Utility Hypothesis, an idea, which goes as far as 1738 from Daniel Bernoulli [10], provides an important view of how to model uncertainty. In the paper, he wrote:

“Somehow a very poor fellow obtains a lottery ticket that will yield with equal probability either nothing or twenty thousand ducats. Will this man evaluate his chance of winning at ten thousand ducats? Would he not be ill-advised to sell this lottery ticket for nine thousand ducats? To me it seems that the answer is in the negative. On the other hand I am inclined to believe that a rich man would be ill-advised to refuse to buy the lottery ticket for nine thousand ducats.”

“...the determination of the value of an item must not be based on its price, but rather on the utility it yields. The price of the item is dependent only on the thing itself and is equal for everyone; the utility, however, is dependent on the particular circumstances of the person making the estimate. Thus there is no doubt that a gain of one thousand ducats is more significant to a pauper than to a rich man though both gain the same amount.”

“If the utility of each possible profit expectation is multiplied by the number of ways in which it can occur, and we then divide the sum of these products by the total number of possible cases, a mean utility [moral expectation] will be obtained, and the profit which corresponds to this utility will equal the value of the risk in question.”

Bernoulli pointed out the problem of focusing only on monetary term and suggested to use expected utility instead. He also suggested to use logarithm to calculated the utility.

In 1944, Von Neumann and Morgenstern provide a formula their book Theory of Games and Economic Behavior [84]. Later in 1954 [71], Savage introduced an alternative framework, subjective expected utility that was followed by the work from Aumann and Anscombe [5].

Von Neumann-Morgenstern Utility Theorem

Before we review the Von Neumann-Morgenstern Utility Theorem (vNM Theorem), we shall explain some preliminary concepts that extend naturally to the context of uncertainty. In the earlier section, we have defined $X$ as a set of op-

---

7 A ducat was a standard gold coin throughout Europe.

8 For the convenience of readers, we will repeat some basic definition of preference that were introduced earlier.
decomposability, a utility function satisfies completeness, transitivity, substitutability, monotonicity, continuity and Theorem 8.

According to decomposability axiom, \( l_1 \) can be substituted with \( l_2 \) separately. To put it formally: Given \( x_1 \rightarrow x_2 \), and \( x_2 \rightarrow x_1 \) or \( x_1 \sim x_2 \), then he is also indifferent between the two lotteries that contain the outcomes. A lottery on \( X \) is a probability distribution: \([p_1 : x_1; p_2 : x_2; p_3 : x_3; \ldots; p_n : x_n]\). In addition, we have: \( \sum_{i=1}^{n} p_i = 1 \) and all \( p_i \geq 0 \).

In the following paragraphs, we will explain the six axioms that are stated by Von Neumann and Morgenstern [84]: completeness, transitivity, substitutability, monotonicity, continuity, and decomposability.

**Axioms** We have introduced the first two axioms (completeness and transitivity) earlier. Therefore we will just simply state them here.

**Completeness:** \( \forall x_1, x_2 \in X, x_1 \rightarrow x_2 \) or \( x_2 \rightarrow x_1 \) or \( x_1 \sim x_2 \)

**Transitivity:** \( \forall x_1, x_2, x_3 \in X, x_1 \sim x_2 \) and \( x_2 \sim x_3 \) \( \Rightarrow x_1 \sim x_3 \)

**Substitutability:** If two outcomes are indifferent for a decision maker, then he is also indifferent between the two lotteries that contain the outcomes separately. To put it formally: Given \( x_1 \sim x_2 \), \([p : x_1; p_3 : x_3; \ldots; p_n : x_n]\) \( \sim [p : x_2; p_3 : x_3; \ldots; p_n : x_n]\), where \( p + \sum_{i=3}^{n} p_i = 1 \). In other words, the outcome \( x_1 \) can be substituted with \( x_2 \).

**Monotonicity:** For all \( x_1, x_2 \in X, x_1 \rightarrow x_2 \) and \( 1 \geq p > q \geq 0 \) \( \Rightarrow [p : x_1; 1-p : x_2] \rightarrow [q : x_1; 1-q : x_2] \)

**Continuity:** For all \( x_1, x_2, x_3 \in X, x_1 \rightarrow x_2 \) and \( x_2 \rightarrow x_3 \) \( \Rightarrow \exists p \in [0, 1] \) such that \( x_2 \sim [p : x_1; 1-p : x_3] \)

**Decomposability:** We denote \( P_{l_j}(x_i) \) as the probability that \( x_i \) is selected by lottery \( l_j \). The axiom states that if we have two lotteries \( l_1 \) and \( l_2 \) over \( X \), and \( P_{l_1}(x_i) = P_{l_2}(x_i) \) for all \( x_i \in X \), then \( l_1 \sim l_2 \).

An example could be following:
\[
\begin{align*}
    l_1 &= [0.7 : x_1; 0.3 : x_1; 0.7 : x_2] \\
    l_2 &= [0.79 : x_2; 0.21 : x_2; 0 : x_3]
\end{align*}
\]
According to decomposability axiom, \( l_1 \sim l_2 \).

**vNM Theorem**

**Theorem 8.** When a binary preference relation \( \succeq \) on a set of outcomes \( X \) satisfies completeness, transitivity, substitutability, monotonicity, continuity and decomposability, a utility function \( u \) exists and fulfills following two properties:

\[
u(x_1) \geq u(x_2) \text{ if only if } x_1 \succeq x_2
\]

\[
u([p_1 : x_1; p_2 : x_2; p_3 : x_3; \ldots; p_n : x_n]) = \sum_{i=1}^{n} p_i u(x_i)
\]
2.2 Empirical Results from Experimental Economics

As we discussed in the introduction, the idea of homo economicus dated as far back as 1836. Sen has written a theoretical objection to the idea and argued about the exact context in which the idea might work. At the same time, the homo economicus assumption was nurturing another branch of research in economics: game theory, which emerged naturally from rational choice theory, and was heavily tested through experimental economics. Although the first reported experiment was done in 1930 by Thurstone, it was not till mid 90’s that experimental economics reached its prime [64]. Through experiments, economists found many deviations in the real world results from the theory. Their initial conclusion was that real humans are not rational.

There is one particular kind of games in which subjects often behave ‘irrationally’ during experiments. Usually, this kind of game has social factors in the process. The social factors are represented in two different types of scenarios. First type is the number of players. For example, in public goods game, there are often more than 2 players. Second type is the implicit social aspect, such as in ultimatum games. In fact, the deviations of the experimental results in both games brings us back to our goal: in search of homo sociologicus. The crucial question we want to ask is: are the subjects really just being irrational or do our models need to be adapted towards the social assumption? Clearly we are in favor of the latter view. So are Bowles et al [28]. They have conducted a few games with social elements in 15 small-scale societies and show that the deviations are systematic. People are not making random social decisions. But the way they make such decision is closely related to their day to day activities, which are shared by the social members.

Reading through such empirical results gives us both motivation and data for building new rational choice models with considerations of social aspects.

In this chapter, we will review literatures in both public goods game and ultimatum game. We are going to explain the basic ideas and procedures of the two games, the theoretical predications, and the empirical results.

2.2.1 Public Goods Game

The theory of public goods is important for economists, policy makers and international organizations. It provides an insight of market failures, e.g. their mechanism, consumers’ and suppliers’ incentives, which are all essential for a well-functioning society. The rise of international governance, like IMF, World Bank, and UN, tells the awareness of the need of public goods. Besides governments and sociologists, economists are also interested in public goods. Do people treat public goods the same as private goods? What do they do when there is a conflict of interests between public and private goods? In 1980, Mas-Colell [40] published his mathematical approach to public goods theory. In the experiments, subjects have to make decisions about their public account and
Such experiments were conducted by different researchers but in one-shot form as Schneider and his co-authors did [72]. Isaac and his colleagues tried a new experiment [33] with the possibility of repetition. We will call their game the basic game in this review paper. The details of the experiment are explained in the following subsections.

The basic game

Public goods experiments have studied standard voluntary contribution mechanisms where groups have a choice to invest in a private or a common account. The private account gives the subject a return of the exact amount invested in it, while the common account gives each group member a marginal per capita return (MPCR)\(^{10}\). The higher the contribution into the common account, the higher the group payoff. Results have shown that contributions start very high and then decrease over time [33]. Also, since the decision process is repeated, participants become more experienced and free riding becomes a strategy, especially when the MPCR is low [14]. Hence, there exists this social dilemma.

To alleviate the problem of free riding and contribution decreases, studies have implemented different mechanisms to drive the contributions higher. In this section, we will look at three pairs of such variations that are particularly interesting for social concerns, namely punishment versus reward, one versus multiple punishers, and strangers versus partners.

Now let us first look at the most basic game. Isaac and his colleagues conducted 9 different experiments for the paper. We will not explain each one of them in detail, but take the essence of these experiments and tailor our ‘basic game’ for theoretical analysis.

Basic Idea  A group of subjects who participate in this experiment have to make some decisions on their contribution to a public account, and later their total private benefit is determined by a pre-set function. No communication is allowed during the whole experiment.

Procedure  Each participant receives an endowment \(y\). He has to decide how much of \(y\) he is willing to contribute to the public account. Then after the experimenter’s calculation with the payoff function, the participant is informed about his own total payoff of the round. The game is repeated for \(t\) rounds.

Payoff determination  We assume the following payoff function at round \(t\) for participant \(i\), who contributed \(g_i\) to the public account, and there are \(n\) participants in total:

\[
\pi_i^t = y - g_i^t + a \sum_{j=1}^{n} g_j^t
\]

\(^{9}\)Details explained in section 3.1.1.

\(^{10}\)A basic example is explained following paragraphs: Basic Idea, Procedure, and Payoff determination.
Note: $1/n < a < 1$. The total payoff for participant $i$ is just simply: $\sum_{t}^{t} \pi_{i}^{t}$.

**Theory and Prediction** In this subsection, we are mainly looking at the basic game and analyzing from a game theoretical point of view. First of all, we have to assume that the following prediction is for the selfish and rational subjects. There are many papers written about altruistic subjects. One of the earliest such proposals was from Andreoni [4]. We will try to relax this assumption later in the paper.

With that assumption, a subject would behave in such a way to maximize his payoff $\pi_{i}$ in each round of the basic game. Looking at the payoff function mentioned before, we clearly see $\frac{d\pi_{i}}{dp_{i}} = -1 + a$. Given $1/n < a < 1$, we know $-1+a < 0$. Therefore the subject should never contribute to the public account, i.e. $g_{i} = 0$.  

**Variations**

As we have mentioned before, there are many different variations of the public goods game. Here we only focus on some of the variations, which are related to our design.

**Punishment vs. Reward** To many people’s surprise, when reward option is given, it actually does not work as well as punishment. In this case we are able to explain the experimental results from game theoretical perspective.

In Fehr and Gächter’s paper [22] that was influenced by [46], the game with punishment was designed in two stages. The first stage was similar to ‘the basic game’. During the second stage, the participants had opportunities to punish other participants at some cost. It turned out that with punishment, the overall contribution to the public account improved in comparison with the game without the punishment.

Sefton, Shupp and Walker wrote a paper in 2007 about the effect of rewards and sanctions [73]. The experiment with rewards has similar structure as the experiment with punishments in Fehr and Gächter’s paper. The rewards are costly just as punishment is costly, i.e. the rewards are not free. Game theoretically speaking, the punishment creates big threat to the free-riders, even without real punishment. However, in the games with rewards, it works differently. If there are expectations of rewards among the participants, then there have to be rewards, otherwise the existence of possibility of being rewarded would actually dis-encourage the high contributors. Indeed Sefton, Shupp and Walker found such a result [73].

**One punisher v.s. Multiple punishers** In order to understand the effects of different punishment structures, O’Gorman et al (2009) [45] have implemented an experiment with three different conditions that have no punisher, one
punisher and all punishers respectively. Their punishment system was based on a 1 : 3 ratio of the cost of punishing for the punisher to the cost for the target. This ratio facilitated in bringing the group with a single punisher to contribute on high levels and produce a larger profit than the group with all being punishers. Profits were larger in the one punisher case because punishment costs were small and punishment was more coordinated, thus reducing inefficient loss. This shows that uncoordinated punishments made by many people will cause inefficiency and unnecessary losses.

**Strangers vs. Partners** It is already game theoretically surprising to find that people punish free-riders at their own cost. Because given the fact that participants know how many rounds there are, one could always benefit more than the others deviating, i.e. contribute low or 0 at the last round. Applying backward induction, a rational participant should never contribute anything from the start. There are people who apply backward induction in a common real life situation. However, Fehr and Gächter pushed it even further. They assigned the group members randomly for each round so that participants are ‘strangers’ to each other. Then there is clearly no future gain by punish someone, because one cannot even ‘educate’ the free-riders to contribute more in the future round. However, people still punish in this setting.

2.2.2 Ultimatum Game

Ultimatum game (UG) is a widely experimented and considered as ‘one of the simplest’ games in experimental game theory. Güth, Schmittberger and Schwarze first wrote an experimental analysis of this kind of take-it-or-leave-it game in 1982. The experimental results often deviate from the game theoretical prediction. Game theoretically speaking, the proposer should claim as much as possible for himself, while the responder should not reject any non-zero offers for him/her. However in reality, people tend to propose much more for the responders, and responders reject offers up to 30% of the total share. One stream of publication, esp. in neuro-economics, tends to claim that emotional activation is the main reason for such a result. Although such a simple claim could not give detailed explanation on cross-cultural results found by Bowles et al. In the experiments ran in 15 small-scaled rather primitive societies reveal some systematic deviations from the game theoretical results across culture. It indicates a potential sophisticated mechanism beneath the behavior in the simple game.

\[13\] From game theoretical point of view, it should be 0, in order to maximize the gain.

\[14\] In next subsetion more detailed the analysis of the variations is provided.
Chapter 3

Social Outliers - Autism

They reveal interesting facts about reasoning, epistemology, more importantly in relation with social situation.

3.0.3 Introduction

People with autism belong to a special group. They could have average intelligence or higher in some cases, but show some very specific deficits. The most obvious one is their asocial tendency. It was Kanner who started recording a special condition, Autism, among children in 1938 and published it in 1943 [34]. Later Rutter [66] summarized three key features of this condition: (1) impaired social development; (2) delayed linguistic development; (3) insistence on sameness. All three prior features can be noticeable as early as 30 months old.

Autistic children are often inaccessible not only because their delayed development in languages, but also because of the absence of interest in any social interaction. They often preoccupy themselves with repetitive movements, such as laying down little objects endlessly. They usually do not have pretense play [36], which is common with normal children. Autistic children can be violent and have rages. All these traits make it very hard for their parents and people around. Some people with autism have lifelong institutionalization.

However there is another side of the autistic condition, savant talent, which is often ignored or overlooked by the parents and researchers$^1$. Many people with Autism have amazing abilities [81] [82]. Many are particularly good with jigsaw puzzles. Some have perfect pitch and can play tunes after hearing them just once. Some have perfect memory of all events as early as several months after they were born. These abilities give important cues of the Autism condition as well as the deficits.

Autism is certainly an intriguing condition. Every individual with autism has an unique set of behaviors. However, the main features which we mentioned

$^1$A search for “Autism” on Google Scholar gives 426,000 results; while for “Autism Savant” it gives only 3,560 results.
above can be reliably identified. In the 80s, it was considered a rare condition which happens 4 in every 10,000 children. However, the frequency has increased dramatically over the decades. According to Centers for Disease Control and Prevention, it is affecting 1 in every 88 children in the U.S. Understanding the matter seems urgently important.

3.0.4 Theory of Mind

Theory of mind, termed by Premack and Woodruff (1978), seems to be a particularly human ability. Theory describes how human generally can think about what other people know, believe, and feel. Some researchers, such as Baron-Cohen, Leslie and Uta Frith, have linked the deficit in Autism with lack of theory of mind.

Leslie [36] argues that pretense is part of the origins of ‘theory of mind’ and develops a meta-representational mechanism to illustrate it. In this section, we will first go through the basics of his theory and then explain its connection with Autism.

Children of age of two or older seem to have a more ‘sophisticated’ play: pretend play. Such ability is a major development because it reveals not only children’s ability of handling distorted reality but also the beginning of a capacity for meta-representation which is seen, by Leslie, as the crucial ingredient for a theory of mind. Firstly Leslie [36] excludes some possibilities which appear to be pretend play, such as acting in error and functional play. Then he further defines three fundamental forms of pretend: objection substitution, attribution of pretend properties, and imaginary objects. In other words, any one (or more) of the following situations would make a play pretend play: a) object substitution: if an object is made to stand for another; b) attribution of pretend properties: a certain unreal property is used for a real object; c) imaginary objects: there are imaginary objects. Leslie continues with definitions of primary representation, which is the first basic representational capacity for infants, and representational abuse which affects all three kinds of pretend and handles some relationships between two primary representation. In order to understand the pretend in others, a child would need a capacity for meta-representation. A major feature of the pretend theory is that it actually represents the beginning of a capacity to understand cognition itself, which is a fundamental idea in the theory of mind. Leslie also explains the isomorphism between the three types of pretend and the logical properties of sentences containing mental state terms. In the article, he gives a model for pretend as well: Decoupling Model. There are three major components in the model: 1) the perceptual processes 2) central cognitive systems 3) the decoupler. The decoupler can be further divided into three parts: the expression raiser, the manipulator and the interpreter. The decoupler is involved when a meta-representation is needed. Besides the detailed explanation of how the model works in a pretend play, Leslie also explains the gap between the two-year-old pretenders and four-year-old children who can pass the false-belief test which indicate the mastery of theory of mind.

Children with Autism are impaired in pretend play (Baron-Cohen 1987;
Rutter 1978; Sigman and Ungerer 1981). With the model from Leslie’s article, it is possible to connect this symptom to the social impairment. Leslie concludes that Autistic children lack the ability in both primitive forms (pretense) and advanced forms (false-belief test) because of the impairment in their decoupling ability.

### 3.0.5 The Affective Foundation Theory

For another group of researchers, the symptoms in Autism are not just cognitive, but also emotional. For example, it is very difficult for autistic children to develop close emotional relationships with people, even with their parents. Hobson [31] believes that understanding such impairment in emotional development is also crucial for understanding the Autism condition as a whole. In fact, he argues further that the cognitive and language abnormalities in Autism are consequences of the emotional impairment. In his opinion, ‘infants are biologically prewired to relate to people in ways that are special to people, and it is through the experience of reciprocal, affectively patterned interpersonal contact that a young child comes to apprehend and eventually conceptualize the nature of persons with mental life’ [31](p.104) The affect from other people directed toward objects or events is a special source of information. Through observing this type of information, young children come to understand self. Hobson believes that his process is a required foundation even for symbolic development. He applies his model to Autistic children, and explains that the difficulty in engaging affective states with others is the origin of the problem.

Then he also ran some experiments [30] which tested their ability to recognize emotion and personal identity. The control group in these experiments are non-Autistic retarded children. In the first experiment, the subjects have to complete some tasks in which two types of faces are presented. The first type is used to observe subjects’ ability of identifying faces, while the second type is to observe their ability of recognizing emotion. Regarding the first type, there were no statistical difference between two groups of subjects. However, the the non-Autistic retarded children are better when the emotion is involved. A second experiment was testing their ability of identifying upside-down faces. The results revealed that Autistic children were better at matching both ‘identity’ and ‘emotion’ in upside-down faces.

The experiments seem to fit with Hobson’s theory about Autistic children’s impairment is in affective connections.

### 3.0.6 Mirror Neurons

Starting from late 90’s, another line of research from neuroscience has been offering an alternative explanation on Autism, especially on social impairment. One of the most representative work is from Gallese and his colleagues [24] [23] [87]. They claim that the newly discovered mirror neurons (MNs) have profound

---

2Here we will just focus on one, Hobson.
implication in understanding theory of mind, and hence in understanding of social impairment in autism.

MNs were first discovered in the macaque monkey premotor cortex. During the experiments, MNs respond in following two situations [24]: first, when a particular action is performed by the recorded monkey; second, when the same action performed by another monkey is observed. In [24], Gallese and Goldman further explain the mirror system in humans. The MNs cannot be directly studied in humans but experiments with alternative approaches have shown strong evidence for a system similar to what is discovered in monkeys. The paper [24] proposes that MNs are an important base for mind-reading process or at least a precursors. This claim supports a second type of model of theory of mind: a simulation based model, comparing with the theory based model proposed by Baron-Cohen and his colleagues which we had looked at in an earlier section. Simulation based model suggests that when people ‘read’ others’ minds, they do not need to have a theory of all the psychological inferences, instead they just put themselves in others’ shoes to simulate the situation. Gallese and Goldman suggest that MNs largely explain this process of mapping behavior and then understanding the others’ minds.

Williams et al continues the direction proposed by Gallese, and draw the connection among imitation, MNs and autism [87]. They argue that imitation is ‘a prime candidate for the building of a ToM (theory of mind)’, and that autistic people have deficit in imitation. The response of MNs system in human shows the ability of imitation. They conclude that such a deficit creates problems in social interaction and contribute to the lack of empathy in autistic people.

3.0.7 Discussion

Autistic people suffer a wide spectrum of problems. The deficits in social interaction and language learning are the most prominent. Both are considered to be related to the lack of ToM. As we have discussed, there are two major competing models for explaining this: Theory Theory (ToM) and Simulation Theory. A second line of problems are emotion-related, such as lack of empathy. The affective foundation theory is the first to look at these problems. Although none of the existing theory has really unveiled the root of stereo-typed behavior, such as sameness and repetition.

The newer theory, namely simulation based theory, has shown promising development by applying properties of MNs in connection with autism. It seems to explain the lack of theory of mind which affects social and language learning, and lack of empathy.

However, if we are in agreement of such deficit in MNs system in autistic people and the fact that it is the prime base for ToM and social interaction, do we get the conclusion that autistic people can never learn social interaction or theory of mind?

I believe there are deeper philosophical questions that need to be answered. What are the natural emergency of theory of mind and social meaning? Is there an un-nature process which purely rely on high-level intelligence as many autistic
people have normal to higher-than-average intelligence? These are questions beyond the scope of current study and subjects to the future research. However, I want to conclude the paper with a quote from the chapter on the autistic professor, Temple Gardin, in “An anthropologist on Mars” by Oliver Sacks:

... *What is it, then, I pressed her further, that goes on between normal people, from which she feels herself excluded? It has to do, she has inferred, with an implicit knowledge of social conventions and codes, of cultural presuppositions of every sort. This implicit knowledge, which every normal person accumulates and generates throughout life on the basis of experience and encounters with others, Temple seems to be largely devoid of. Lacking it, she has instead to ‘compute’ others’ intentions and states of mind, to try to make algorithmic, explicit, what for the rest of us is second nature. ...*

The subject, Temple Grandin, shows clear self-introspection, ideas of ToM and understands that her way of perceiving social interaction is different. She actually wrote a book on social rules [25] “The Unwritten Rules of Social Relationships: Decoding Social Mysteries Through the Unique Perspectives of Autism” with another autistic adult. Who would imagine that when she was diagnosed with autism as a child, the doctor had suggested life-long institutionalization for Temple?

So maybe both Theory Theory and Simulation Theory are right. However, they are two different approaches to understanding of ToM and social interaction. I would call simulation-based a more natural kind. More philosophical research needs to be done in this direction.
Chapter 4
Signaling

4.0.8 Sender-Receiver Model

In this section, we will briefly review some important developments in sender-receiver configuration. I will group different theses into two types: the theory type and model type, and disregard the chronological order. We will start with some modelers, and continue with a few theoretical developments. At the end of this section, I will summarize the key features that are important to this paper from both groups.

Although Shannon developed Information Theory, I still consider him more of a modeler in the context of sender-receiver model and in comparison with other theorists whom I will discuss later. Shannon defined the early version of sender-receiver process to illustrate the information delivery from the world to a sender who signals through channels to a receiver. Lewis further developed the model in a more game theoretic setting [37] to show that meaning is a result of sender-receiver interaction given common interest and knowledge. Skyrms pushes the model a couple of steps further in the direction of naturalistic approach. He shows that without the assumption of common knowledge and high level of intelligence, the meaning of a signal can still evolve. In fact, his model shows a continuous application of sender-receiver configuration at all biological levels.

Both Dretske [16] and Millikan develop detailed theoretical accounts for how information is being processed. Dretske gives enriched explanation of information in a natural and objective manner, in a sense that a signal with information is not tied to a time, a person or a history. Millikan opposes this type of signals. In her book ‘Varieties Of Meaning’ [42], she further defines natural signs in a local and recurrent sense. As Millikan’s theory on varieties of meaning (signs) is quite important for my discussion, we shall come back to the details in the later sections.

So which aspects of the models and theories are important for this proposal? I will call it a quasi-Lewis-Skyrms configuration. It certainly is a sender-receiver configuration. I want to use the game theoretic aspect, in which the interaction
between the sender and receiver, and re-enforcing reward, naturally stabilize a particular meaning of the signals. However, at the same time, I will incorporate Millikan’s various refined definition of signals into the game theoretical framework. Eventually I hope such a quasi-Lewis-Skyrms configuration can be a tool to distinguish the two kinds of perceptive meaning(signals) between people. In addition, I hope we can illustrate how misinformation can happen in both case, and why only in one case it can be corrected.

### 4.0.9 The Lewis-Skyrms Signaling Game

I will start with Lewis’ signaling game. In a typical setting, there are two agents who want to achieve a common goal, for example the sexton of the Old North Church and Paul Revere. A crucial difficulty they are facing in the Revere Story is that neither can achieve the goal alone, in this case to inform the American defenders against the British army. The sexton can observe the arrival of the army, but cannot defend; while Revere can help with defending, but cannot observe the arrival. Hence, they established a signaling game for coordination.

In the game, there are three states (invasion by land, invasion by sea and no invasion), three signals accordingly (one lantern, two lantern and no lantern), and each could result in an action (preparation for invasion by land, preparation for invasion by sea and no preparation). Table 1 represents a normal-form game matrix for the Revere story.

<table>
<thead>
<tr>
<th>States</th>
<th>Acts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A1</td>
</tr>
<tr>
<td>S1</td>
<td>2</td>
</tr>
<tr>
<td>S2</td>
<td>0</td>
</tr>
<tr>
<td>S3</td>
<td>0</td>
</tr>
</tbody>
</table>

More formally speaking, there are two functions \( f_c \) and \( f_a \), which are used by the communicator(sender) of the signal and the audience(receiver) respectively. The sender can observe a set of states, and has a set of signals available; while the receiver can not observe the states, but has a set of actions he can use to act on each signal. \( f_c \) is a mapping from states to signals. \( f_a \) is a mapping from each signal to at most one action. Coming back to the Revere story, the sexton as the sender can observe three states \( S_1 \) (by land), \( S_2 \) (by sea), and \( S_3 \) (no invasion). He also has three possible signals to use \( M_1 \) (one lantern), \( M_2 \) (two lantern) and \( M_3 \) (no lantern). Revere as the receiver can have three different acts \( A_1 \) (preparation for land), \( A_2 \) (for sea), and \( A_3 \) (none). If \( f_c \) is a one-to-one function, it is called admissible. Similarly, \( f_a \) can be admissible. A pair of admissible functions \( < f_c, f_a > \) is a signaling system. In the game matrix, such

\[1\] I will use ‘sender’ and ‘receiver’ from now on as they are more conventional.
pairs of functions give a set of equilibriums\(^2\) \(\langle S_1, A_1 \rangle, \langle S_2, A_2 \rangle, \langle S_3, A_3 \rangle \). Hence there could be multiple signal systems. In the Revere story, we could have \(M_2\) for \(S_1\), \(M_1\) for \(S_2\), and \(M_3\) for \(S_3\). Receiver’s act would change accordingly too. \(A_1\) for \(M_2\), \(A_2\) for \(M_1\), and \(A_3\) for \(M_3\). However, the equilibrium does not change.

Now let us look at a few important assumptions in Lewis signaling games and then connect them to Skyrms’ book. The first two assumptions are common interest and common knowledge. Common interest is reflected in the payoff structure, i.e. both get maximum payoff when a signaling system is established. The fact that both parties are assumed to have a certain level of reasoning ability, which enables them to make an agreement and have higher order expectation of each other, reflects the common knowledge assumption. The third assumption is made on how a particular signal system is chosen. As we have shown above, multiple signal systems can exist in one game. Lewis assumes the players either have a prior agreement or a natural salience when choosing a particular signaling system.

Skyrms book ‘Signals’ responds to all these three assumptions and pushes the game in a more naturalistic direction. First, common interests are not necessary, as he shows how senders and receivers may have partial or opposite interests. He also argues that signaling games can be useful in the analysis of a wider biological spectrum without a high level reasoning capacity, thus eliminating the need for common knowledge. By using evolutionary dynamics, he shows that without any natural salience\(^3\) a signal system can emerge.

### 4.0.10 Information in ‘Signals’

Under the naturalistic assumptions, Skyrms emphasizes the importance of the information flow from the beginning of the book (page 2). He argues that the meaning of signals can be studied through the information carried by signals. Such information is further measured by its *quantity* and *content*. Skyrms derives the quantity of information from the mathematical theory of information which was originated by Claude Shannon.

As we have briefly pointed out above, Skyrms uses evolutionary dynamics to show the emergence of signaling systems. Initially, nature creates necessary states with equal probabilities. The quantity of information is how much a signal moves probabilities. If we have two states \(S_1\) and \(S_2\) with initial probabilities \(p_{S_1} = 0.5, p_{S_2} = 0.5\), and a signal \(A\) moves the probabilities of the two states to \((p_{S_1} = 0.2, p_{S_2} = 0.8)\), then the information quantity is same as another signal \(B\) which moves the probabilities to \((p_{S_2} = 0.2, p_{S_1} = 0.8)\). Skyrms defines information content as a vector which indicates not only how far a signal moves the states’ probabilities, but also in which direction. In our example, information

---

\(^2\)The pair of state and act which gives both sender and receiver best payoff when the interests of the two are aligned.

\(^3\)Clearly, it also assumed to have no prior agreement since high level reasoning capacity is not assumed.
content for signal $A$ is different from that of signal $B$, since they affect the probabilities in opposite directions.
Chapter 5

Social Software: From Social Procedure to Homo Sociologicus

Although we are approaching the conclusion, we are far from concluding if we do not review the seminal work on the topic from Parikh. We have purposefully excluded his work thus far hoping to explain the seemingly-unrelated past chapters through introducing a series of his past important and highly interdisciplinary works. Therefore, in this chapter we will review a few key works from Parikh related to the search of homo sociologicus. Then we will continue with a section that explains the connection between his work and the direction of this review paper.

Trained as a physicist and mathematician and did much of the research works as a logician, Parikh has had interest in social sphere for decades, such as [56] [47]. In 1995, he officially coined the term, Social Software [48] [49], and called for an interdisciplinary collaboration between computer scientists and social scientists. In [49], he points out that computer science concepts, such as algorithms and data, were already described in early philosophical works that concerns social context such as Wittgenstein’s and Grice’s. At the same time, he also gives a few real life social problems, for example car parking problems, which could have been improved by using an algorithm like social procedure. He gives a famous game theory example, the Santa Fe bar problem, which was discussed first by Brain Arthur [8] in 1994, later also by Greenwald, Mishra and Parikh [26] in 1998. As Parikh explains in [49], when the demand of a public good exceeds the supply in a group of people, the conventional free market approach is not always the most desirable one socially. He states the “striking contrast between the efficiency with which purely computational processes are carried out, and the inefficiency of the social processes which are intended to be... ”. Therefore he argues that the role of algorithms and game theory in social procedures should be properly studied and positioned. [49] is an important first
step that opens the research area of social software.

Like the beginning of any theory, key concepts have to be properly raised, discussed, and defined. In the work of social software, there are a few such concepts that were elaborately explained by Parikh in [50] [51] [52]. Being an expert on the topic of knowledge, Parikh focuses on the technical aspects of knowledge in [50] based on his past work in distributed computing together with Krasucki [57] and Ramanujam [58]. He suggests that understanding the levels of knowledge under different circumstances is crucial. Later in [51], Parikh brings attention to the knowledge and logic structure of algorithms in social context. The paper displays the necessity of logical structure in the car parking problems and logical conditions in using a key.

Starting from “Logic of Society”[54], Parikh starts to list more directions that could be taking in the social software research besides the previous focuses, knowledge, logic and planning. It sheds some light on ‘softer’ issues such as rationality, incentives and preferences, culture and tradition. To be more specific, the social software project is extending to the human elements. [60] continues this extension to game theory. In the paper, Parikh, Tasdemir and Witzel illustrate how different types of people use the same utility function differently. The knowledge of your opponent’s type in a game is a ‘game-changer’. Two relatively recent papers of Parikh, [55] and [59], are extending quest of human elements in a more philosophical direction.

The works mentioned above are just highlights of Parikh’s research. Starting from “Logic of Society”[54], Parikh starts to list more directions that could be taking in the social software research besides the previous focuses, knowledge, logic and planning. It sheds some light on ‘softer’ issues such as rationality, incentives and preferences, culture and tradition. To be more specific, the social software project is extending to the human elements. [60] continues this extension to game theory. In the paper, Parikh, Tasdemir and Witzel illustrate how different types of people use the same utility function differently. The knowledge of your opponent’s type in a game is a ‘game-changer’. Two relatively recent papers of Parikh, [55] and [59], are extending quest of human elements in a more philosophical direction.

The works mentioned above are just highlights of Parikh’s research. They appear to follow a direction of from system to components, from the hard issues to soft issues, and from social procedure to homo sociologicus.

5.1 Conclusion and Discussion

5.1.1 The Jigsaw Puzzle: in Search of Homo Sociologicus

We first raised the need for better theory for predicting people making social related choices through reviewing two widely experimented games: Public Goods Game and Ultimatum Game. Then we reviewed a theoretical pillar: the rational choice theory. Parikh points out in [54], “actual behavior differs from theoretical prediction and seems to follow some pre-existing cultural pattern”. Getting back to the queuing example in [49], besides FIFO, there are other different possibilities. For example, people who come late can still go to the front of the queue. What stops him from doing that? The pure ethical concern? Or fear of others’ scolding? He suggests that the work from Lewis on convention [37] has great relevance to the subject. We, therefore, reviewed the famous signaling game from [37]. Skyrms provides a more evolutionary and biological view of the game, which could help us understand how signals and meaning form over time. Potentially, we could blend his ideas into defining social signals, which is the next step that we want to take in our research.

\footnote{Much of Parikh’s works deserves further detailed reviewing, which we will continue in the research proposal stage.}
This review paper is in the direction of converging with the social software project led by Parikh. The focus is how to define the sociality within each individual. More importantly how is such sociality related to our rationality and choices.

Rational choice theory serves a theoretical pillar to all concerning choices including making choices in a group, which is the focus here. However in such situation, sociality is reflected on two levels, namely the group level and individual level. Most of the work in the plethora of social software research concerns the first. We want to argue that the second level is equally important. What is sociality on an individual level? Although this is a crucial point that requires more detailed explanation in the future research, we can simply see it as a kind of social influence when individuals are making decisions. For example, in both Public Goods Game and Ultimatum Game experiments, people deviated from classical game theoretical predictions. We shall not easily conclude that they were irrational because their behaviors do not fit within conventional rational choice framework. As we have seen from the experimental results in [28] from the 15 small scale societies, such deviations are highly correlated with their social structure. Clearly, there is some kind of rationale in these people’s choice. Of course, these subjects didn’t study different social systems and then decide one choice for a situation like Public Goods Game or Ultimatum Game. So how did they decide? What is the social influence here?

5.1.2 Possible Future Direction: Defining and Modeling Social Signals?

Autism seems at first glance to be an unrelated topic. However the research of the autism spectrum gives us various of examples that are philosophically important for modeling social signals [88]. Autistic people have distinctive social reasoning patterns that are different from normal people [59]. Yet the patterns seem to be closer to animal’s [67]. How does this help us in modeling social signal or evolution of social signals? Is autism an intermediate step? How does modeling social signal help us understand social choice? ...

There are many questions we want to answer, but in the next stage of our research.
Chapter 6

Knowledge and Epistemic Logic

6.1 Knowledge

We can trace original epistemology, the study of knowledge, back to the Ancient Greek philosophy. The *Theaetetus*, which is considered Plato’s greatest work on epistemology [11], focuses on the question “What is knowledge?” In general, the study concerns following questions: “What are the necessary and sufficient conditions of knowledge? What are its sources? What is its structure, and what are its limits?” [80].

Here in this proposal, we are more focused on the formal aspects of the knowledge. More specifically, we will review some existing frameworks in epistemic logic and recent development in dynamic epistemic logic such as friends’ influence. At the end of this chapter, we will also discuss some technical results on language splitting which is useful for our new logic system.

6.2 Formal Frameworks

Von Wright’s work (1951), *An essay in modal logic* [85], is widely considered to be the first proper formal treatment of epistemic logic. In 1962, Jaakko Hintikka extended the work in his book *Knowledge and Belief: An Introduction to the Logic of the Two Notions* [29]. Saul Kripke made a technical breakthrough in 1963 [35] with his Kripke semantics. In the Kripke semantics, $< W, R, |= >$ is called a Kripke-model. $W$ represents a possible set of worlds that are accessible for the agents. $R$ represents the relation between the worlds, which can be symmetric, reflexive, transitive, etc. $|= $ reads as “models”. We can use this to express the Alice example again. $m, w |= K_a P$ says that there is a model $m$ in a world $w$ such that the statement “Alice (a) knows statement $P$” is true. The
following axiom $K_a(A \to B) \to (K_a A \to K_a B)$ is called $K$ axiom. This axiom constitutes System $K$, the most basic reasoning system in epistemic logic. More reasoning axioms can be added to the system. Hence we also have System $D$, System $T$, System $S4$, etc.

In the work of the Logic of Belief Revision, AGM [2] model is the most dominant theory. A belief state includes a set of sentences (statements), which is logically closed and is called “theories”. Let us assume $K$ represents a belief state of an agent, i.e. $K$ is a logically closed set of sentences. There are three types of belief changes that can happen to a belief state $K$.

- **Expansion:** A sentence $p$ is added to $K$ and nothing in $K$ is removed. The new belief set is $K + p$.

- **Contraction:** A sentence $p$ is removed from $K$. The new belief set is $K \div p$ with some adjustment needed.

- **Revision:** A sentence $p$ is added to $K$, and at the same time some sentences are removed if they contradict with $p$. The new belief set is $K * p$.

The last type, belief revision, can have an absurd consequence. When a new sentence that is inconsistent with $K$ is added, all the information in $K$ can be discarded. This is clearly unrealistic. In real life, when we learn something about the weather we retain our beliefs about politics. In the last section of this chapter, we will see how this can be improved by using splitting languages.

### 6.3 Friends’ Influence

The most recent development of dynamic epistemic logic is to apply belief revision in social networks. We always have a group of people that are close to us and who can also influence our beliefs in many ways. More importantly understanding the belief influence in a group of people will help us understand how social belief changes happen at a local level. Liu et al [38] do suggest a simple yet normative model for social belief change, called ‘threshold influence’.

In their framework, they look at friendships which is taken to be a symmetric and irreflexive relation. So an agent is a friend of any friend of hers, but she is not her own friend. In addition, they do not assume that friendship is transitive. Therefore friends of friends are not necessarily friends. Friendships among a group of people create a social network, or community. Agents start with some initial belief towards an issue $p$. They can be $Bp$ (believe $p$), $Bp$ (believe the negation of $p$), or $Up$ (no opinion on $p$). In the process of communicating with friends, this belief can be changed in two ways. When the agent is strongly influenced by friends ($Sp$), it leads to belief revision ($Rp$). Similarly, when she is weakly influenced ($Wp$), it leads only to belief contraction ($Cp$).

1. If an agent $a$ knows the implication between the two statements $A$ and $B$, then if she knows $A$, she may also know $B$.
2. This notation differs in different literature.
Symbol $F$ means ‘all my friends’. Therefore $FBp$ means ‘all my friends believe $p$’. In the framework, Liu et al further define the dual operator $\langle F \rangle$ as ‘some of my friends’. $(F)Bp$ means some of my friends believe $p$. Then they add the aspect of a threshold. For instance assuming my threshold is 100%, if I believe $p$ but all of my friends believe $\neg p$, then I will change my own belief to $\neg p$. but if only some of my friends believe $\neg p$ then I will become undecided about $p$.

Within a group of friends, beliefs tend to be adopted from one to the other. It creates a distribution of beliefs in the community. Liu et al [38] look at the influence one could gain from a socially connected agent.

The three possible doxastic states of a proposition $p$ can be defined through following axioms:

- **Strong influence**: $Sp \iff (FB\varphi \land \langle F \rangle B\varphi)$
- **Weak influence**: $Wp \iff (F\neg B\neg \varphi \land \langle F \rangle B\varphi)$

Liu et al also look at the stability of beliefs in different friendship-based networks. They defined a program, $Ip$ with following rules:

- if $Sp$ then $Rp$ else if $Wp$ then $C\neg p$
- if $S\neg p$ then $R\neg p$ else if $W\neg p$ then $Cp$

They call a community **stable** when the program $Ip$ has no effect on the belief states of the members. Some communities become stable after a few applications of $Ip$. Some can never be stable. Then these communities are said to be in flux. Figure 3.1 is one such example.

6.4 **Language Splitting**

Parikh [53] shows that a person’s beliefs can be uniquely divided into different subject matters and so if his beliefs are a theory $T$ in $L$ then $L$ will split naturally into sublanguages $L_1 \cup L_2 \cup L_n$ and $T$ will split into $T_1 \cup T_2 \cup ...T_n$, each $T_i$ in $L_i$.

This idea of splitting language was naturally applied to belief revision. If I learn some political fact, it will not change my views about my teeth or the
location of my children. Moreover, if have trouble with my car, I will not consult a dentist. Conversely I will not consult a garage mechanic about my teeth.

Before we look into Parikh’s axioms for language splitting, let us review the original AGM axioms which can lead to some forgetful updates.

**AGM Axioms**

1. \( T \ast A \) is a theory
2. \( A \in T \ast A \)
3. If \( A \leftrightarrow B \), then \( T \ast A = T \ast B \).
4. \( T \ast A \subseteq T + A \)
5. If \( A \) is consistent with \( T \), i.e. it is not the case that \( \neg A \in T \), then \( T \ast A = T + A \)
6. \( T \ast A \) is consistent if \( A \) is.
7. \( T \ast (A \land B) \subseteq (T \ast A) + B \)
8. If \( \neg B \notin T \ast A \) then \( (T \ast A) + B \subseteq T \ast (A \land B) \)

If \( A \) is consistent with \( T \), then \( T \ast A = T + A \), otherwise \( T \ast A = \text{Con}(A) \).

Given these axioms, a forgetful update can happen in which case when \( A \) is inconsistent with \( T \), all information in \( T \) is discarded.

Clearly this kind of updates are unrealistic since in real life when we learn something about weather we retain our beliefs about politics. Parikh improves this update by using Craig’s Interpolation Theorem.

**Craig’s Interpolation Theorem**

**Theorem 9.** Let \( L_1, L_2 \) be first order languages, \( L = L_1 \cap L_2 \) and \( T_1, T_2 \) be theories in \( L_1, L_2 \), respectively such that \( T_1 \cup T_2 \) has no model (is inconsistent). Then there is some formula \( \psi \) of \( L \) such that \( T_1 \vdash \psi \) and \( T_2 \vdash \neg \psi \). In particular, if \( \phi \) is an \( L_1 \) formula and \( \xi \) is an \( L_2 \) formula and \( T_1 \cup T_2, \phi \vdash \xi \) then there exists an \( L \)-formula \( \psi \) such that \( T_1, \phi \vdash \psi \) and \( T_2, \psi \vdash \xi \).

### 6.4.1 P Axioms

Parikh proposed following axioms to improve the belief updates, which prevent the discarding all the information in the way we mentioned in a forgetful update.

1. **Axiom P1:** If \( T \) is split between \( L_1 \) and \( L_2 \), and \( A \) is an \( L_1 \) formula, then \( T \ast A \) is also split between \( L_1 \) and \( L_2 \).

2. **Axiom P2:** If \( T \) is split between \( L_1 \) and \( L_2 \), \( A, B \) are in \( L_1 \) and \( L_2 \) respectively, then \( T \ast A \ast B = T \ast B \ast A \).
3. **Axiom P2g:** If $T$ split between $L_1$ and $L_2$, $A$, $B$ are in $L_1$ and $L_2$ respectively, then $T \ast A \ast B = T \ast B \ast A = T \ast (A \land B)$

4. **Axiom P3:** If $T$ is confined to $L_1$ and $A$ is in $L_1$ then $T \ast A$ is just the consequences in $L$ of $T \ast' A$ where $\ast'$ is the update of $T$ by $A$ in the sub-language $L_1$.

All these axioms follow from axiom $P$ below:

**Axiom P:** If $T = Con(A, B)$ where $A, B$ are in $L_1, L_2$ respectively and $C$ is in $L_1$, then $T \ast C = Con(A) \ast' C \ast B$, where $\ast'$ is the update operator for the sub-language $L_1$. 
Chapter 7

Network Science

7.1 Networks

The Network Science is truly interdisciplinary. Euler’s solution of the Königsberg bridge problem in 1735 is considered the first true proof in the theory of networks [44]. The vast body of knowledge has been applied physics, mathematics, computer science, biology, economics and sociology. With the access to cheaper computing power at the turn of this century, the science has been focusing on different problems [86], many of which are supported with empirical observations [9].

The 2003 Northeast blackout is considered a famous example of cascading failure in a network [9]. In that event, 55 million people lost their power. A more recent example (see figure 4.1) would be Google. The search giant’s service was inaccessible for millions due to a cascading failure caused by a routing leak which originated from an Indian ISP. Both cases expose the vulnerability due to interconnectivity in a network.

For our purpose of understanding social issues, this following example which is explained by Barabási in the very beginning his book Network Science, seems more relevant.

In March 2003, American forces entered Iraq. But they were unsuccessful at capturing high ranking officials including Saddam Hussein. Later, the US military reconstructed the social network of Hussein, relied on gossip and family trees instead of government documents. Using the social network diagram, there were a few successful raids including one that led to an important piece of intel: a family album. It dramatically helped with further understanding of Hussein’s trusted network. The military was eventually able to figure out the hiding place of Saddam Hussein.

Through the example, Barabási points out a few important observations of network theory:

- The predicative power: it allows even non-experts to extract useful information.

41
Figure 7.1: Google went down for millions of users due to a routing leak from an Indian ISP

Figure 7.2: Image from *Network Science* by Albert-László Barabási

Image 1.2b

The network of Saddam Hussein.
• The remarkable stability: the social network was not constructed through updated intelligence but dated information such as gossips and family photo albums.

• The choice of network can be crucial. In the case of Hussein, the military tried for months to find him through the network of the Iraqi government. But it was personal network that eventually helped.

7.2 Basic Concepts

Often ‘networks’ and ‘graphs’ are often used interchangeably by researchers. However in Network Science, we use network, node, and link, while in Graph Theory, we use graph, vertex and edge. In this chapter, we will use the terminologies from network science.

Let us review some basic concepts in network science. \( N \), Number of Nodes is the total number of components in a network. \( L \), Number of Links, is the total number of interactions between the nodes. A link can be either directed or undirected. When we trace how a message is being sent in a network, the links are directed from senders to receivers. When we look at hand-shaking in a group of people, the links are undirected since a handshake always takes two people at the same time.

Degree, \( k \), is an important concept in network science. It is the number of links from a node to other nodes. In undirected networks, we have following relationship between \( L \) and \( k \):

\[
L = \frac{1}{2} \sum_{i=1}^{N} k_i
\]

The average degree, \( \langle k \rangle \), of a network can be calculated in the following method:

\[
\langle k \rangle = \frac{1}{N} \sum_{i=1}^{N} k_i = \frac{2L}{N}
\]

In directed networks, we have incoming degree, \( k_i^{in} \) and outgoing degree, \( k_i^{out} \). Therefore the total degree of a node \( i \) is:

\[
k_i = k_i^{in} + k_i^{out}
\]

Following formula shows the relationship between \( L \) and \( k_i^{in} \) and \( k_i^{out} \):

\[
L = \sum_{i=1}^{N} k_i^{in} = \sum_{i=1}^{N} k_i^{out}
\]

In directed network, the average degrees of \( k_i^{in} \) and \( k_i^{out} \) are:
\[ (k_i) = \frac{1}{N} \sum_{i=1}^{N} k_i = (k_{out}) = \frac{1}{N} \sum_{i=1}^{N} k_{out} = L/N \]

\[ p_k, \text{ the degree distribution, is the probability that a randomly selected node in the network has degree } k. \text{ Generally, we have } \sum_{k=1}^{\infty} p_k = 1. \text{ In a fixed network that has } N \text{ nodes, we have: } p_k = \frac{N_k}{N^2}, \text{ where } N_k \text{ represents the number of nodes that has degree } k. \text{ We can also express the average degree of a network in terms of } p_k: \]

\[ \langle k \rangle = \sum_{k=0}^{\infty} kp_k \]

In a network that is a complete graph, which means every node is connected to all others, then we have \( L_{max} = \left( \frac{N}{2} \right) = \frac{N(N-1)}{2} \). However, most of the networks observed in real life are sparse [9].

Often it is easier to talk networks in the adjacency matrix. Therefore, we will review some concepts and formula related to it. In a directed network with \( N \) nodes, the adjacency matrix has \( N \) rows and \( N \) columns. \( A_{ij} = 1 \) when there is a link pointing from node \( j \) to node \( i \). \( A_{ij} = 0 \) when there is no link between node \( j \) and node \( i \). In an undirected network, \( A_{ij} = A_{ji} \). Since \( A_{ij} \) represents the links, we can get the degree \( k_i \) of node \( i \) from the adjacency matrix. For undirected networks:

\[ k_i = \sum_{j=1}^{N} A_{ij} = \sum_{i=1}^{N} A_{ij} \]

For directed networks:

\[ k_{in}^i = \sum_{j=1}^{N} A_{ij} \]

\[ k_{out}^i = \sum_{i=1}^{N} A_{ij} \]

A path is a route from one node to another in a network. It can pass through the same link more than once. It can also intersect itself. A shortest path, \( d_{ij} \), is a path between \( i \) and \( j \) with fewest number of links. In undirected networks, \( d_{ij} = d_{ji} \), while in directed networks, \( d_{ij} \neq d_{ji} \). A network diameter, \( d_{max} \), is the maximal shortest path in a network. Average path length, \( \langle d \rangle \) is \( \frac{1}{N(N-1)} \sum_{i,j=1,N} d_{ij} \) in a directed network; is \( \frac{2}{N(N-1)} \sum_{i,j=1,N} d_{ij} \) in an undirected network.

When there is a path between two nodes \( (i \text{ and } j) \), we say that they are connected, and otherwise they are disconnected with \( d_{ij} = \infty \). A network is connected if all the pairs of nodes are connected.

In figure 7.3 (a), it is a disconnected network with two components. If we place a single link between 2 and 4, the network becomes (b) which is connected. This link is called a bridge.
In real life networks, e.g. social networks, highly density ties are often observed. Let us review a concept that is related to it. The local clustering coefficient represent the degree of how neighbors of a node are connected to each other. For a node $i$ with degree $k_i$, we can express the local clustering coefficient as following:

$$C_i = \frac{2L_i}{k_i(k_i - 1)}$$

where $L_i$ is the total number of links between the neighbors of node $i$.

$C_i$ measures the local density in a network. It is between 0 and 1. When $C_i = 1$, the neighbors of $i$ are fully connected and form a complete graph. When $C_i = 0$, none of the neighbors of $i$ are connected.

### 7.3 The Scale-Free Property

When the WWW was first mapped out in 1999 [1], it was discovered that the network had many highly connected nodes unlike in a random network. In fact, many real networks have the same property [9]. It is called the scale-free property. We can represent the degree distribution of the WWW in following manner:

$$p_k \sim k^{-\gamma}$$

We call this a power law distribution with a degree exponent $\gamma$. After taking a logarithm of the formula above, we get: $\log p_k \sim -\gamma \log k$.

In directed networks, we will have following:

$$p_{ki} \sim k^{-\gamma_{in}}$$

$$p_{kout} \sim k^{-\gamma_{out}}$$
When we compare a scale-free network with a random network whose distribution is a Poisson distribution, we observe that the probability of high-degree nodes is much higher in a scale-free network. (See figure 7.5 and 7.6) In addition, there are many small degree nodes in scale-free networks while they are absent in random network. Figure 7.6 is an example of a scale-free network with \( \langle k \rangle = 3 \). “The more nodes a scale-free network has, the larger are its hubs.” [9].
Figure 7.5: A Poisson function compared with a power-law function with $\gamma = 2.1$. Both have $\langle k \rangle = 10$. [9]

Figure 7.6: Same two functions as above shown on a log-log plot. [9]

Figure 7.7: A scale-free network with $\langle k \rangle = 3$. [9]
Chapter 8

Proposed Work

8.1 Introduction

Logic is a study which was initially associated with theoretical studies, such as philosophy, mathematics, and later computer science and linguistics. In the recent decades, different branches of logic have blossomed into many modern logic systems, many of which are applicable to areas like cognitive science, neuroscience, and biology.

Epistemic logic is one group of such systems that captures how we reason about knowledge and beliefs. In such systems, we can express the logic of individual knowers. For example, Alice knows a computer science statement: the problem of NP versus P is not solved. We can also show that Alice knows a political statement: The current president is Obama. The systems can be extended to a group of knowers. Then we can express a statement like, “Bob knows that Alice knows the statement.”. It can be written in the following format: $K_bK_aP$, with $b$ as Bob, $a$ as Alice and $P$ as the statement we mentioned earlier. In such multi-agent systems, we can also express a statement that is known by all agents simultaneously, common knowledge. For example, a public announcement generally creates common knowledge.

A more recent development is to have active agenthood in the logic systems. Such systems can show a more dynamic process of how beliefs and knowledge are updated. This branch of logic systems is often called dynamic epistemic logic. Each agent has a set of initial beliefs (or knowledge). For one particular statement, an agent can believe it, disbelieve it or have no opinion. When new evidence presents, an agent usually has three options: revise, contract, and expand. With these settings, agents now can communicate and influence each other. We can see how belief convergence may lead to greater group coordination.

The latest logic systems have placed such active agents in a social network. The belief and knowledge updates depend not only on the initial state of an agent and her influencer, but also on the network structure of their community and
their positions. This development, therefore, shows the changes of knowledge and belief on a community level.

In this dissertation, I will build a logic system based on the latest development in dynamic epistemic logic for a community. I will additionally include possibilities of sub-languages\(^1\). A sub-language used by some members of the community can be seen as a language for a particular topic or a sub-culture. For example, a community of people all speak English. Among them, some of them are physicians. These physicians may share a sub-language that is filled with jargon and medical statements. A sub-language on a topic can also be shared among non-professionals. A good example can be a sub-language shared by foodies.

There are a few benefits of splitting a language into multiple sub-languages. First of all, we can express a more realistic process of day-to-day belief or knowledge updates. Due to the restrictions of the mathematical axioms in dynamic epistemic logic, when an agent revises her belief or knowledge on one particular statement, she might disregard all her other beliefs in the same language\(^2\). With a splitting language scenario, the agent would only revise all her beliefs or knowledge in related topics. For example, if I learn, at some point, that there is a two-term limit for presidency in the US, then I only revise my belief in the topic of politics. I disregard my past belief that Obama will be the president of the US in 2017. However I will not revise any of my beliefs about my dentist.

Second benefit is that we will be able to capture the multiple sub-cultures. Often people who are knowledgeable about politics are not necessarily the ones who are also into food. Therefore, knowledge influences may spread in different directions within a network of agents. A splitting language allows us to demonstrate this.

Last but not least, when we separate knowledge into different topics, we naturally have experts in each topic. These are the people who have more knowledge on a topic than most of the others in the network.

My dissertation contains three sections: 1. Expert Influence, 2. Distributed Common Knowledge, and 3. Simulations of Knowledge Influence. The first section lays the ground work for the new epistemic logic system that allows multiple sub-languages. In this new system, we can show how experts influence other agents’ beliefs on a particular topic. The second section shows some additional results in this new logic system with regard to common knowledge. One of the most important aspects is to show how group coordination can be achieved even with localized common knowledge in the social network. The last section will be a series of simulations of this logic system. Different network structures will be simulated and compared to the actual social networks that are observed in real life.

\(^1\)Past work have assumed one language without any sub-languages for the whole community.
\(^2\)This point will be elaborated in the following two sections: Methodology - relevant literature and Projected results.
8.2 Expert Influence

Work on belief revision can be traced back to Peirce’s classical papers, “the Fixation of Belief” in late 19-th century and “how to make our ideas clear”. A more recent, more technical, and a widely recognized work is by Alchourron, Gärdenfors and Makinson [2]. However, neither Peirce nor AGM discuss the influence on our beliefs by other people. They are interested in belief change but not on its source in other people. How do we adopt beliefs from people and adjust our beliefs accordingly?

Liu et al [38] do suggest a simple yet normative model for social belief change, called ‘threshold influence’. They contrast strong influence which leads to belief revision and weak influence which leads only to belief contraction. For instance if I believe $p$ but all of my friends believe $\neg p$, then I will change my own belief to $\neg p$. but if only some of my friends believe $\neg p$ then I will become undecided about $p$. Within a group of friends, beliefs tend to be adopted from one to the other. It creates a distribution of beliefs in the community. Liu et al [38] look at the influence one could gain from a socially connected agent.

We are extending the setting to show how influence on a subject travels in general. It differs from the friends network in following ways:

- Not everyone we talk to are friends. Some are connected to us professionally such as co-workers. Some are connected to us geographically such as neighbors.
- We are being influenced not just by friends. Often we are influenced by people we consider who know more on a subject matter. In other words, they are experts or more close to experts than we are.
- Even though we talk about many subjects in the community, we do ask advice from certain people on particular subjects. More importantly, these people might not be friends.

Therefore, we propose an alternative model that captures expert influence. Parikh [53] shows that a person’s beliefs can be uniquely divided into different subject matters and so if his beliefs are a theory $T$ in a language $L$ then $L$ will split naturally into sub-languages $L_1 \cup L_2 \cup L_n$ and $T$ will split into $T_1 \cup T_2 \cup ... T_n$, for each $T_i$ in $L_i$. This idea of splitting language was naturally applied to belief revision. If I learn some political fact, it will not change my views about my teeth or the location of my children. Moreover, if have trouble with my car, I will not consult a dentist. Conversely I will not consult a garage mechanic about my teeth. So a community can have different experts on different areas of concerns.

In this dissertation, the new logic system employs all AGM axioms, P axioms from [53], and two new additional axioms that instruct agents what to do when they are influenced by an expert or a non-expert.

- **Influence from Expert to Non-expert**: Before a non-expert talks to an expert on $p$, he can believe $p$, believe $\neg p$ or have no opinion about $p$. 
Once the non-expert talks to a neighboring expert about \( p \), he adopts the beliefs of the expert and become an expert on \( p \) himself.

- **Influence between Non-experts**: Between two non-experts, if they have different beliefs about \( p \), they both contract their beliefs on \( p \) since none of them are authoritative on the matter.

We are aware that this logic system has some limitations. Firstly, we currently assume one expert per subject at the starting stage. In real life, it is not hard to imagine that there are multiple experts. Secondly, the influence is currently dominated by experts. However we observe in real life that non-experts can also teach experts a thing or two. These will be our future research directions.

## 8.3 Distributed Common Knowledge

When all agents know a statement \( p \), it is called mutual knowledge. However under some circumstances that each agent knows that all other agents know \( p \) as well, then it is a common knowledge. Much of our social interactions assumes common knowledge as Lewis pointed out in his book *Convention* [37]. However, the technical definition of common knowledge, which requires infinite reasoning levels of all agents, rarely happens in real life [56]. At the same time, behavioral scientist do show that cooperation happens with unexpectedly high probability [28] [18] [3]. Therefore, understanding how groups actually coordinate is important.

In the logic system we build, we can show how beliefs can be adopted throughout the whole network from a single source. In our construction, communication can happen at any time between any two connected agents. But once the expert starts to communicate with her neighbors, the network will converge to her beliefs fairly quickly. We are also able to show that this convergence is pretty stable.

This result allows us to think about a concept that shares some characteristics of common knowledge. We call it *Distributed Common Knowledge*.

1. **Locally**: From the process of the expert influence, we can see that once the expert influenced all his neighbors, there exists pair-wise common knowledge that the non-experts’ beliefs are updated.

2. **Globally**: Evidently, every agent obeys the two axioms. If each agent also knows the number of total agents, they can deduce that within certain amount of time\(^3\) the network will converge.

With these two kinds of knowledge, even though they don’t know the exact structure of the network and the exact knowledge of each agent’s belief, it is still clear how group coordination can happen.

\(^3\)This is one of our results. It shows precisely how quickly the influence and convergence can take place.
8.4 Simulations of Knowledge Influence

We will simulate our new logic system in a software called NetLogo. We want to study the following aspects:

- Different network structures: Does the network structure impact the expert influence?
- Different positions for experts: Does the position of an expert impact her influence?
- Multi-cultural networks: How do we simulate multiple influences in one network?
- Scaling up: When the number of agents increases dramatically, how does influence play out?

The last aspect of the simulation is particularly interesting, especially in connection to distributed common knowledge. A recent real life example at a large scale would be using FireChat, a one-to-one communication phone app, during the protests in HongKong. During a few protests, the number of participants surged. It had led to complete overload of both wifi and telecommunication networks. However, protesters were still able to coordinate the actions through FireChat, which relies simply on the Bluetooth or Wi-Fi links that exist between phones.
Chapter 9

Time-line

The first and second sections of my dissertation include work that were done with Rohit Parikh. Work from the first section on *expert influence* has appeared on 14th Asian Logic Conference (ALC) in Jan. 2015 in Mumbai, India. Some preliminary results of the second section on *distributed common knowledge* is presented in the 11th International Conference on Technology, Knowledge and Society in Feb. 2015 at University of California, Berkeley, California. The conference has also invited us to submit the journal version, which we plan to finish before April.

Currently, I am also focusing on the last section of my dissertation, which is the simulation of knowledge influence. I have already learned the software, NetLogo, which is often used in such agent-based simulations. In addition to that, I have found a few related existing software libraries for my own simulations. I expect to have some results of the work in the spring or early summer of 2015.

Following is a list of conferences that are suitable for publishing this work:

- LORI-15: 5th Conference on Logic, Rationality, and Interaction
- IJCAI-15: International Joint Conferences on Artificial Intelligence
- ECKM-15: 16th European Conference on Knowledge Management
- ICAART-16: 8th International Conference on Agents and Artificial Intelligence

I plan to spend the fall semester of 2015 to write my dissertation. With this schedule, I should be able to defend my dissertation during the spring semester of 2016.
Bibliography


