

Chapter 4

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 discoveries in physics, biology, chemistry and mathematics. Observing the rapid increase of computer use in science, however, it becomes natural to ask whether there are any scientific domains out of reach for machine discovery. For example, could machines also make discoveries in qualitative social science? Is there something about humans that makes us uniquely suited to studying humans? Is there something about machines that would bar them from such activity? A close look at the methodology of interpretive social science reveals several abilities necessary to make a social scientific discovery (such as cognitive empathy and the ability to assign meaning) and one capacity necessary to possess any of them is imagination. Novel and significant interpretations required by social scientific discovery require imagination. For machines to make discoveries in social science, therefore, they must possess imagination algorithms.

Keywords Discovery · Ethnographic interpretation · Ethnographic method · Ethnographic novelty · Ethnographic significance · Imagination algorithm · Interpretive science · Machine discovery · Machine interpretation · Philosophy of social science

The question of whether machines could discover arose early in the history of artificial intelligence.¹ Since then, machine learning algorithms have been developed,

¹Especially 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), Shragar

M. T. Stuart (✉)

Centre for Philosophy of Natural and Social Science, London School of Economics and Political Science, London, UK

e-mail: m.stuart@lse.ac.uk

and there are now many putative examples of machine discoveries, for example: the BACON program that discovered 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 Automated Mathematician discovering Goldbach's conjecture (Lenat, 1982) (see also e.g. Giza, 2002; Lane, Sozou, Addis, & Gobet, 2014).

A common way to frame the possibility of machine discovery has been functionalist: if a machine can carry out some crucial set of processes, such as generating hypotheses, performing experiments, writing papers that pass peer-review, etc., it can discover. The thought is that scientists make discoveries with certain methods, so if machines can use those same methods, they can discover. This might be an effective framing if we want to suggest that machines can make some discoveries as opposed to none at all, but it will not tell us much about the limits of machine discovery. By analogy, teaching a computer to buy a canvass, paint with certain brushstrokes and sell the painting to a gallery might justify the claim that machines can do some art, but this would not tell us what kinds of art machines are capable of making.

A different kind of approach (call it "transcendental") would seek a set of necessary conditions for scientific discovery instead of sufficient ones. This approach presupposes that we do discover, and asks what makes this achievement possible. Applied to the case of machine discovery, we ask what capacities machines must possess as agents in order to discover in science. This begins in a piecemeal way since the features necessary for discovery in one domain might not be necessary in another, but the full account should provide the (perhaps disjunctive) combined set of features necessary for discovery in all domains of science. Since I am interested here in the *limits* of machine discovery, I propose we look at the social sciences, where no documented cases of machine discovery yet exist.

I argue in Sect. 4.1 that we should characterise discovery as an action (part of a discovery event). In Sect. 4.2, I develop this characterisation of discovery. In Sect. 4.3, I extract what is necessary for agents to discover in social science by an analysis of social science textbooks and methodology papers. In Sect. 4.4, I show that the ability to imagine is necessary for social scientific discovery, and 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).

and Langley (1990), Langley, Shrager, and Saito (2002), Langley (2000), Dzeroski, Langley, and Todorovski (2007).

4.1 Reasons to Pursue an Action-Centered Account of Discovery

There are many ways of characterising scientific discovery,² but each of them portrays discoveries either as *events* or *objects*. According to the first, the *process* or *act* of discovery is emphasized. Think of Newton's discovery of universal gravitation: we typically emphasize what Newton did, and how he did it, rather than focusing on gravity itself. According to the second, a particular *object* is emphasized. Think of penicillin; we talk about what it is and why it is important. We should not claim that only one of these is the "true" sense of discovery, that would be to introduce a false dichotomy. Inquiry into both processes and products can illuminate the phenomenon of scientific discovery. Nevertheless, I will focus on discovery as an event for three reasons.

First, objects of discovery are not counted as discoveries until there is some recognition of those objects *being* discoveries, and this recognition takes place at a certain time. It makes little sense to say that penicillin was a discovery in 30,000 BCE, or that it was always a discovery. The number 0, democracy and Snell's law do not exist in time, although they seem to have been discovered at certain times. The temporal element of objectual discoveries suggests that we might take discovery events as conceptually primary to discovery objects, since events can be indexed to times and this is not true of all objects of discovery.

Second, the temptation to think of discovery as objectual 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 & Shapin, 1985). Discoveries are made by science itself, that is, by no one in particular, in order to distance those discoveries from the doubts that might otherwise attend them if they were 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 (or could be) discovered and ask if machines could produce them, than to ask what sort of action a discovery is and whether machines could be the sort of agent to perform it.

So, what kind of action is discovery?

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

4.2 Elements of the Discovery Event

I propose we distinguish the following elements of a discovery event: (a) an agent (who discovers), (b) an object of discovery (that which is discovered), (c) a trigger event (that which prompts the discovery) and (d) 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, an entity (concrete or abstract), 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 categorisation 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 Fleming's walking into the lab and seeing the area surrounding the mould in which there was no *Staphylococcus*. All of these objects and events 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 therefore the key *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. Novelty 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 *rediscover* it. Something may be a personal discovery for an agent though only a rediscovery for her 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. Lastly, we should note that an interpretation can be more or less novel.

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 in a scientific domain, while practical problems obfuscate our solving such problems. The distinction is contextual, but it will do for our purposes.

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; Hudson, 2001). The kinds of solution that count as successful for a given problem will vary according to context.

Fourth, in addition to producing a novel and successful solution to a theoretical scientific problem, we require that the problem solved be significant. This is to preserve the intuition that discoveries are in some sense special: not all 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 adequacy, coherence with previous knowledge, fruitfulness, beauty and simplicity.

To summarise, scientific discovery events consist of an agent’s novel interpretation of an object which successfully solves a significant theoretical scientific problem. A discovery is more or less momentous depending on how novel and complete the solution is, and how significant the problem solved is.

Can machines discover in this sense? Trigger events and objects of discovery can be almost anything, and tests for a solution’s satisfactoriness can be programmed into computers in advance, so I will leave these to one side and focus on the question of whether machines can produce novel interpretations that provide solutions to significant theoretical problems. According to some characterisations 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. Second, machines can interpret if by interpretation we mean categorisation. In this minimal sense of interpretation, computers already categorise certain states as solutions by checking them against programmed desiderata. Finally, machines can have significance encoded into them, insofar as they are designed by scientists to address problems that are antecedently deemed significant. 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 senses 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.

³He 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, p. 51).

Good candidates can be found by considering qualitative discoveries like those made in the social sciences.

4.3 Social Scientific Discovery

A broad range of methods are used by social scientists, from surveys and statistics to interviews and observations. Naively, we can draw a continuum from positivist to interpretivist social science methods. Positivists focus on observables such as physical movements and questionnaire responses, and explicitly endorse the more “objective” methods of science, including standardised surveys and statistical analyses. They aim to draw generalisable lessons concerning human behaviour. Interpretivists argue that positivist methods cannot capture the rich complexity of human social life. To understand this complexity, researchers must be deeply immersed in the target system and recognise that they can only uncover a limited amount from a limited perspective. In light of this, interpretivists claim that their discoveries will not be widely generalisable.

Insofar as positivist social science employs quantitative methods of science, machines can make positivist social scientific discoveries if they can make quantitative discoveries in other fields of science, and I assume they can. A harder question is whether machines can discover in interpretivist social science. That is, can machines produce new interpretations of the shared experiences and meanings 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 teach.

It would be impossible to review the notion of discovery across all interpretivist social sciences, so in what follows I focus on ethnography, which I take to be a paradigm interpretivist social science method 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 observations 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 the objects and events that surround them. 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” (Kanter, 1977, p. 98). The skills developed by the secretaries studied by Kanter were highly specific to the needs of their particular bosses. While such specialisation might have provided job security, it also ensured that bosses typically would not let their secretaries move into other (higher) positions.

Another example is Annette Lareau's *Unequal Childhoods* (2003), which focused on how differences in social class affected parenting styles among 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 organised 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 organised structure on daily life. As a result, children are encouraged 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* interpretations that solve *significant* problems?

4.3.1 *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, Snow, Anderson, & Lofland, 2006, p. 173; see also Snow, Morrill, & Anderson, 2003, p. 186). The second, 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" (Lofland et al., 2006, p. 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 in general this is something machines have been doing more and more effectively, especially in mathematics (Lenat, 1977), physics (Langley, 1981), chemistry (Żytkow & Simon, 1986) and biology (Kulkarni & Simon, 1990). We can expect this progress to continue, and it is only a matter of time before machines can apply existing models of human behaviour to new sociological data.

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, p. 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 analyse data to create or refine models, which in this sense of theory, satisfy the requirement (for examples of programs that are capable of such novelty, see Valdés-Pérez, 1995, and Kocabas & Langley, 1998).

These are not 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 analyse a new social phenomenon (e.g. Twitter behaviour). 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 theoretical extension. What is interesting about the second and third is the production of a new interpretation. To give an account of the novelty relevant for social scientific discovery, therefore, we need to look at the nature of ethnographic interpretation. But first, a quick look at ethnographic significance.

4.3.2 *Ethnographic Significance*

According to Lofland et al. (2006, pp. 177–181), a significant ethnographic solution should do at least some of the following: (a) go against “common sense” or the “modern mind-set,” (b) develop ideas that “establish broader implications,” (c) 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 (d) refine or extend existing social science ideas. Let us address these in turn with machine discovery in mind.

First, because computers are not typically programmed to reason as humans do, they generally go against common sense and the modern mind-set. We might nevertheless worry that they have their own computer common sense: patterns of reasoning and expression that they cannot deviate from. Machine novelty could then be thought of as the power to break free from such reasoning styles. This is something machines can currently 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 flexibility in problem solving.

The second way of achieving significance, namely, establishing broader implications, is an instance of theory extension, which we granted was within the purview of machines.

The third, which Lofland et al. (2006) call “developed treatment,” merely requires that work be done “well”: 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. However, there is no (unique, finite, exhaustive) list of methods that are the “good” or “scientific” ones. The only list we have is 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 sense of significance, namely, to refine an existing theory, is equivalent to theoretical refinement, which we granted above is achievable for machines.

In sum, to recognise 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 actions, 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 and problems respectively, and they are attributed *by interpretation*. If we want to uncover the cognitive requirements of producing a novel, significant interpretation, therefore, we will be off to a good start if we can identify the requirements for interpretation in general.

4.3.3 *Ethnographic Interpretation*

We can identify three main interpretive methodologies: analytic induction, grounded theory and the extended case method.⁴

To pursue analytic induction, 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 a novel solution to a significant theoretical problem) 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 focus more generally 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" (Glaser & Strauss, 1967,

⁴For 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 (1991, 1998, 2000). In what follows, I try to distil the methods of ethnographic interpretation, but I cannot do them complete justice. Interested readers are encouraged to look at the sources listed for more details.

p. 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 & Corbin, 1994, p. 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 categorisations 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 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, pp. 17–18). Things not relevant to the background theory and initial set of questions can and should be ignored.

On a loose reading, these 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 in the senses of theory extension and refinement respectively. What we really want are cases where new theoretical understanding is born 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, pp. 198–199).

This kind of interpretation “occurs continuously throughout the life of any qualitatively oriented project” (Miles & Huberman, 1994, p. 10). It begins with coding the data, which is “the process of defining what the data are all about” (Charmaz, 2001, p. 340) or “relating (those) data to our ideas about them” (Coffey & Atkinson, 1996, pp. 45–47) by “sorting your data into various categories that organise it and render it meaningful from the vantage point of one or more frameworks or sets of ideas” (Lofland et al., 2006, p. 200). The codes themselves are “names or symbols used to stand for a group of similar terms, ideas, or phenomena” (LeCompte & Schensul, 1999, p. 55), “tags of labels for assigning units of meaning to information compiled” (Miles & Huberman, 1994, p. 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, p. 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, pp. 122–132), for example, a practice

(like getting ready for work), an episode (like divorce), an encounter (like a cocktail party), an organisation (like a school), or a larger community (like a refugee camp). An aspect of a unit might be the beliefs, norms, ideologies, emotions, relations, etc., of the people in the unit. These combine to form a topic (e.g. the faith of people 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, p. 347). This is where we generate and develop possible explanatory relationships between data (organised in codes) and the topic. 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, p. 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 performed 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, p. 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 organise the codes into answers about a topic. But the thought might be that a machine cannot do this well. Interpretation can be a simple act of rule-governed categorisation, 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 are missing from machines. In the next section, I will try to identify some of the cognitive powers that make the most difficult acts of interpretation possible.

4.4 Machine Interpretation

Building on the work of Peter Winch (1958) and Charles Taylor (1971) I will argue that there are at least five abilities any agent must possess in order to interpret well. I leave out abilities like collecting data and performing calculations, which I take machines already to possess.

To begin with, explaining human social behaviour requires that we discern the meanings of utterances and actions. And this requires that we recognise the possibility of certain behaviours having meanings at all. A statement is not just the

production of a sound wave, it is also the expression of a thought. 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 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 recognise 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 indeterminacy of translation or interpretation here: 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, perhaps by means of a principle of charity and some informed guesses (Stuart, 2015). This is especially difficult where metaphors, loose speaking, body language or implicature 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 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 something has a given meaning.

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, p. 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 reasons for action are.

Sometimes we can discover someone’s intentions simply by asking. But to interpret the answers we receive again requires knowledge of intention, because we must know whether our subject intends to be deceptive before we can consider taking their answers at face-value. In other words, to uncover someone’s intentions by asking, we must already be able to interpret intentions. A second difficulty is that we cannot determine what a subject intends based on observation alone. Contributions of irrationality and luck must be recognised, otherwise, we interpret

a gambling addict as intending to lose money, and people acting under cognitive biases in general as intending to ignore pertinent evidence or deceive themselves.

Finally,

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 (a) distinguish presentation from meaning, (b) identify meaning, (c) identify the “owner” of a meaning, (d) identify reasons for behaviour (while leaving room for irrationality and luck) and (e) 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. Specifically, I want to argue that each ability requires at least the faculty of 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 sensory experience (see Stuart [forthcoming](#)). Let us go in order.

1. To distinguish between presentation and meaning, an agent must recognise that there are always several possible meanings we could attribute to any given presentation (and vice versa). Such recognition requires looking beyond the presentation, in other words, we must conjure and consider states of affairs not currently present to experience. To distinguish between the statement “I’m fine” and the actual meaning someone intends with that statement, we must be able to imagine that the person could mean different things by that statement.
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. Sometimes this is a straightforward practice that could be made algorithmic. But at some point we hit bedrock, and to break through we require a special sort of experience and acquaintance. Consider an emotion term like “shame”. This

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, p. 13)

In other words, many basic pieces of the human semantic puzzle can only be grasped by taking part in common actions, values and experiences. Such participation is ultimately the source of many of our own meanings (Winch, 1958, pp. 81ff), though as Taylor mentions, this participation can also take place in imagination. For example, I possess many important concepts that I could not have gained through actual participation in the home-world of those concepts, because those worlds are fictional or in the past. It is therefore only through exercises of imagination that some instances of semantic understanding can be had, and this will be especially true for machines that cannot (yet) experience many of the things humans do. In any case, even when we have all the relevant experience, we still need to be able to come up with reasonable guesses about what someone means, and find ways to test those hypotheses (Stuart, 2015). And this requires imagination because to test hypotheses we must imagine different experimental setups (in non-trivial cases) and decide which would be more effective for testing by reasoning through possible outcomes of these tests. In other words, we must reason through states of affairs that do not (yet) exist.

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. Taking up a different perspective requires a cognitive departure from our present experience of objects and events, and this requires imagination.
4. The only way to identify someone's intentions (without being told what they are) is to imagine that you have the personal and contextual properties of the agent, and then ask yourself what reasons you would have for acting if you were them. In other words, the ability to interpret others depends both on the ability to interpret yourself (Jackman, 2003) – which requires seeing that your own mental actions have more than one possible meaning – *and* the ability to convert yourself mentally into an approximation of someone else. Both of these abilities were discussed above, and both require imagination.

Finally,

5. Tracking the differences between one worldview and another and establishing semantic links that would enable translation between them requires experience of both worldviews. However, we can only occupy one substantial worldview at a time (otherwise we would have to attribute conflicting properties to the same object). So, to determine and 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 the same objects and events to ourselves from different perspectives, which is to cognise objects and states of affairs otherwise than they are given to us.

Imagination is what enables us to recognise that there are several options for what someone might mean, hypothesise a number of plausible candidates, choose ways to test those candidates, and participate in the otherwise inaccessible action-worlds of others and thereby gain new concepts. It helps us put ourselves in another's position or worldview by seeing ourselves acting under different constraints with some of our existing properties strengthened, and others diminished or removed. Each of the five abilities needed to interpret in difficult cases requires cognitive interaction with objects and states of affairs not currently present to sensory experience. Imagination is therefore a fundamental capacity underlying ethnographic interpretation.

4.5 Conclusion

I have discussed the possibility of ethnographic machine discovery and I have argued that interpreting human behaviour and natural language systems of meaning requires imagination. A fortiori, imagination is necessary for producing some of the novel interpretations that solve significant problems in social science. Therefore, some discoveries in social science are only possible if the discoverer possesses imagination. This implies that for machines to be able to discover in social science to the same extent that humans can, 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 sensory "experience"? Under some interpretations, 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, e.g., something fictional, since in the case of the logic program, the machine is interacting only with symbols that *are* present to its "experience". Perhaps we could reserve (sense) "experience" for machines that have more human forms of data input, like sight and hearing, but this seems unduly restrictive since most machine discovery systems do not currently have such input.

Concerning more substantial senses of imagination, like those required for perspective shifting and empathy, things are even murkier. I conclude therefore on what I think is a surprising note. The necessary conditions for scientific discovery (conceived of as an action) include providing novel interpretations that solve significant theoretical scientific problems, and in order to say whether machines can produce such interpretations we first need a better understanding of the imagination and what cognitive powers are required for its operation. Unfortunately, philosophers and cognitive scientists are still very far from possessing such an understanding.

To conclude, those who feel skeptical *or optimistic* about the extent to which machines can discover in science might profitably focus that skepticism (or optimism) on the nature and possibility of imagination algorithms. And to do this, we require a better understanding of the nature and cognitive requirements of imagination in humans, and imagination in general.

Acknowledgments Thanks to the organisers and participants of the conference “Scientific Discovery in the Social Sciences” at the London School of Economics, as well as Nancy Nersessian, Marco Buzzoni, Markus Kneer, Maël Pégnny 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. This work was funded by a postdoctoral fellowship from the University of Pittsburgh, and a postdoctoral fellowship from the Social Sciences and Humanities Research Council of Canada.

References

- Achinstein, P. (2001). *The book of evidence*. Oxford, UK: Oxford University Press.
- Boden, M. A. (2004). *The creative mind: Myths and mechanisms*. London: Routledge.
- Bradshaw, G. L., Langley, P., & Simon, H. A. (1980). BACON4: The discovery of intrinsic properties. In *Proceedings of the third national conference of the Canadian Society for Computational Studies of Intelligence*, pp. 19–25.
- Bradshaw, G. L., Langley, P., & Simon, H. A. (1983). Studying scientific discovery by computer simulation. *Science*, 222, 971–975.
- Burawoy, M. (1991). *Ethnography unbound*. Berkeley, CA: University of California Press.
- Burawoy, M. (1998). The extended case method. *Sociological Theory*, 16, 4–33.
- Burawoy, M. (2000). *Global ethnography*. Berkeley, CA: University of California Press.
- Charmaz, K. (2001). Grounded theory. In R. M. Emerson (Ed.), *Contemporary field research: Perspectives and formulations* (2nd ed., pp. 335–252). Prospect Heights, IL: Waveland Press.
- Coffey, A., & Atkinson, P. (1996). *Making sense of qualitative data: Complementary research strategies*. Thousand Oaks, CA: Sage.
- 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 S. Dzeroski & L. Todorovski (Eds.), *Computational discovery of communicable scientific knowledge*. Berlin, Germany: Springer.
- Gillies, D. (1996). *Artificial intelligence and scientific method*. New York: Oxford University Press.
- Giza, P. (2002). Automated discovery systems and scientific realism. *Minds and Machines*, 22, 105–117.
- Glaser, B. (1978). *Theoretical sensitivity*. Mill Valley, CA: Sociological Press.
- Glaser, B., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.
- Hudson, R. G. (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*. New York: Basic Books.
- Katz, J. (2001). Analytic induction. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences* (pp. 480–484). Amsterdam, The Netherlands: Elsevier.
- Kocabas, S., & Langley, P. (1998). Generating process explanations in nuclear astrophysics. In *Proceedings of the ECAI-98 Workshop on Machine Discovery* (pp. 4–9). Brighton, UK.
- Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.
- Kulkarni, D., & Simon, H. A. (1990). Experimentation in machine discovery. In J. Shrager & P. Langley (Eds.), *Computational models of scientific discovery and theory formation*. San Mateo, CA: Morgan Kaufmann.

- Lane, P., Sozou, P., Addis M., & Gobet, F. (2014). Evolving process-based models from psychological data using genetic programming. In R. Kibble (Ed.), *Proceedings of the 50th Anniversary 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. *International Journal of Human-Computer Studies*, 53, 393–410.
- Langley, P., & Jones, R. (1988). A computational model of scientific insight. In R. Sternberg (Ed.), *The nature of creativity*. Cambridge, MA: Cambridge University Press.
- Langley, P., Shrager, J., & Saito, K. (2002). Computational discovery of communicable scientific knowledge. In L. Magnani, N. J. Nersessian, & C. Pizzi (Eds.), *Logical and computational aspects of model-based reasoning*. Dordrecht, The Netherlands: Kluwer Academic.
- Langley, P., Simon, H. A., Bradshaw, G. L., & Żytkow, J. M. (1987). *Scientific discovery: Computational explorations of the creative processes*. Cambridge, MA: MIT Press.
- Lareau, A. (2003). *Unequal childhoods: Class, race, and family life*. Oakland, CA: University of California Press.
- LeCompte, M. D., & Schensul, J. J. (1999). *Journeys through ethnography: Realistic accounts of fieldwork*. Boulder, CO: Westview Press.
- Lenat, D. B. (1977). Automated theory formation in mathematics. In *Proceedings of the fifth international joint conference on artificial intelligence* (pp. 833–842). Cambridge, MA: Morgan Kaufmann.
- Lenat, D. B. (1982). AM: Discovery in mathematics as heuristic search. In R. Davis & D. Lenat (Eds.), *Knowledge-based systems in artificial intelligence*. New York, NY: McGraw-Hill.
- Lindesmith, A. (1947). *Opiate addiction*. Bloomington, IN: Principia Press.
- Lofland, J., Snow, D., Anderson, L., & Lofland, L. H. (2006). *Analyzing social settings* (4th ed.). Belmont, CA: Wadsworth.
- McArthur, D. J. (2011). Discovery, theory change and structural realism. *Synthese*, 179, 361–376.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks: CA: Sage.
- Newell, A., Shaw, J. C., & Simon, H. A. (1958). Elements of a theory of human problem solving. *Psychological Review*, 65, 151–166.
- Poincaré, H. (1914). *Science and method*. London: Thomas Nelson and Sons.
- Schaffer, S., & Shapin, S. (1985). *Leviathan and the air-pump*. Princeton, NJ: Princeton University Press.
- Schickore, J. (2014). *Scientific discovery*. *Stanford Online Encyclopedia of Philosophy*. Available at <http://plato.stanford.edu/entries/scientific-discovery>. Accessed 15 Nov 2016.
- Schickore, J., & Steinle, F. (Eds.). (2006). *Revisiting discovery and justification: Historical and philosophical perspectives on the context distinction*. Dordrecht, The Netherlands: Springer.
- Schindler, S. (2015). Scientific discovery: That-whats and what-thats. *Ergo*, 2, 123–148.
- Shrager, J., & Langley, P. (Eds.). (1990). *Computational models of scientific discovery and theory formation*. San Mateo, CA: Morgan Kaufmann.
- Simon, H. A. (1977). *Models of discovery*. Dordrecht, The Netherlands: Reidel.
- Simon, H. A. (1979). *Models of thought*. New Haven, CT: Yale University Press.
- Snow, D., Morrill, C., & Anderson, L. (2003). Elaborating analytic ethnography linking fieldwork and theory. *Ethnography*, 4, 181–200.
- Spangler, S., Myers, J. N., Stanoi, I., Kato, L., Lelescu, A., Labrie, J. J., ... Comer, A. (2014). *Automated hypothesis generation based on mining scientific literature*. Association for Computing Machinery's Digital Library.
- Strauss, A. (1987). *Qualitative analysis for social scientists*. New York: Cambridge University Press.
- Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: Sage.
- Strauss, A., & Corbin, J. (1994). *Grounded theory methodology: An overview*. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 273–285). Thousand Oaks, CA, USA: Sage Publications, Inc.

- Stuart, M. T. (2015). Philosophical conceptual analysis as an experimental method. In T. Gamerschlag, D. Gerland, R. Osswald, & W. Petersen (Eds.), *Meaning, frames and conceptual representation* (pp. 267–292). Düsseldorf, Germany: Düsseldorf University Press.
- Stuart, M. T. (Forthcoming). Towards a dual process epistemology of imagination. *Synthese*. <https://doi.org/10.1007/s11229-019-02116-w>.
- Taylor, R. (1971). Interpretation and the sciences of man. *The Review of Metaphysics*, 25, 3–51.
- Valdés-Pérez, R. E. (1995). Machine discovery in chemistry: New results. *Artificial Intelligence*, 74, 191–201.
- Winch, P. (1958). *The idea of a social science and its relation to philosophy*. London: Routledge.
- Znaniecki, F. (1934). *The method of sociology*. New York: Farrar and Rinehart.
- Żytkow, J. M., & Simon, H. A. (1986). A theory of historical discovery: The construction of componential models. *Machine Learning*, 1, 107–137.