Making social representations machine-readable: An algorithm design

Alec McGail

May 14, 2018

1 Introduction

This is a motivational essay for myself. My aim is to convince myself of the "whys" and "hows" of an ambitious set of projects for Summer 2018. The ambition I will motivate in this paper is to design a computer program which interacts with individuals one-on-one over online chat. This program will be linguistically fluent enough to *understand* what individuals tell it. It will understand what claims its interlocutor uses to support which others, and how one person's claims may or may not relate to other people's claims. Its understanding will be culturally formed, dynamic, and learning, with the hope of building a single mind to understand qualitatively many individuals, bridging the gap between qualitative and quantitative sociology.

Using this algorithm, I ideally want to collect machine-readable datasets which codify *representations* people maintain about their world, and especially about their social world. Conveniently, the language of Social Representation Theory describes well what I mean by "representation". Moscovici is seen to have initiated the field, building on Durkheim's concept of collective or common consciousness (for instance Durkheim, 1895, p. 40). According to Lopes and Gaskell (Sammut et al., 2015, chapter 3),

> [Moscovici] proposed that social representations are systems of values, ideas and practices which give order and meaning to the material and social world, with which members of a community exchange views, and make sense of their world and their individual and group history.

Social representations are systems of common sense which are used to justify certain human practices (ibid., Chapter 1)[1]. The specific representations I will analyze are individuals' *beliefs* and *systems of belief*. As is typical, contention about the proper definition of beliefs does arise, but I will bracket a deep dive into this discussion and define beliefs as, "enduring, unquestioned ontological representations of the world [which] comprise primary convictions about events, causes, agency, and objects that subjects use and accept as veridical," following Connors and Halligan (2015). I do not necessarily worry about whether held beliefs are objectively true or false, or shared by others, although a restriction along these axes is interesting and useful for more specific substantive investigations. In this quite broad definition I refer to conscious and unconscious beliefs, metaphysical beliefs, political beliefs, the meaning of words or phrases, beliefs about past events, about other individuals or their beliefs, about relations between ideas or beliefs in the abstract, etc.

By systems of belief, I refer to lay logic, commonsense reasoning, and to the practice of justification. I am interested in how individuals justify or condemn abstract beliefs – those of morality and ethics, platitudes and political understandings – by means of reason, analogy, or whatsoever other discursive tool developed for the purpose. I am also interested in the relationship understood by individuals between these more abstract beliefs and their actions or the actions of people around them[2]. I will consider primarily which beliefs are used as justification for which others, which beliefs seem to the individual to contradict which others, and other psychological and discursive analogues to formal logic[3]. Beliefs exhibit other forms of structure, however, which are also of interest. For instance, beliefs can refer to other beliefs (for example, a person's belief that "A implies B", or a belief about the social impact of holding a belief). They also may have the same referent, mobilize the same cultural symbols, or what is not always the same thing, simply be expressed using the same words. Of course beliefs can also be completely unrelated. In all this, it is useful to keep in mind that this structure is here defined in the mind or practice of a single person, begging for extension to collective belief systems[4].

[1] Abric (1993) has built an elaborate evolutionary theory of representation maintenance within societies. The author distinguishes between the core and peripheral system of a representation. The core is built on shared beliefs, linked to collective memory and the history of the group, while peripheral beliefs are flexible and heterogeneous, and these systems have distinct and identifiable functions related to the maintenance of the system.

[2] I simply adopt Weber's definition on p. 4 of *Economy and Society* of action and social action: "Sociology (in the sense in which this highly ambiguous word is used here) is a science concerning itself with the interpretive understanding of social action and thereby with a causal explanation of its course and consequences. We shall speak of 'action' insofar as the acting individual attaches a subjective meaning to his behavior – be it overt or covert, omission or acquiescence. Action is 'social' insofar as its subjective meaning takes account of the behavior of others and is thereby oriented in its course."

[3] I readily admit that when handled by individuals, the way propositions are formed and held, justified and manipulated differ wildly from what one would see in the proofs of formal logic. E.g. Converse (1964): "There is a broad gulf between strict logic and the quasi-logic of cogent argument." The analogy is useful, however, as long as it is recognized as such.

[4] See Zerubavel (1997) for an introduction to "cognitive sociology," which refers to collective beliefs, symbols, ways of thinking, history, etc.)

2 The pragmatics and problems of collecting social representations

Language is intimately connected to, and indeed inseparable from, meaning. In most of the statements I make in my everyday life, I feel I am successfully communicating, and that I am being understood by my interlocutor. I may mean multiple things, and even different things to different listeners. Sometimes I am incentivized to communicate clearly, for example in order to prevent financial loss or the loss of a valued relationship, or to explain to myself why this is a good summer project. Other times I am incentivized to remain vague, for example so that a joke will get more laughs when the punch line comes. But there remains the fact that we as individuals are able to communicate, and many times are convinced we are doing so successfully.

I would like to emphasize that I do not believe it is a simple matter to "collect" social representations, and that in some cases it is not possible. Whatever method we as researchers use to collect social data, it is not possible to "collect a representation," with exact precision. Qualitative and survey methods, personal anecdotes and news reports, none will convince us with full certainty what is *really* going on in people's minds. Did this even need to be stated? It is perfectly agreeable. But this does not stop us from pursuing such methods, or sometimes thinking we understand something about people! These sources of information are valuable, and we as social scientists feel they teach us something about people. When collecting datasets of social representations, I will be unable to fully understand people, and know with certainty that I do in fact understand them, or in fact that there is even something to understand. This problem, however, is inescapable and ubiquitous in social research, and I do not think any sociologist or anthropologist would claim that this should prevent researchers from attempting to collect them.

Language is the only means through which researchers can access social representations, i.e. the systems of meaning which organize people's lives. With the possible exception of measuring a person's brain activity, there seems no reasonable way to gain access to social representations without the use of language. Think about the sort of non-linguistic behavior we can observe in people. We can observe what products they buy, where they drive, count them, etc. These behaviors can be representative of meaning, if we acknowledge how influential meaning is to behavior, but we as researchers can never really get at meaning unless we go and ask them why, or what they are thinking about. We could imagine that one or another of these behaviors, for example a specific signal someone makes with their hands, has significant semantic meaning to the individual making the signal, and presumably to the person to which the signal is directed. But again, seeing this signal would give the researcher no information about what is going on in the person's head unless they explain it to the researcher in words. Along the same lines, sign language is no less a language than a spoken language. Thus the distinction between *just* behavior and what we consider speaking a language is exactly the distinction between non-meaning and meaning. Any behavior which contains meaning *becomes* a form of language, of communication. Thus language, spoken language in particular, is *the best way* to get at people's representation of their world.

Admittedly, understanding and explanation are both quite hard for humans to do. These take quite a bit of effort and attention. If they do not, it is only because the effort has already been expended earlier, and the individuals have developed a functioning and clear language. But we humans are able to explain and understand if we expend the effort and ingenuity necessary. We have learned tactics to make sure communication is actually happening. For example, understanding requires a careful and repeated accounting of whether or not the listener did in fact understand. As listener, you are constantly examining the consistency of what is being said with those things you already know. If something seems amiss, you can ask a clarifying question to regain consistency. Likewise, if I am explaining a long series of events to you, it is very comforting for me to hear you ask questions which indicate your understanding of what I am saying. Without confirmations such as these, the conversation simply would not go on. I as story-teller would quickly realize that my words are not being understood, and cease telling my story.

Most common among currently accessible machine-readable datasets which capture meaning is the structured survey. Such surveys rest fundamentally on language, because when answering surveys respondents must be able to understand and respond to questions. The pitfalls of survey research are complicated and somewhat well-known. Surveys answers must be rotated to ensure answer ordering is not systematically affecting the answers individuals give. Those who administer surveys must follow strict scripts, being careful not to influence their answers in any systematic way (or at least not in a way systematically differing from person to person). The questions themselves are prone to misunderstandings, and must be tested first on a small sample, accompanied by qualitative interviews to confirm that the questions are correctly and consistently measuring what was intended. The wording of the questions can be to the interviewee, and even when clear can be understood heterogeneously among different individuals (see Groves et al., 2004, section 7.3.2, for examples).

3 Bridging quantitative and qualitative social science

This paper is a call for the quantitative coding of so-called social representations. I should explain why it is important to collect social representations, i.e. systems of meaning, in a way that is machine-readable. The distinction between quantiative and qualitative sociology consists largely in the former's coding of social acts into data-points – of making information about people machine-readable. This analogy, along with the success of quantitative sociology (where it succeeds, of course) should sensitize the reader to the benefits of extending such coding to social representations. The main outcry against quantitative sociology is that it misses some crucial aspects of the situation. In coding there has been a loss of information. Even worse, this loss of information may be systematic, skewing observation, and generating false knowledge.

The same criticism, however, can be easily laid against qualitative social science. And in some sense this criticism can be made harsher, as biases in qualitative social science are less detectable and not subject to direct analysis. They lie in the cognitive categories and biases of the researcher, including their theories and what they expect to find. Bryman (1988) considers this topic in depth, and summarizes (p. 73): "What has proved to be disquieting to some commentators, both within and outside the qualitative approach, is whether researchers really can provide accounts from the perspective of those whom they study and how we can evaluate the validity of their interpretations of those perspectives." The benefits of a quantitative social science, where it is possible, are numerous. Theories can be stated precisely, even in mathematical or formal-logical notation, and tested against the aforementioned data. A single analyst can consider thousands, or tens of thousands of individuals ("cases"), and verify theories by observation on a scale not possible if the data were not machine-readable. I intend to make exactly these sorts of large-scale observations about the interviewees' representations of the world.

As a further motivation for this sort of data to be collected, I take a theoretical stance about social understanding and behavior. I claim that we as human beings work unceasingly to understand what is going on around us[5]. If we are not in a constant effort to understand, it is only because we have already figured it out. It is not confusing to us anymore, it is already understood.

More radically, I claim that interviewees know more about why they are acting in this or that way than a sociologist can hope to reveal based on behavioral data. That is, the only way we can hope to get to the actual reasons for actions (or if such reasons exist, and when they become conscious or are constructed by the individual) is by collecting individuals' perceptions of what they are doing and why, what others are doing and why, etc. It is by collecting their social representations through interactive dialogue.

4 Previous attempts and the promise of a more effective method

There are methods for the understanding of the structure of beliefs which only require opinion polls, which are widely available. Prototypically, they will identify the beliefs' structural relationships by identifying positive or negative correlations between the affirmation of two beliefs. When the correlation is quite high, the researcher infers (correctly) that whenever a person holds belief A they also hold belief B, and many times continue to surmise that $A \implies B$ is also believed among the population. In this way, they uncover a relational network structure of beliefs. In one of the first such analyses, Converse (1964) examined the coherence of belief systems by demographic traits, specifically socio-economic status. Converse defines the coherence of a belief system as how strongly correlated the beliefs are in aggregate among a group[6], and shows that elites' belief systems are more coherent than the masses. Similar analyses re-applying this methodology in other substantive contexts followed (see Axelrod, 1967; Luttberg, 1968; Bennett, 1973).

[5] In saying we as living human beings attempt to understand, and come to understand, I only mean that we hold beliefs, bracketing the veracity or reasonableness of these beliefs. For example, the schizophrenic patient who maintains delusions that he is being followed by government agents has a very clear *subjective* understanding of what is going on. He has strong beliefs, and reasons for believing them, and it does not matter much for an explanation of his behavior that his beliefs seem false to us. This subjective understanding is always what I will mean by an understanding. It's a system of meaning that makes the world make sense.

[6] Converse used the Goodman-Kruskal gamma for aggregation. The more standard measure later became the average of pairwise association measures.

This sort of work is able to uncover homogeneous patterns of belief across a large population who have been surveyed, and benefits from its nonparametric approach to understanding belief systems. That is, these methods let the data speak for itself. Unfortunately there are many things this data simply cannot say. For instance, the researcher cannot identify that there is some *causal*, *logical*, or *argumentative* relationship between the beliefs A and B in the minds of individuals when they show "$A \implies B$". The dataset simply does not speak to this. Furthermore, the possibility to identify heterogeneity in the structure of belief systems across the population is very limited. It is true that the methods of Goldberg (2011) focuses exactly on this heterogeneity, identifying internally coherent "collective logics" within a population (see also Im, 2013; Baldassarri and Goldberg, 2014), however this method can at most find a handful of distinct collective logics. In the case of more extreme heterogeneity in the structure of beliefs across individuals, these methods fall mute. This limitation is largely due to the original surveys' collection of *beliefs*, and no corresponding collection of discourses of justification, or in general the perceived relationships between beliefs.

In proposing a more effective methodology, I first wish to clarify what I believe to be possible. I will explain the ideal goal to aim at in the creation of machine-readable datasets which codify representations people maintain about their world. But also I want to spend a few paragraphs fanatically imagining possibilities.

Essentially, I believe it is possible to write a specific sort of algorithm. This algorithm can interact with a human being, and come to understand what they mean when they express themselves. In fact, a human who is really trying could explain anything (as long as it can be explained) to this algorithm. The algorithm would come to understand another person in the same way humans are able to understand each other – through language. And the reverse is also possible. Such an algorithm, once it has come to understand something, would be able to explain it just as well to others. Indeed, if a machine is truly able to read representations, so defined, it would be possible to write an algorithm to express these representations *in words*. A crucial competency of such an algorithm would be to not only learn facts based on interaction with humans, but to learn new ways to communicate.

I propose this with knowledge that such endeavors have been undertaken before (see Abu Shawar and Atwell, 2007, for a history of the "chatbot"), and I feel strongly that any lack of success in conversational fluency is not inherent. There is no fundamental technical limitation preventing a computer from becoming conversationally fluent in the same way that humans are. Indeed, the sociologist of science Harry Collins, who is highly skeptical of the possibility for machines to be actually conversationally competent, insists the reason why is because the current way that the machine is trained is not by immersing it in the everyday interactions of human society[7]. I wish to rectify exactly this, by immersing a conversational machine in human interaction. This isn't new. For example, CleverBot "learns" by interacting with people, yet it is still not able to *understand* in any meaningful way what is being said. I claim that the lack of success of previous attempts, of real machine understanding, is due to a lack of an explicit and reflexive understanding of understanding. We simply do not have a good account of how we come to understand each other, and what exactly it means to come to such an understanding. But this is not an insurmountable

[7] Taken from a second-hand account of a talk given at Cornell University in 2018.

hurdle, and I believe crucially that we can understand understanding.

Let me describe more concretely the sort of analyses we could perform given well-organized datasets of belief systems on an individual level (i.e. meaning systems, lay logic, common sense). Representations are not necessarily static qualitites of individuals, and may change over time, or never be defined whatsoever. Now imagine if one was to collect high-density longitudinal data on lay logic and representations via smart-phone or other easily accessible device. Through the device, individuals are asked to clarify and defend what they say or have said. They are asked about their previous understandings of things, and if these understandings seem flawed to them now, to explain why. With this sort of data we can perform many analyses completely inaccessible to current methodologies, at scale. For example, we can identify stable representations (such as the tenets of a morality or words defined in a language), dynamic representations (such as conceptions of others), or even representations which might have no discernable regularity over time.

On another more radical note, the algorithm and how it works are understood by its programmers, and thus should be explainable to the algorithm. Much of the algorithm's functionality, especially with regard to linguistic fluency, could be constructed based on what it learns from interacting with people. As a simple example, the algorithm could understand from speaking to people what are the rules of grammar, and how these rules relate to how the algorithm goes about constructing sentences.

5 The intermediary as facilitator of more efficient communication

I will end the paper on a final theoretical note, which I believe to be quite an important motivation for the construction of an algorithm which understands. There is a deeper and more pragmatic need for such an algorithm in today's society. In the up-swing of the technological age, it is undeniable that the world has become extremely complicated in ways never before seen (as far as we know). And as we grow more complex, there is an urgent need for cooperation, communication, and understanding. I argue that the capability of a computer a translator, interpreter, understander, and explainer – in whole, as intermediary – increases greatly the efficiency of communication.

A fundamental property of explanation is that it must be done by someone to someone else. That is, the person explaining must consider their audience. Textbooks in Calculus are written *to* researchers, undergraduate students, children, biologists, chemists, etc. They are written in a language specialized to the readers, attending to the body of knowledge which defines the audience, with motivating examples in the field of the audience. Typically, one person constructs an explanation specifically for another person, or at the very least for a type of person. Explanation is best done in person, one-on-one, as it allows for clarification and questions, and these opportunities make this *explaining to* much more direct and effective. But if, for example, 20 different people all want to communicate what they find to each other, each person must communicate clearly their understandings *to each* *other person*[8]. Thus $40 * 39 = 1590$ directed communications would ensue, as each *explained to* the other, accepting questions and comments. This is where the benefit of intermediary comes in. If every person is all able to explain to an intermediary, and that intermediary is able to understand the language and understandings of each, the algorithm would be an indefatigable resource for explanation of any understanding by any person to any other person. Each person would need explain themselves only once, and crucially would give access to their explanations and understandings to researchers.

References

Abric, Jean Claude. Central System, Peripheral System: their Functions and Roles in the Dynamic of Social Representations. *Papers on Social Representations*, 2(2):75–78, 1993.

Abu Shawar, Bayan and Atwell, Eric. Chatbots: are they really useful? *LDV Forum*, 22 (1):29–49, 2007. ISSN 0175-1336. doi:10.1.1.106.1099.

Axelrod, Robert. The Structure of Public Opinion on Policy Issues. *The Public Opinion Quarterly*, 31(1):51–60, 1967. doi:10.1086/267481.

Baldassarri, Delia and Goldberg, Amir. Neither Ideologues nor Agnostics: Alternative Voters' Belief System in an Age of Partisan Politics. *American Journal of Sociology*, 120 (1):45–95, 2014.

Bennett, Stephen Earl. Consistency Among the Public's Social Welfare Policy Attitudes in the 1960s. *Midwest Political Science Association*, 17(3):544–570, 1973.

Bryman, Alan. *Quantity and Quality in Social Research*. Routledge, New York, NY, 1988.

Connors, Michael H. and Halligan, Peter W. A cognitive account of belief: A tentative road map. *Frontiers in Psychology*, 5(1588):1–14, 2015. ISSN 16641078. doi:10.3389/fpsyg.2015.01588.

Converse, Philip E. The nature of belief systems in mass publics (2006). *Critical Review*, 18(1-3):1–74, 1964. ISSN 0891-3811. doi:10.1080/08913810608443650.

Durkheim, Emile. *The Rules of Sociological Method (1982)*. The Free Press, New York, NY, 1895.

Goldberg, Amir. Mapping Shared Understandings Using Relational Class Analysis: The Case of the Cultural Omnivore Reexamined. *American Journal of Sociology*, 116(5): 1397–1436, 2011.

Groves, Robert M.; Fowler, Floyd J.; Couper, Mick P.; Lepkowski, James M.; Singer, Eleanor, and Tourangean, Roger. *Survey Methodology*. John Wiley & Sons, Hoboken, New Jersey, 2004. ISBN 0-471-48348-6.

[8] In describing "different types of person," I am implicitly assuming that each could not communicate in the same way to all others.

Im, Dong-Kyun. Attitudes and Beliefs about Distributive Justice in China A dissertation presented. 2013.

Luttberg, Norman R. The Structure of Beliefs Among Leaders and the Public. *Public Opinion Quarterly*, 32(3):398–409, 1968.

Sammut, Gordon; Andreouli, Eleni; Gaskell, George, and Valsiner, Jaan, editors. *The Cambridge Handbook of Social Representations*. Cambridge University Press, Cambridge, 2015. ISBN 9781107323650. doi:10.1017/CBO9781107323650.

Weber, Max. *Economy and Society*. 1922.

Zerubavel, Eviatar. *Social Mindscapes: An Invitation to Cognitive Sociology*. Harvard University Press, Cambridge, Mass, 1997.

Note to self: This document contains 16 references.