The Role of Imagination in Social Scientific Discovery: Why Machine Discoverers Will Need Imagination Algorithms

Michael T. Stuart

Abstract

When philosophers discuss the possibility of machines making scientific discoveries, they typically focus on physics, biology, chemistry and mathematics. Observing the rapid increase of computer-use in science, however, it becomes natural to ask whether there are any limits to the kinds of things that machines can discover. For example, could machines also make discoveries in the most interpretive (i.e., least quantitative) regions of social science? Is there something about humans that makes us uniquely suited to studying humans? Is there something about machines that bars them from such activity? A close look at the methodology of interpretive social science reveals several abilities necessary for social scientific discovery, and one capacity is necessary to possess any of them, namely, imagination. In this paper I discuss what it would mean for machines to be discoverers in social science, in terms of the imagination algorithms that (I argue) they must possess.

The question of whether machines could discover arose early in the history of artificial intelligence.[1] Since then, machine learning algorithms have been developed, and there are now many putative examples of machine discoveries, for example: the BACON program (re)discovering Kepler’s third law, Coulomb's law and Ohm's law (Langley 1981); the KnIT program discovering features of a molecule important for cancer-prevention (Spangler et al 2014), and the

---

[1] Especially in the work of Herbert Simon and his students. See, e.g., Newell, Shaw and Simon (1958), Simon (1977, 1979), Bradshaw, Langley and Simon (1980), Bradshaw, Langley and Simon (1983); Langley, Simon, Bradshaw and Żytkow (1987); Langley and Jones (1988); Shrager and Langley (1990); Langley, Shrager and Saito (2002); Langley (2000); Dzeroski, Langley and Todorovski (2007).

Automated Mathematician discovering Goldbach's conjecture (Lenat 1982) (see also, e.g., Giza 2002; Gobet, Addis, Lane and Sozou 2014).

A common way to frame the possibility of machine discovery has been functionalist: if a machine can do all (or some crucial set of) the things required by the scientific method, such as generate hypotheses, perform experiments, write papers that pass the peer-review process, etc., it can discover. For this functionalist approach to succeed, it must be realistic about the scientific method. Since there is no (unique, finite) set of methods that counts as "the" scientific method, there is no list we can simply gather up and check off to determine whether machines can discover in general. The best a functionalist approach can do is take a specific scientific context and identify the conditions sufficient for discovery in that context. Perhaps several replications and a certain number of citations qualifies something as a discovery in one domain, while the five sigma standard applies in physics. Functionalist accounts like these are incapable of shedding light on the nature of scientific discovery itself, however, since we only learn what a set of scientists count as discovery.

There is another approach—which might be labelled transcendental—that would seek necessary conditions instead of sufficient ones. This sort of approach presupposes that we do discover, and asks what makes that possible. In the case of machine discovery, we ask what machines would have to be like, that is, what capacities they must possess as agents, in order to discover in science. Again, we must admit some contextuality: what is necessary for discovery in one domain might not be necessary in another. For our inquiry, I propose we choose a domain of science in which machines are at a disadvantage: social science. The idea is that if they can make discoveries about the complicated world of human behaviour, surely they can discover in less qualitative domains like physics and chemistry. After all, if it turned out that machines could *only* make quantitative discoveries in the "hard" sciences, the general question of machine scientific discovery would go unanswered.

To employ this transcendental approach, I argue in section 1 that we should characterize discovery as an action. In section 2, I develop this agent-centered characterization. In section 3, I extract what is necessary for agents to discover in the social sciences by an analysis of social science textbooks and methodology papers. And in section 4, I show that the ability to imagine is necessary for social scientific discovery. I conclude that machine discoverers must possess imagination algorithms if they are to discover in the full sense of scientific discovery (which must include social science). A second conclusion is that discussions about the future of machine discovery must address the possibility of machine imagination.

1 Reasons for an Agent-Centered Account of Discovery

There are many ways of characterizing scientific discovery,[2] but each of them portrays discoveries as either *events* or *objects*. According to the first (“agential discovery”), the *process* or *act* of discovery is emphasized. Think of Newton’s discovery of the law of universal gravitation; we talk about what Newton did, and how he did it. According to the second (“objectual discovery”), a particular *object* is emphasized. Think of penicillin; we talk about what it is and why it was important. To claim that only one of these is the “true” sense of discovery would be to introduce a false dichotomy: inquiry into both processes and products can illuminate the phenomenon of scientific discovery. Nevertheless, I focus on agential discovery for three reasons.

---

2 Important accounts include Kuhn (1962), Achinstein (2001), Hudson (2001), McArthur (2011), Schindler (2015), and the entries in Schickore and Steinle (2006). See Schickore (2014) for an overview.

First, objects of discovery are not counted as discoveries until there is some recognition of those objects *being* discoveries. It makes little sense to say that penicillin was a discovery in 30 000 BCE, or worse, that it was always a discovery. The number 0, the concept of democracy, and Snell's law do not exist in time, although they *became* objectual discoveries over some period of time. The temporal element of objectual discovery encourages us to take agential discovery as primary, since actions can be indexed to times and this is not true of all objects of discovery.

Second, the temptation to think of discovery as objectual instead of agential is at least partially a result of scientific rhetoric. Since the foundation of the Royal Society, science has been portrayed as objective by removing the traces of particular agents (see e.g., Schaffer and Shapin 1985). Discoveries are made by science itself—that is, by no one in particular—and this distances those discoveries from doubts that might otherwise attend them if portrayed as products of a practice carried out by biased and imperfect humans. This rhetorical move masks important agential aspects of science, and we need not take the mask for the face.

Finally, the products of discovery can be anything from bacteria to equations to methods, and it seems far more difficult to look for commonalities in the set of all things that have been discovered and ask if machines could produce them, than to ask what sort of action discovery is and whether machines can be the sort of agent to perform it.

So, what kind of action is discovery?

2 Elements of Agential Discovery

I propose we distinguish the following elements of agential discovery: 1) an agent (who discovers), 2) an object of discovery (that which is discovered), 3) a trigger event (that which prompts the discovery), and 4) an act of discovery (the agent's interpretation of the object, prompted by the trigger event).

The agent can be an individual or a community, whose mind can be extended or distributed. *The object of discovery* can be an idea, a fact, a value, a concrete entity, a process, a problem, a kind, an ability or a method. Anything, really. *The trigger event* often takes the form of an observation, inference, experiment, simulation, model manipulation, statistical analysis, or combination of these. It need not be intentional and can even be accidental. Generally speaking, any event can be a trigger event. Finally, *the act of discovery* is the agent's interpretation of the object of discovery, prompted by the trigger event. In the simplest cases, this interpretation is a mere categorization of the object of discovery. We must not confuse the act of discovery with the trigger event, however. The discovery of penicillin was not the Petri dishes left uncovered, or the mould growing on the dishes, or Flemming's walking into the lab and seeing the halo of empty space surrounding the mould in which there was no Staphylococcus. All of these objects and events (and others) jointly constitute the trigger event, and the discovery of penicillin must be different from this event because we want to be able to praise a discoverer for their discovery, and we cannot do this if the discovery simply *is* the trigger event, which need not include the intentional action of any agent.

Interpretation is the key to the action of discovery. I propose four specifications of interpretation that when satisfied (and combined with the other elements) yield what I think is a plausible explication of scientific discovery.

First, the interpretation of the object of discovery has to be *novel*—whether to the agent (personal discovery) or to the agent's epistemic community (historical discovery, see Boden 2004; this is also a requirement for Kuhn 1962, Schindler 2015, and Hudson 2001). An agent or community who discovers the same thing again still discovers it, but after the first instance we say that they *re*discover it. Something may be a personal discovery for an agent though only a rediscovery for the epistemic community. And one epistemic community can rediscover what another has discovered already. The same discovery can therefore be a discovery or a rediscovery depending on how broadly we understand the agent's epistemic community.

Second, for an interpretation to count as a scientific discovery, it must interpret the object of discovery as the solution to a scientific problem. This is too loose, however, because a scientist who learns that her assistant has been stealing lab equipment interprets an object of discovery (the lab assistant's actions) as the solution to a problem (the equipment going missing) in a scientific domain. So let us focus on acts of *theoretical* scientific discovery, which are those that solve theoretical scientific problems. Theoretical problems concern the phenomena studied by a scientific domain, while practical problems obfuscate our solving such problems. The distinction is contextual, but there are clear cases.

Third, a scientific discovery must not merely *appear* to solve a theoretical problem; it must actually solve it. (Or partially solve it for a partial discovery). Thus Poincaré wrote that a mathematical discovery has three steps, an unconscious combination of ideas, a flash of insight that suggests that one of these combinations solves the problem, and then the most important step: verification that the solution is correct.[3] Since we do not want to discuss the mere *feeling* of discovery (what William Whewell called "happy thoughts"), we must include some criterion of success (Achinstein 2001 and Hudson 2001). The kinds of solution that count as successful for a given problem will vary according to the context.

Fourth, in addition to successfully solving a theoretical scientific problem, we should require that the problem solved be significant. This is to preserve the intuition that discoveries are *special*; not all

3 He writes, "Discovery consists precisely in not constructing useless combinations, but in constructing those that are useful, which are an infinitely small minority. Discovery is discernment, selection" (Poincaré 1914, 51).

novel solutions to theoretical problems are discoveries. A problem is significant when its solution possesses some minimum value in whatever the relevant set of weighted scientific values are. Examples of such values include descriptive and predictive empirical adequacy, coherence with previous knowledge, explanatory power, fruitfulness, beauty, and simplicity.

To summarize, we have explicated agential scientific discovery in terms of an agent's novel interpretation of an object which successfully solves a significant theoretical scientific problem.

Can machines discover in this sense? Since trigger events and objects of discovery can be almost anything, I will leave these to one side and focus on the question of whether machines can produce interpretations that provide novel and satisfactory solutions to significant theoretical problems. According to some characterizations of NOVELTY, INTERPRETATION and SIGNIFICANCE, the answer will be, yes. As regards novelty, machines can do things that are novel (in the sense that they have not been done before) given the use of random number generators.[4] Second, machines can interpret objects if by interpretation we mean categorization. In this minimal sense of interpretation, sufficiently well-programmed computers do categorize certain states as solutions by checking them against programmed desiderata. And machines have significance encoded into them insofar as they are designed by scientists to address problems that are antecedently deemed significant, which means that machines can satisfy the third requirement as well. In sum, there are senses of NOVELTY, INTERPRETATION and SIGNIFICANCE that justify the use of the concept DISCOVERY as applied to machine behaviour.

But there are notions of NOVELTY, INTERPRETATION and SIGNIFICANCE that will be more difficult for machines to satisfy. These can be found in many corners of science. We want to know what is required if machines are to satisfy even the most difficult.

---

4 Thanks to Peter Sozou for this point.

Good candidates are qualitative discoveries, like those achieved in the social sciences.

3 Social Scientific Discovery

A broad range of methods fall under the banner of social science, from surveys and statistics to interviews and observations. Naively, we can draw a continuum from positivist to interpretivist social scientists. Positivists focus on observables such as behaviour, and explicitly endorse the more “objective” methods of science including standardized surveys and statistical methods. They aim to discover and understand generalizable patterns of human behaviour. Interpretivist approaches argue that positivist aims are in principle unachievable because the object of study and the researcher are of the same nature (agents with desires, beliefs and emotions). We can only have perspectives on perspectives. In light of this, interpretivists claim that social science is limited to subjective (but deep and honest) objects of discovery with limited generalizability.

If positivist social science employs the methods of “objective” science, then machines can make positivist social scientific discoveries if they can make “objective” discoveries in other fields of science. A harder question is whether machines can discover in the sense of interpretivist social science. That is, can machines produce new interpretations of the experiences and behaviour of agents as solutions to significant theoretical problems concerning how communities form, function and fall apart? The answer depends on the relevant senses of NOVELTY, INTERPRETATION and SIGNIFICANCE, which can be extracted from what sociologists do and how they teach.

Because it would be impossible to review the notion of discovery across all interpretivist social sciences, in what follows I focus on ethnography, which I take to be a paradigm case of an interpretivist social science *method* of discovery, shared across many subdisciplines of social science including sociology, anthropology, international relations, economics, and history.

The ethnographic method discovers by means of field studies, which include participant observation and interviews. The main goal of this kind of research method is to tell us why people think and behave in the ways they do in terms of the meanings they ascribe to surrounding people, objects and events. For example, Rosabeth Kanter's famous study, *Men and Women of the Corporation* (1977) found that secretaries in the 1970s had little or no upward mobility because of "trained incapacity," that is, "training that makes people fit for one position [but] progressively less fit for any other" (1977, 98). The skills developed by the secretaries studied by Kanter were highly specific to the needs of their particular bosses. While these skills might have provided job security, it also ensured that bosses typically would not let their secretaries move into higher positions.

Another example is Annette Lareau's *Unequal Childhoods* (2003), which focused on how differences in social class affected parenting styles among African American families. Middle class parents seemed to favour a style Lareau called "concerted cultivation," according to which children are allowed negotiation power concerning their life trajectories, are put into organized activities, and are taught to question authority. Lareau dubbed the other style “accomplishment of natural growth,” which she found to be favoured by working class families. This style gives far less negotiation power to the child, but also imposes less organized structure on daily life. As a result, children learn to respect authority figures while developing a sense of personal independence; both thought to be beneficial character traits in the context of working class life.

The above examples are typical of interpretivist social scientific discoveries; an agent or team performs observations and interviews motivated by a few general questions, and interprets the data to find patterns that explain why certain social phenomena take the forms they do. Now, in what sense are such discoveries novel and significant interpretations?

3.1 Ethnographic Novelty

According to two widely-used ethnography textbooks, “A report may be perceived as new and noteworthy…in at least three ways: through theoretical discovery, extension, or refinement” (Lofland et al. 2006, 173; see also Snow et al. 2003, 186).

Theoretical extension “involves extending pre-existing theoretical or conceptual formulations to groups or settings other than those in which they were first developed or intended to be used” (173). There are difficult cases where it is not clear how to extend a conceptual formulation or how to determine what counts as a new domain of application, but machines have been making more and more progress in emulating human scientific discovery using theoretical extension, for example in mathematics (Lenat 1997), physics (Langley 1981), chemistry (Żytkow and Simon 1986), and biology (Kulkarni and Simon 1990). We can expect this progress to continue.

Theoretical refinement is “the modification of existing theoretical perspectives through the close inspection of a particular theoretical proposition or concept with new field data” (Lofland et al. 2006, 173). Again, there is no reason to deny this sense of novelty to machines. If something like the dominant view in philosophy of science is correct—that theories are collections of models—then we already have programs that can analyze data to create or refine models, which in this sense of theory, satisfy the requirement (see, e.g., Valdés-Pérez 1995, and Kocabas and Langley 1998).

These aren't the most interesting sorts of novelty. Indeed, it has been argued that if computers are limited to these sorts of novelty, they cannot really discover (Gillies 1996). So let us turn to theoretical discovery.

Theoretical discovery requires categories to be devised and used in a way that interprets and explains data. This could be novel because the data is novel, the interpretation is novel, or both. For instance, we could use an old system of interpretation (e.g., looking at power imbalances) to examine at a new social phenomenon. Or we could look at old data through a new interpretive scheme. Or we could produce a new interpretation of new data.

The first of these options can be achieved by computation as a special case of theoretical extension. What is interesting about the second and third is the production of a new interpretation. In sum, to give an account of the novelty relevant for social scientific discovery, we need to look at the nature of ethnographic interpretation. But first, significance.

3.2 Ethnographic Significance

According to Lofland et al., a significant ethnographic thesis should do at least some of the following: 1) go against "common sense" or the "modern mind-set," 2) develop ideas that "establish broader implications," 3) be well-developed, that is, use or generate concepts that are elaborated in detail, with a good balance of conceptual elaboration and data presentation, and a high degree of interpenetration between the two, or 4) refine or extend existing social science ideas (2006, 177-181). Let us address these in turn with machine discovery in mind.

First, because computers are not typically programmed to reason as humans currently do, they have always been good at going against common sense and the modern mind-set. We might worry nevertheless that they have their own “computer common sense”: patterns of reasoning and expression from which they cannot deviate. Machine novelty could then be thought of as the power to break free from such reasoning styles. This is something machines can do, as programmers regularly soften the criteria that define problem solutions and appropriate methods, as well as adding stochastic elements and evolutionary algorithms that encourage such novelty.

The second way of achieving significance concerns theory extension, which we granted was within the purview of machines.

The third, which Lofland et al. call “developed treatment,” is something like the typical sense of “good” science: a significant study will be well-researched, have conclusions that are empirically or theoretically well-supported and were arrived at using scientific interventions that were carefully thought-out and cleverly brought-about. To repeat, however, there is no (unique, finite) list of methods that are the “good” or “scientific” ones. The only list we can have is necessarily open-ended. So, how could a computer go about choosing the *best* evidence and the *best* methods, when we cannot say in advance what those are? The computer will have to answer these questions by interpretation. Given the theoretical context, methods available, and data collected, it must interpret one or some of the methods as the most appropriate. And it must interpret one of the many possible explanations as the best or most plausible. In other words, for a machine to satisfy this requirement it must interpret well.

The fourth requirement is related to the second, which concerned theory extension. Here we are extending not from theory to phenomena, but from phenomena to theory. This is theoretical refinement, which, again, we granted, should be achievable for machines.

In sum, to recognize and achieve significant solutions to theoretical social scientific problems, the only crucial element that machines do not yet obviously possess is the capacity to interpret. Just as an ethnographer is able to interpret the significance of the acts, questions, explanations, and so on, that she observes, the social scientific community is able to interpret the significance of the ethnographer's results.

The cognitive requirements for producing a novel, significant interpretation must therefore be a superset of the cognitive requirements for interpretation alone. Novelty and significance are features of interpretations, which are attributed to particular interpretations *by interpretation*. If we want to uncover the cognitive requirements of producing a novel, significant interpretation, therefore, we will be off to a very good start if we can identify the requirements for interpretation in general.

3.3 Ethnographic Interpretation

We can identify three main interpretive methodologies: analytic induction, grounded theory and the extended case method.[5]

Analytic induction is a methodology according to which we produce claims of universal generality that we aim to refute using particular cases, over and over, until only one irrefutable universal explanation remains. To reach this final end point (if it was also novel and sig-

5 For a statement of analytic induction, see Znaniecki (1934) and Lindesmith (1947). For statements of grounded theory, see Glaser and Strauss (1967), Corbin and Strauss (1990), Glaser (1978), Strauss (1987), and Strauss and Corbin (1990). For statements of the extended case method, see Burawoy (1998, 1991, 2000). In what follows, I try to do my best to distil the methods of ethnographic interpretation, but I inevitably will not do them complete justice. Curious readers are urged to look at some of the sources listed for more details.

nificant) would be to discover. However, since we could never establish that any universal statement was forever immune to future disconfirmation (Katz 2001), more recent versions of analytic induction have relaxed this requirement, and simply focus on the method of hypothesis and counterexample. The role of interpretation in analytic induction is to turn data into counterexamples and to determine how to refine theory to avoid those counterexamples.

Grounded theory, in its strongest (and original) form, claims that theory must come from data and never the other way around. An often-quoted phrase is: “An effective strategy is, at first, literally to ignore the literature of theory and fact on the area under study, in order to assure that the emergence of categories will not be contaminated by concepts more suited to different areas” (Glazer and Strauss 1967, 37). As with analytic induction, criticism has softened grounded theory over time. For example, Strauss (one of the theory’s originators) came to admit that it is not realistic to think we could generate theory *purely* from data and data alone (Strauss and

Corbin 1994, 277). The main idea is now something like the following. As much as possible, we must try to let themes and patterns present themselves to us instead of imposing existing categorizations and theoretical assumptions on our data. Then, we test the emerging notions against future observations and interviews, until we feel sure we have understood them correctly. The relevant notion of interpretation here is quite complicated, and we will return to it in a moment.

In the extended case method, the emphasis is on extending and developing theory through qualitative methods. Given some background theory, a researcher enters the field with a host of specific hypotheses inspired by theory. Fieldwork is then “a sequence of experiments that continue until one’s theory is in sync with the world one studies” (Burawoy 1998, 17-18). Things not relevant to the theory can and should be ignored.

On a loose reading, these interpretive methodologies are not mutually exclusive. One can begin with a theory in mind to inform an investigation (as in the extended case method), but look for empirical counterexamples (as in analytic induction) and be ready to create new conceptual resources as necessary (as in grounded theory). However, the extended case method and analytic induction produce new interpretations only in the senses of theory extension and refinement discussed above. What we really want are cases where new theoretical understanding is born de novo from the data. This is the promise of grounded theory.

A great deal has been written on the process of interpretation in grounded theory. In general terms,

> You get *from* data, topics, and questions, on the one side, *to* answers or propositions, on the other, through intensive immersion in the data, allowing your data to interact with your disciplinary and substantive intuition and sensibilities as these latter are informed by your knowledge of topics and questions. (Lofland et al. 2006, 198-199)

This kind of interpretation "occurs continuously throughout the life of any qualitatively oriented project" (Miles and Huberman 1994, 10). It begins with coding the data, which is “the process of defining what the data are all about” (Charmaz 2001, 340), or “relating (those) data to our ideas about them” (Coffey and Atkinson 1996, 45-47) by “sorting your data into various categories that organize it and render it meaningful from the vantage point of one or more frameworks or sets of ideas” (Lofland et al. 2006, 200). The codes themselves are “names or symbols used to stand for a group of similar terms, ideas, or phenomena” (LeCompte and Schensul 1999, 55), “tags of labels for assigning units of meaning to information complied” (Miles and Huberman 1994, 56) or just “the labels we use to classify items of information as pertinent to a topic, question, answer, or whatever” (Lofland et al. 2006, 200).

Once some codes are established, we move to “focused” coding. One way to do this is to sort the codes into *units* and *aspects*, which combine into *topics*. The unit is the scope of the sample (Lofland et al. 2006, 122-132), for example, a practice (like getting ready for work), an episode (like divorce), an encounter (like a cocktail party), an organization (like a school), or a larger community (like a refugee camp). An aspect of a unit might be the beliefs, norms, ideologies, rules, self-identities, emotions, relations, etc., of the people in the unit. These combine to form a topic (e.g., the faith of fans in a sports team or the norms governing drug dealers). Topics should emerge and change naturally as the ethnography progresses.

These reflections ready us for writing “memos,” which are “the intermediate step between coding and the first draft of your completed analysis” (Charmaz 2001, 347). This is where we generate and develop possible explanatory relationships between data (organized in codes) and the topic. At this stage it is often helpful to draw concept maps and other sorts of diagrams to link all the ideas together. Again, coding and memoing must be done simultaneously with the data collection process, so that ideas can be brought back to the field, tested and updated.

But how do we select units and aspects and generate meaningful codes and memos? “Field researchers too rarely elaborate how they get from their data, topics, and questions to their findings and conclusions. The result is a kind of ‘black box’ or…‘analytic interruptus’… between the data-gathering and writing phases of the fieldwork enterprise that contributes to the sense that qualitative analysis is often the result of a mystical process or romantic inspiration” (Lofland et al. 2006, 211). While we can identify the parts or milestones of this process (coding, memoing, etc.), there still appears to be some extra cognitive leap that is left undescribed. And this is why, admitting that some parts of this process can be done by machines, Lofland et al. claim that data interpretation “is not a process that can be farmed out to independent analysts nor…to computers and various software programs” (2006, 196).

Why not? I think it has to do with interpreting *well*, as opposed to merely interpreting. Perhaps a machine can select units, aspects and topics, generate codes from data, and organize the codes into answers about a topic. But the thought might be that a machine cannot do this well. Perhaps this is because interpreting well consists in maintaining cognitive flexibility. We must not “become too locked in or committed to a particular theoretical perspective or line of argument too early in the analysis process" (Lofland et al. 2006, 217). We must not limit our analysis "to a single form of theoretical development or to one analytic model to the exclusion of another… [or to] allow the considerable computerized data filing and storage possibilities and the kindred qualitative data analysis programs to lull [us] into thinking that the hard analytic work is done once [our] logged notes are coded and stored" (218).

A good way to avoid such fixedness is to rephrase memos and codes, since "the sheer way a question (or answer) is phrased or worded can greatly facilitate or deter your thinking" (218). Another method is "periodic distancing": "Good field research is partly contingent on reaching a chronic yet healthy tension between closeness and distance or involvement and detachment" (219). An ethnographer must approach their fieldnotes "as if they had been written by a stranger" (Emmerson et al. 2011, 174).

Perhaps because computers operate using rules written in formal language, it is thought that they cannot maintain such flexibility. But despite current limitations, there are already programs that do not commit to a single interpretation, but continuously update their interpretations on receiving more data. Natural-language machine learning and data mining programs like Google’s DeepMind do this already, and will continue to improve. The formality of their programming language does not appear to be hindering them.

The issue is a deeper one. As mentioned above, interpretation can be a simple act of rule-governed categorization, but it can also be one of the most difficult cognitive acts that an agent can perform, requiring creativity, patience, imagination and insight. Perhaps it is some of these underlying cognitive powers that ethnographers suspect is missing from machines. In the next section, we will try to identify some of those cognitive powers that make the most difficult acts of interpretation possible.

4 Machine Interpretation

Building on the work of Peter Winch (1958) and Charles Taylor (1971) I will present five abilities any agent must possess in order to interpret well. I leave out abilities like collecting data and performing calculations, which machines can surely perform.

First, a set of collected data must admit of the possibility of making more sense than it currently does. In other words, it must be possible that an interpretation of the data could produce something that renders the data clearer, more intuitive, better understood, etc., because to interpret in social science is to do more than merely report a set of data as the meaning of these data. Following from this, a machine must have at least the following two abilities:

> 1) It must be able to distinguish between the *presentation* of a datum and the *meaning* of that datum.

In other words, it must recognize that there can be a difference between a thing and what it means, and also when such a difference obtains. When someone says "I'm fine," they *present* themselves as being fine. They might also *mean* that they are fine. But they might not. If we always assumed that speakers meant exactly what they said, no additional sense could ever be made. So, for a machine to discover in ethnography it must make this distinction and be able to recognize cases where presentation differs from meaning.

> 2) Once this distinction is made and instances are identified in which meaning (seems to) differ from presentation, an interpreter needs a method for determining meaning.

We do not need to overcome the infamous indeterminacy of translation or interpretation here, since partial interpretation or partial grasp is perfectly fine in ethnography as an intermediate step towards understanding. But some way of getting from presentation to meaning is necessary. This is especially difficult where metaphors, loose speaking, body language or implicature in general are involved.

Next, meaning is only ever meaning *for*. There are no absolute meanings, or meanings *in vacuo*. An action might have one meaning for the actor, and a totally different meaning for the researcher, who looks at it in a different way. Because of this,

> 3) An interpreter must be able to identify the subject for whom a given meaning holds.

Without being able to say *who* means what, a machine interpreter cannot interpret, not least because the properties of the specific agent are needed to inform the interpretation.

To understand human behaviour we must understand not only the meanings attributed by actors to events and objects, but also the purposes for which actions are performed and the normative constraints that govern those actions (Winch 1958, 77). This is necessary if we are to give a full explanation of any behaviour: the purpose of intentional action is to achieve some end, which is desired for some reason. Therefore, in order to perform ethnographic interpretation,

> 4) An interpreter must be able to tell the difference between actions performed intentionally and unintentionally, and identify what the relevant reasons for action are.

One way to discover someone's intentions is simply to ask. But the answers we receive themselves require knowledge of intention to interpret. For example, we must know whether our subject intends to be deceptive before we can consider taking their answer at face-value. A second difficulty is that we cannot assume what a subject intends based on what they achieve. The contributions of irrationality and luck must be recognized. Otherwise, we interpret a lousy basketball player as intending to miss their shots, and people with cognitive biases as intending to ignore pertinent evidence or to deceive themselves.

Finally, and perhaps most importantly,

> 5) An interpreter must observe and track the differences between their worldview and the worldviews encountered in the field.

To understand someone, we must allow that they might not mean what we mean, see things as we do, desire what we desire, attribute the same level of importance to the same things, and so on. Because of this it is crucial for ethnographers to know what their own worldview is, so that they can tell when and how it informs their interpretation of the worldviews under study. “Do *they* mean *A* by *B*, or do I only think so because *A* is what *I* would mean by *B*?”

In sum, we have five abilities required for an agent to interpret in the most difficult cases of ethnographic discovery: the ability to 1) distinguish presentation from meaning, 2) identify meaning, 3) identify the “owner” of meaning, 4) identify reasons for behaviour (while leaving room for irrationality and luck), and 5) distinguish, track and translate worldviews. There are surely other relevant abilities, but at least these five are necessary.

Can machines possess these abilities? Instead of pretending to know what future machines will be capable of, I want to say what they would have to be like if they were to possess them. That is, we apply the transcendental approach again; first, we used it to find what abilities were necessary for an agent to discover in difficult cases of interpretivist social science, and now we use it to find what is necessary in order to have those abilities. Let us take each of the five abilities, one by one. 1) To distinguish between presentation and meaning, an agent must allow that there are several possible meanings for any given presentation (and vice versa). 2) To identify meaning, the machine must be able to present to itself options for semantic ascription other than what is immediately inferable from the data alone, and choose the best option. 3) To identify the owner of a meaning, the machine must be capable of taking up the perspective of an agent to see if a given meaning attribution is reasonable. This (like 2) requires what is called cognitive empathy: taking someone else's perspective (as opposed to emotional empathy, which is mirroring someone's emotions). Cognitive empathy also underlies abilities 4 and 5. The only way to identify someone's reasons for acting without being told what they are is to imagine that you have the properties (personal and contextual) of the actor, and then ask yourself what reasons you would have for acting. In other words, the ability to interpret others depends both on the ability to interpret yourself (Jackman 2003) *and* the capacity for cognitive empathy (to convert yourself mentally into an approximation of someone else). Finally, tracking the differences between one worldview and another and establishing semantic links that would enable translation between them requires cognitive empathy because we cannot compare *x* to *y* if we only understand *x*. We must first experience *y* (to some extent) to relate it to *x*. However we can only occupy one substantial worldview at a time. So to compare worldviews, we must be able to distance ourselves from our current worldview, get into another, and then switch back and forth to make comparisons. And this requires presenting ourselves with various interpretations of the same objects from different perspectives, which requires cognitive empathy.

Underlying all these abilities are two things. The first is the imagination. Imagination has no commonly accepted definition, but the basic idea is the ability to interact cognitively with objects and states of affairs not currently present to experience. This is what enables all acts of cognitive empathy: to put yourself in another's position, you *imagine* yourself acting under different constraints with some of your existing properties strengthened, and others diminished or removed. Imagination is therefore a fundamental capacity underlying ethnographic interpretation.

The second thing required by abilities 1) - 5) is the ability to interpret meaning, which is closely tied to our ability to create meaning. To interpret someone we often consider what we would mean if we were them. Consider an emotion term like "shame," which

> can only be explained by reference to other concepts which in turn cannot be understood without reference to shame. To understand these concepts we have to be in on a certain experience, we have to understand a certain language, not just of words, but also a certain language of mutual action and communication, by which we blame, exhort, admire, esteem each other. In the end we are in on this because we grow up in the ambit of certain common meanings. But we can often experience what it is like to be on the outside when we encounter the feeling, action, and experiential meaning language of another civilization. Here there is no translation, no way of explaining in other, more accessible concepts. We can only catch on by getting somehow into their way of life, if only in imagination. (Taylor 1971, 13)

In other words, we learn how to mean things in the context of acting and reacting, feeling, experiencing and "getting on" with the meanings of others, which are the ultimate source of our own meanings (see also Winch 1958, 81ff). And grasping meanings would be impossible without imagination. We must see that several meanings are possible for any given behaviour, and so we must guess which the correct attribution is and confirm this (in difficult cases) by something like experiment (Stuart 2015). And we communicate meaning to others when we believe that our audience will understand us, which is something we track by putting ourselves in their shoes, among other things.

To summarize, we have discussed the possibility of ethnographic machine discovery and I argued that the issue of interpreting natural language systems of meaning requires cognitive empathy and the ability to assign meaning, both of which require imagination. Perhaps it goes without saying that novelty and significance require imagination as well. To create something novel, we need imagination, which is "important for even the most minimally creative thought"

(Stokes 2014). And to tell which problems are significant requires knowing how much better off we are with a given solution, compared to where we would have been without it. And this requires imagination. Therefore, *if machines are to make the most difficult novel and significant interpretations required by social scientific discovery, they will require imagination algorithms*.

Could such algorithms exist? That is, could a machine cognitively interact with objects and states of affairs that are not currently available to their "experience"? In some senses, yes. Computers can propose counterfactual hypotheses to make certain inferences. Logic software does this for reductio ad absurdum proofs and conditional derivations. But this is not the same as entertaining something that is not present, something fictional, since in the case of the logic program, the machine is interacting only with symbols that *are* present to its "experience."

Concerning more substantial senses of imagination, like those required for cognitive empathy and the sharing of meanings, things are murkier. I conclude therefore on what may be a surprising note. The necessary conditions for agential scientific discovery include providing novel interpretations that solve significant theoretical scientific problems. And in order to say whether machines can meet the conditions necessary for such action, we first need a better understanding of the imagination and what cognitive powers are required for its operation, whether in human or machine. Unfortunately, philosophers and cognitive scientists are still very far from possessing such an account.

To conclude, those who feel skeptical *or optimistic* about the extent to which machines can discover in science have good reason to focus that skepticism (or optimism) on the nature and possibility of imagination algorithms. To do this in a detailed way, however, we first require a better understanding of the nature, operation and cognitive requirements of imagination.

Acknowledgments

Thanks to the organizers of the conference “Scientific Discovery in the Social Sciences,” and to the participants for helpful discussion. Thanks also to Nancy Nersessian, Marco Buzzoni, Markus Kneer, and Peter Sozou for comments on earlier drafts of this paper, as well as Susan Staggenborg (and Nancy Nersessian again) for generosity in sharing their knowledge of social scientific methodology.

References

Achinstein, P (2001) The book of evidence. Oxford University Press, Oxford

Atkinson P (1990) The ethnographic imagination: Textual constructions of reality. Routledge, New York

Boden MA (2004) The creative mind: Myths and mechanisms. Routledge, London

Bradshaw GL, Langley P, Simon H (1980) BACON4: The discovery of intrinsic properties. Proceedings of the third national conference of the Canadian society for computational studies of intelligence, pp 19-25

Bradshaw GL, Langley P, Simon HA (1983) Studying scientific discovery by computer simulation. Science 222:971-975

Burawoy M (1991) Ethnography unbound. University of California Press, Berkeley

Burawoy M (1998) The extended case method. Sociological Theory 16:4-33

Burawoy M (2000) Global ethnography. University of California Press, Berkeley

Charmaz K (2001) Grounded theory. In: Emerson RM (ed) Contemporary field research: Perspectives and formulations, 2nd ed. Waveland Press, Prospect Heights, pp 335-252

Coffey A, Atkinson P (1996) Making sense of qualitative data: Complementary research strategies. Sage, Thousand Oaks

Corbin J, Strauss A (1990) Grounded theory method: Procedures canons, and evaluative criteria. Qualitative Sociology 13:3-21

Dzeroski S, Langley P, Todorovski L (2007) Computational discovery of scientific knowledge. In: Dzeroski S, Todorovski L (eds) Computational discovery of communicable scientific knowledge. Springer, Berlin

Emmerson R, Fretz R and Shaw L (1995/2011) Writing ethnographic fieldnotes, 2nd ed. University of Chicago Press, Chicago

Gillies D (1996) Artificial intelligence and scientific method. Oxford University Press, New York

Giza P (2002) Automated discovery systems and scientific realism. Minds and Machines 22:105-117

Glaser B (1978) Theoretical sensitivity. Sociological Press, Mill Valley

Glaser B, Strauss AL (1967) The discovery of grounded theory: Strategies for qualitative research. Aldine, Chicago

Hudson RG (2001) Discoveries, when and by whom? The British Journal for the Philosophy of Science 52:75-93

Jackman H (2003) Charity, self-interpretation, and belief. Journal of Philosophical Research 28:143-168

Kanter R (1977) Men and women of the corporation. Basic Books, New York

Katz J (2001) Analytic induction. In: Smelser NJ, Baltes PB (eds) International encyclopedia of the social and behavioral sciences. Elsevier, Amsterdam, pp 480-484

Kocabas S, Langley P (1998) Generating process explanations in nuclear astrophysics. Proceedings of the ECAI-98 workshop on machine discovery, Brighton, pp 4-9

Kuhn TS (1962) The structure of scientific revolutions. University of Chicago Press, Chicago

Kulkarni D, Simon HA (1990) Experimentation in machine discovery. In: Shranger J, Langley P (eds) Computational models of scientific discovery and theory formation. Morgan Kaufmann, San Mateo

Lane P, Sozou P, Addis M, Gobet F (2014) Evolving process-based models from psychological data using genetic programming. In: Kibble R (ed) Proceedings of the 50th Aaniversary convention of the AISB

Langley P (1981) Data-driven discovery of physical laws. Cognitive Science 5:31-54

Langley P (2000) The computational support of scientific discovery. Int J Human-Computer Studies 53:393-410

Langley P, Jones R (1988) A computational model of scientific insight. In: Sternberg R (ed) The nature of creativity. Cambridge University Press, Cambridge

Langley P, Simon H A, Bradshaw G L, and Żytkow J M 1987 Scientific discovery: Computational explorations of the creative processes Cambridge, MA: MIT Press

Langley P, Shrager J, Saito K (2002) Computational discovery of communicable scientific knowledge. In: Magnani L, Nersessian NJ, Pizzi C (eds) Logical and computational aspects of model-based reasoning. Kluwer Academic, Dordrecht

Lareau A (2003) Unequal childhoods: Class, race, and family life. University of California Press, Oakland

LeCompte MD, Schensul JJ (1999) Journeys through ethnography: Realistic accounts of fieldwork. Westview Press, Boulder

Lenat DB (1977) Automated theory formation in mathematics. Proceedings of the fifth international joint conference on artificial intelligence. Morgan Kaufmann, Cambridge, pp 833-842

Lenat, DB (1982) AM: Discovery in mathematics as heuristic search. In: Davis R and Lenat D (eds) Knowledge-based systems in artificial intelligence. McGraw-Hill: New York

Lindesmith A (1947) Oppiate addiction. Principia Press, Bloomington

Lofland J, Snow D, Anderson L, and Lofland LH (2006) Analyzing social settings, 4th ed. Wadsworth, Belmont

McArthur DJ (2011) Discovery, theory change and structural realism. Synthese 179:361-376

Miles MB, Huberman AM (1994) Qualitative data analysis: An expanded sourcebook, 2nd ed. Sage, Thousand Oaks

Newell A, Shaw JC, Simon HA (1958) Elements of a theory of human problem solving. Psychological Review 65:151-166

Poincaré H (1914) Science and method. Thomas Nelson and Sons, London

Schaffer S, Shapin S (1985) Leviathan and the air-pump. Princeton University Press, Princeton

Schickore J (2014) Scientific discovery. Stanford online encyclopedia of philosophy. Available here: http://platostanfordedu/entries/scientific-discovery. Accessed 15 Nov 2016

Schickore J, Steinle F (eds) (2006) Revisiting discovery and hustification: Historical and philosophical perspectives on the context distinction. Springer, Dordrecht

Schindler S (2015) Scientific discovery: That-whats and what-that’s. Ergo 2:123- 148

Shrager J, Langley P (eds) (1990) Computational models of scientific discovery and theory formation. Morgan Kaufmann, San Mateo

Simon HA (1977) Models of discovery. D Reidel, Dordrecht

Simon HA (1979) Models of thought. Yale University Press, New Haven

Snow D, Morrill C, Anderson L (2003) Elaborating analytic ethnography linking fieldwork and theory. Ethnography 4:181-200

Spangler S, Myers JN, Stanoi I, Kato L, Lelescu A, Labrie JJ, Parikh N, Lisewski AM, Donehower L, Chen Y, Lichtarge O, Wilkins AD, Bachman BJ, Nagarajan M, Dayaram T, Haas P, Regenbogen S, Pickering CR, and Comer A. (2014) Automated hypothesis generation based on mining scientific literature. Association for Computing Machinery's digital library. DOI: 10.1145/2623330.2623667

Stokes D (2014) The role of imagination in creativity. In: Paul ES, Kauffman SB (eds) The philosophy of creativity: New essays. Oxford University Press, Oxford

Strauss A (1987) Qualitative analysis for social scientists. Cambridge University Press, New York

Strauss A, Corbin J (1990) Basics of qualitative research: Grounded theory procedures and techniques. Sage, Newbury Park

Stuart MT (2015) Philosophical conceptual analysis as an experimental method. In: Gamerschlag T, Gerland D, Osswald R, Petersen W (eds) Meaning, frames and conceptual representation. Düsseldorf University Press, Düsseldorf, pp 267-292

Taylor R (1971) Interpretation and the sciences of man. The Review of Metaphysics 25:3-51

Winch P (1958) The idea of a social science and its relation to philosophy. Routledge, London

Valdés-Pérez RE (1995) Machine discovery in chemistry: New results. Artificial Intelligence 74:191-201

Znaniecki F (1934) The method of sociology. Farrar and Rinehart, New York

Żytkow JM, Simon HA (1986) A theory of historical discovery: The construction of componential models. Machine Learning 1:107-137