




# Ontology, neural networks, and the social sciences

David Strohmaier<sup>1</sup> 

Received: 2 September 2020 / Accepted: 14 December 2020 / Published online: 28 December 2020  
© The Author(s) 2020

## Abstract

The ontology of social objects and facts remains a field of continued controversy. This situation complicates the life of social scientists who seek to make predictive models of social phenomena. For the purposes of modelling a social phenomenon, we would like to avoid having to make any controversial ontological commitments. The overwhelming majority of models in the social sciences, including statistical models, are built upon ontological assumptions that can be questioned. Recently, however, artificial neural networks (ANNs) have made their way into the social sciences, raising the question whether they can avoid controversial ontological assumptions. ANNs are largely distinguished from other statistical and machine learning techniques by being a representation-learning technique. That is, researchers can let the neural networks select which features of the data to use for internal representation instead of imposing their preconceptions. On this basis, I argue that neural networks can avoid ontological assumptions to a greater degree than common statistical models in the social sciences. I then go on, however, to establish that ANNs are not ontologically innocent either. The use of ANNs in the social sciences introduces ontological assumptions typically in at least two ways, via the input and via the architecture.

**Keywords** Neural networks · Philosophy of social science · Prediction · Statistical models · Ontological assumptions

## 1 Introduction

The field of social ontology remains riddled with controversies. It is a matter of persistent disagreement which social objects exist and which ontological dependence

---

✉ David Strohmaier  
davidstrohmaier92@gmail.com; ds858@cam.ac.uk

<sup>1</sup> Department of Computer Science and Technology, University of Cambridge, 15 JJ Thomson Avenue, Cambridge CB5 8DT, UK

relations<sup>1</sup> hold between them. To get a taste of these ontological controversies, consider the following questions:

- Are social facts (partially) grounded in facts about non-human objects? (Epstein 2015)
- What is the nature of social groups? (Uzquiano 2004; Sheehy 2006; Ritchie 2013, 2015, 2020; Epstein 2015; Thomasson 2019; Hawley 2017; Strohmaier 2018; Uzquiano 2018; Epstein 2019)
- Do group agents exist? (List and Pettit 2011; Huebner 2014; Tollefsen 2015; Epstein 2019; Strohmaier 2020)

These are fundamental and substantial questions that also affect more narrow issues. Consider the nature of families as a type of social group. There is no straightforward agreement on who composes or should compose a family (cf. Satz 2017; Kane 2019). Can there be families with more than two parents per child? Is recognition according to social norms or laws required to make a group of people a family? The fundamental disagreements about the nature of groups also render these more specific ontological issues controversial.

Given the controversial nature of social ontology, there is a demand for ontologically neutral approaches to predictive models. Models that allow social scientists to make predictions without committing to any controversial ontological assumptions, would be of great value, especially if their results could then be compared to models with different ontological commitments. The more ontological assumptions a model can avoid, the better it meets this demand.

Common statistical and machine learning methods, however, require social scientists to specify features, which often come loaded with ontological assumptions.<sup>2</sup> For example, to model the impact of family structure on a child's educational achievement, the structure needs to be encoded using selected features such as how many parents are present. This feature selection is bound into ontological controversies.

Increasingly, however, artificial neural networks (ANNs) have made their way into the social sciences, raising the question of whether they fare any better. In contrast to other statistical and machine learning methods, an ANN does not require feature selection. Instead of imposing our controversial metaphysics on the social, we seem to leave ontology almost completely to the empirical data. Or, as two computational social scientists used to working with the explicit ontologies of agent-based models have put it:

Neural networks have the absolute minimum in the way of ontological structure it is possible to have. Their 'content' comes from the data they are trained to fit. (Polhill and Salt 2017, p. 144)

I will explore whether neural networks live up to these high hopes of ontological neutrality. While I will conclude that ANNs allow social scientists to avoid some

<sup>1</sup> I write here and later generically of ontological dependence relations, which I take to include grounding, constitution, composition, and anchoring if such relations exists. I do not include causation.

<sup>2</sup> Agent-based and game theoretical models do even worse insofar they require an explicit ontology. In the following these types of modelling will not be considered.

ontological assumptions, they only do so partially. To establish this thesis, I will first introduce the functioning of ANNs and their use in the social sciences. After this setup, I will compare neural models to other statistical and machine learning models and argue that neural networks achieve greater ontological neutrality. I will then, however, consider two ways in which neural networks are not ontologically innocent. The input and the architecture of neural networks provide openings for ontological assumptions. Before concluding, I will sketch how mistaken ontological assumptions can lead to multiple problems.

## 2 Neural networks

In the past decade, artificial neural networks have made impressive progress. While some of the most prominent examples of this progress—such as winning in go against a professional human player (Silver et al. 2016)—have mainly been of show-value, neural networks have also become a valuable tool for scientific investigations (e.g. Schmidt et al. 2019; Guest et al. 2018; Lakhani and Sundaram 2017). The social sciences have been no exception.

To see how neural networks figure in the work of social scientists, we need to understand how they work. Accordingly, I will begin by sketching the history and functioning of the most common types of neural networks. Then, I will discuss the current use of ANNs in the social sciences. On this basis we will be able to assess the role of ontological assumptions for neural networks.

### 2.1 History and functioning

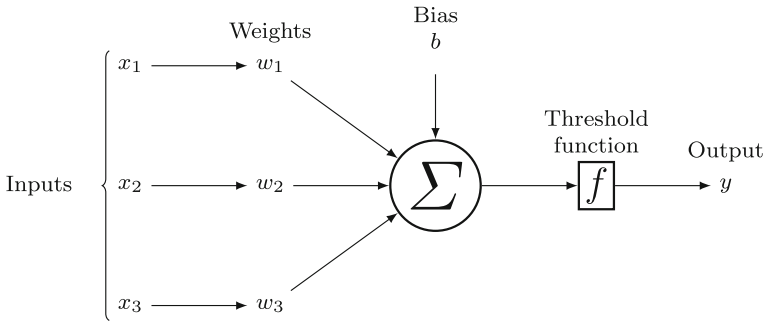
ANNs are an old technology with deep roots in the history of artificial intelligence research. Drawing upon the description of neuronal activity by McCulloch and Pitts (1943), Rosenblatt (1958) popularised the perceptron as an early form of neural artificial intelligence. A single perceptron can be understood as a neural unit which takes multiple inputs, multiplies them with weights, adds a bias, and feeds the result through a threshold function, the output of which is the prediction (see Fig. 1). Thus, the computation by a single unit with three input values can be represented as

$$Y = f(w_1x_1 + w_2x_2 + w_3x_3 + b) \quad (1)$$

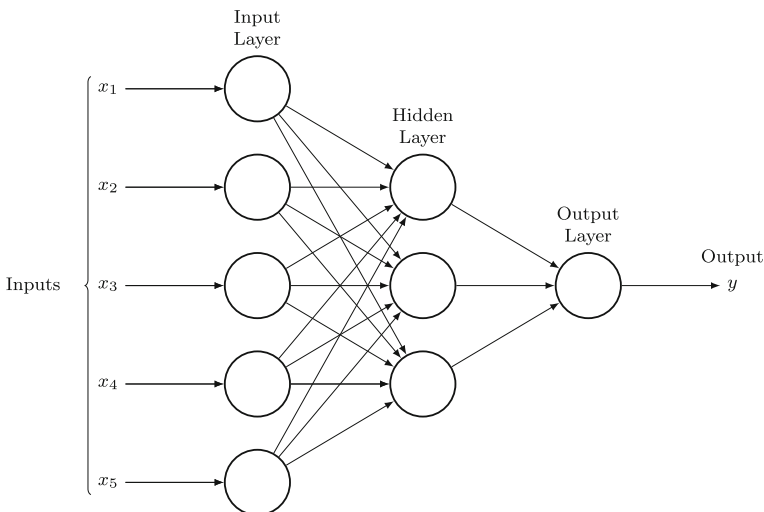
where  $x_1, x_2, x_3$  are the input values,  $w_1, w_2, w_3$  their respective weights,  $b$  a bias term, and  $f$  a threshold function such as the step function. In the neural unit of a contemporary ANN, the threshold function is replaced by a non-linear activation function such as the sigmoid function.<sup>3</sup>

After influential and critical work by Minsky and Papert (1972), interest in neural networks declined for a period only to rise and fall again in the 80s and 90s with the connectionism movement in cognitive science (cf. Buckner and Garson 2019). By then, multiple layers of units were connected to each other (see Fig. 2). These nodes

<sup>3</sup> The sigmoid function is defined as  $\sigma(x) = \frac{1}{1+e^{-x}}$ .



**Fig. 1** Structure of a perceptron, similar to a unit in a neural network. The main difference is that the threshold function is typically replaced by a different activation function



**Fig. 2** Structure of a fully connected feed-forward network with five input values. Each node represents a neural unit, including the activation function. Current networks are rarely fully connected and contain additional computational mechanisms

receive either the original input or the output of the previous layers, perform a linear operation on them, and apply a non-linear activation function to them.

During the era of connectionism, ANNs were shown to be universal function-approximators. Theoretical results have established that ANNs can in principle approximate continuous functions with very few restrictions (see Cybenko 1989; Hornik et al. 1989) and that they can simulate all Turing machines (Siegelmann and Sontag 1992). One should keep in mind, however, that these results do not show that it is feasible to *train* a neural network to approximate any function or Turing machine (cf. Goodfellow et al. 2016, pp. 192–193). That is a much harder task.

One of the most important innovations in training ANNs is the backpropagation-algorithm (Rumelhart et al. 1986), which allows researchers to train multi-layer neural networks. Training a neural network is the single greatest hurdle on the way to deploy-

ment and remains so despite the development of ready-made toolkits. The training process can be varied in many ways and I will describe the mere outlines. In the most common procedure, labelled training data is fed into the neural network and then used to calculate a loss function, that is a measure of error. By way of backpropagation, this error is used to adapt the parameters of the neural layers, most importantly their weights.

Advances in engineering of the training procedure and the development of new architectures enabled the recent wave of ANNs under the name “deep learning”, which refers to the multiplicity of layers. How the layers of an ANN are connected and how other computations, such as so-called attention-mechanisms (Bahdanau et al. 2014) and softmax-functions, are applied, determine its architecture.<sup>4</sup> Different architectures have proved valuable for different purposes. For example, the convolutional neural network architecture might be more appropriate for image recognition than for document classification.

## 2.2 The use of neural networks in the social sciences

The use of neural networks in the social sciences goes back decades (Bainbridge 1995; Garson 1998; Herbrich et al. 1999), but only with the latest wave have computational power, software, and data collection reached a point at which training neural networks on massive social datasets has become a feasible and tempting endeavour. They have become a valuable method in the toolbox for data mining in the social sciences (e.g. Attewell and Monaghan 2015).

ANNs have shown promise for predicting the default risk of countries (Cooper 1999) and have been used to study international conflict (Beck et al. 2000; de Marchi et al. 2004), poverty (Jean et al. 2016), and public corruption (López-Iturriaga and Sanz 2018). In the *Fragile Families Challenge*, researchers competed to predict variables such as GPA, grit, and material hardship using data collected as part of the *Fragile Families & Child Wellbeing Study* (Waldfoegel et al. 2010).<sup>5</sup> Neural network models were one type of model used in this challenge (Davidson 2019), although the performance of machine learning models was overall disappointing (cf. Salganik et al. 2020).

The aim in these applications is to predict a variable of direct interest to social scientists, usually on the assumption that the available data are especially suited for this methodology. In addition, there have been many adjacent uses of ANNs with bearing on the social sciences, such as enriching datasets with demographic information based on facial recognition (Mancosu and Bobba 2019), classifying messages on Twitter (Gambäck and Sikdar 2017; Liao et al. 2019), and recognising collective actions in image sequences (Bagautdinov et al. 2017). Such applications of neural networks can form part of a larger investigation employing more traditional types of models.

<sup>4</sup> For a taste of the variations, see <https://www.asimovinstitute.org/neural-network-zoo/>.

<sup>5</sup> For more information about the challenge, see <https://www.fragilefamilieschallenge.org/>, the results in Salganik et al. (2020), and the dedicated special collection of the journal *Socius*: [https://journals.sagepub.com/topic/collections-srd/srd-1-fragile\\_families/srd](https://journals.sagepub.com/topic/collections-srd/srd-1-fragile_families/srd).

Although the use of neural networks in the social sciences is growing, it has been mostly confined to subfields of economics (Li and Ma 2010; Falat and Pancikova 2015) and minor exceptions in other fields. A major reason for this reluctance of social scientists towards using neural networks is the difficulty of interpreting them. While this issue is not the main focus of my investigation, it has such influence that a few comments are required to understand the current use of ANNs in the social sciences.

Social scientists largely consider ANNs to have a black-box character (but see Lipton (2018) on this issue).<sup>6</sup> A standard ANN might classify examples by providing a probability distribution over possible classes, but it will not provide explicit reasons for doing so. The weights and biases of the neural network have no direct interpretation.

For many social scientists, it would be at best disappointing to be able to predict the outcome of a social situation but unable to provide any reasons for the prediction. Practitioners in the computational social sciences have responded by stressing the value of prediction and connecting it to more classical goals of the social sciences, such as identifying causal connections (Hofman et al. 2017; Watts et al. 2018).<sup>7</sup>

Independently of the value of prediction, however, there has been much work into making ANNs interpretable, often motivated by ethical concerns such as the need to detect unwanted social biases.<sup>8</sup> For example, some technologies can indicate which part of the original data was especially important in reaching the classification (Ribeiro et al. 2016). These techniques can also be applied to making neural networks interpretable for the purposes of the social sciences, rendering them more similar to common statistical methods (see the discussion in Davidson 2019).

While trained neural networks are initially black boxes to social scientists, additional tools and effort can shed light on their internal workings. Given that they perform well, which is not yet sufficiently shown (see the disappointing results in the *Fragile Families Challenge*), this effort might be justified. If one model predicts social developments better than any other, then the best way to advance the social sciences might lie in investigating how the model achieves such a performance, rather than ignoring them and working on more limited approaches. A sufficiently powerful predictor of X is part of the research domain for those studying the regularities of X.

A potential reason for social scientists to take neural networks seriously despite the challenges of interpretability is to avoid ontological assumptions. In the next section, I will suggest that ANNs in fact have such an advantage compared to other models.

### 3 Comparing neural networks to other statistical and machine learning methods

Neural networks are both statistical and machine learning models. A comparison with other methods in these fields, however, reveals differences suggesting greater

<sup>6</sup> For one debate about this issue in the social sciences, see Beck et al. (2000) and de Marchi et al. (2004).

<sup>7</sup> For a discussion of similar issues in cognitive science, see Cichy and Kaiser (2019).

<sup>8</sup> An article in *Pro Publica* by Angwin and Lawson (2016) has particularly raised interest in this issue. For a rejoinder to this article, see Flores et al. (2016). For some general treatments of algorithms and fairness see Corbett-Davies et al. (2017), Obermeyer and Mullainathan (2019) and Glymour and Herington (2019).

ontological neutrality on the side of neural networks. I will discuss both comparisons in order.

### 3.1 Comparison to other statistical models

Neural networks, at least in their most common forms, are statistical models (cf. Lee 2004, p. 21). That being said, ANNs differ greatly from the standard statistical methods employed in the social sciences in the assumptions they have to make.

For a simple example of standard models, consider a linear regression investigating whether the variable of favourability in Gallup polls is predictive of success in the US Presidential election (e.g. Lewis-Beck and Rice 1982). The form of such a linear regression resembles that a single neural unit, except for the lack of an activation function. For the case where we only estimate the presidential vote share based on the Gallup favourability rating on the day of the election, the equation can be written as:

$$Y = wX + b \quad (2)$$

where  $X$  is the Gallup rating and  $w$  and  $b$  are the parameters to be estimated.

Linear regression and similar standard approaches require explicit choices by the modellers about the relationships between the features of the input. In a linear regression, the relevant variables between which a correlation is suspected need to be specified. Lewis-Beck and Rice (1982) went at their research with a specific hypothesis in mind and selected the features—in this case only the Gallup rating—to which they fitted the regression line. From all the available data, they select one feature and encode it. As a consequence of the explicit selection of predictive features in the data, constructing such models tends to be guided by controversial ontological assumptions.<sup>9</sup>

Common statistical models can be taken to mediate between the underlying substantive model, which is supposed to represent the actual explanatory factors, and the observed data, which are partially the result of chance patterns. As a results, statistical models can be substantively adequate or inadequate (this understanding is based on Spanos (2006) and Spanos and Mayo (2015)).<sup>10</sup>

In the case of the regression predicting voting behaviour, the underlying substantive model assumes that a favourable impression of the candidates could be an explanatory factor for the presidential vote. In virtue of the connection the substantive model, the weight  $w$  estimated by the model is interpretable as the impact of the feature Gallup favourability rating. The size and direction of this weight is the subject matter of the investigation using linear regression.

In the sketched instance, these assumptions pose little problem, because it is widely accepted that the outcome of the vote is largely the result of the decisions by the indi-

<sup>9</sup> Linear regressions in particular also make statistical assumptions that ANNs can avoid, such as homoscedasticity, i.e. that the random variables have the same variance. For an early sociological work on these issues, see Zeng (1999).

<sup>10</sup> I thank Aris Spanos for clarifying their work to me. Any remaining misunderstandings are due to me.

vidual voters who are supposedly polled. In other instances, however, the underlying assumptions are bound to be more controversial.

For example, a linear regression might be used to evaluate the correlation between the family structure and the educational outcomes of a child. In this case, a need to provide a coding for the different types of family structures arises. Such a coding is likely to be ontologically controversial and requires assumptions about what constitutes a family. Is only the nuclear family included in the structure? What about families in which the primary parents practice polyamory or have separated and entered new relationships?

The creation of such feature codings for a statistical model is dependent on ontological assumptions about families. The prevalent forms of statistical modelling in the social sciences carry an ontological burden, because they rely on decisions about coding the data.<sup>11</sup>

When using an ANN, however, researchers typically do not build upon an underlying substantive model, i.e. they do not parametrically nest such a substantive model by deliberate design. The researchers do not estimate the weight and bias for specific features, but instead the network serves predictive purposes and estimates non-interpretable weights and biases. Consequently, no feature selection is required. While not picking out features for estimation reduces the interpretability of ANNs, it also means that the issue of ontological bias in coding does not arise.

To see the difference between statistical methods such as linear regression and ANNs even more clearly, consider how each of these approaches can make use of open-ended survey questions. Open-ended questions are supposed to allow survey respondents to draw from a broader range of possible responses. To make statistical use of these open-ended questions, however, they are often coded again, which creates challenges (e.g. Behr 2015) and partially undermines the purpose of employing open-ended questions in the first place. By contrast, the open-ended responses can be directly used by ANNs without an intermediate coding step. The open response can be used as a text input into an ANN without the need for a unified coding. Similar to open questions, other collected texts and images, even video evidence can be used more directly without a coding step, as currently common in the social science.

But if the researchers do not specify the features on which to train ANNs, the question arises of how they represent data at all. I address this question in the next subsection by comparing ANNs to other machine learning models.

### 3.2 Comparison to other machine learning models

ANNs are distinguished from many other machine learning methods, such as Support Vector Machines and Decision Trees, by being a representation-learning technique.<sup>12</sup>

<sup>11</sup> But do these assumptions in fact matter as long as the model is predictive? This issue is addressed for the case of ANNs in Sect. 5 and similar considerations apply here.

<sup>12</sup> For an already somewhat dated but nonetheless informative introduction to different representation-learning techniques, see Bengio et al. (2013). Goodfellow et al. (2016, chapter 15) also includes a discussion of representation learning.



Text	Bag-of-Words Vector
“I prefer Shakespeare to Goethe”	[1, 1, 1, 1, 1]
“I prefer Goethe to Shakespeare”	[1, 1, 1, 1, 1]

**Fig. 3** Despite expressing a reversed preference ranking, the bag-of-words representations are identical. Vocabulary: [I, prefer, Shakespeare, to, Goethe]

For most machine learning techniques researchers need to painstakingly identify relevant features of the data to represent it, just as in the case of a linear regression.

To give an example from natural language processing, consider the task of classifying social media messages according to whether they express a positive or negative sentiment.<sup>13</sup> For such classification with a classical machine learning method, one has to define various features. As a simple approach one might simply select the presence of word tokens as features, a so-called bag-of-words representation. In this case, the message is encoded as a vector with as many dimensions as word types in the vocabulary and each dimension includes the count of word tokens belonging to the respective type (see Fig. 3).

But to achieve better results the engineer has to select more sophisticated features. For example, one might want to take into account whether the word “not” appears before other words such as “great”. To achieve this one can use positional encodings as another feature. The selected features together determine a complete vector of a fixed dimensionality which is then put into a Support Vector Machine to classify examples, but only on the basis of pre-created representations.

Even though the goal of machine learning techniques is typically not to estimate weights and biases, they require feature selection like linear regression. Accordingly, the process of hand-crafting such representations is prone to be led by ontological assumptions. In the case of natural language processing, the assumption might be that the meaning of a sentence depends on the syntactic arrangement of words. Such ontological assumptions are bound to be much more controversial in the social sciences.

Assume again that the aim is to predict the impact of a child’s family structure on its educational outcomes. Creating a coding for the structure is dependent on social ontological assumptions. The choice between features such as “divorced parents” and “hours spent with grandparents” offers an opening for ontological assumptions to influence the models. For example, the features used to represent family structures could exclusively be drawn from properties of the nuclear family. Such a selection would taint the encoding of data with a controversial ontological assumption.

By contrast, ANNs are an instance of representation-learning and therefore do not rely on so-called feature engineering. They create their own internal representations guided by the data and the loss function. In the case of natural language processing tasks, neural networks often only receive the string of tokens (or even characters) as input without the need to select any further features. On the basis of a selected vocabulary, these tokens are then mapped to a vector in an initial layer of the neural network, the so-called embedding layer. This approach results in word embeddings,

<sup>13</sup> This task is known as sentiment classification or sentiment analysis (Pang et al. 2002).

dense vectors created in many natural language tasks. Such embedding vectors serve as representations of word meanings for a pre-selected vocabulary.<sup>14</sup>

In the case of the social sciences, ANNs can use trace data, that is raw data that is found rather than created for research purposes (see Howison et al. (2011) for a discussion of digital trace data). Such trace data can include behavioural traces left on social networks or image data. On the basis of such data, neural networks can create representations and predict social phenomena. Hence, there is no need to design an encoding of the family structure to predict the educational outcomes of children, instead an ANN can create such a representation internally for the purpose of the prediction task.<sup>15</sup> All that is required is that the relevant data are either directly available as real-valued vectors or that they can be fitted into a vocabulary so that embedding representations can be trained for them. For text and image data there are standard ways of doing so, which can be readily used in the social sciences.

Having seen that neural networks stand out from other modelling approaches in virtue of being a representation-learning technique, the next section will discuss where ontological assumptions nonetheless affect ANNs.

#### 4 The ontological assumptions of neural networks

Modelling remains a contentious practice in the social sciences. All approaches to modelling social phenomena face criticisms, but the approaches differ in regard to the assumptions they make. As discussed, neural networks are an instance of representation-learning and thereby distinguished from common statistical and other machine learning approaches. In principle, ANNs are trained on extensive data, learn how to represent it internally, and employ these representations for approximating a function. This picture suggests that no or very few ontological assumptions are built into neural networks. To quote again the computational social scientists Polhill and Salt:

Essentially, apart from the labels assigned to the input and output units of a neural network, neural networks don't have an ontology at all. (Polhill and Salt 2017, p. 142)

While it is correct that neural networks don't have much of an explicit ontology compared to agent-based models—the main point of comparison for Polhill and Salt—the impression given by this quote is misleading at best. The choice of labels is not the only point where ontology can make a difference. Typically, neural network approaches

---

<sup>14</sup> In some cases, such as the popular *Word2Vec* implementations (Mikolov et al. 2013) embeddings have been used for many purposes other than the original networks in which these embeddings are trained. For example, these representations can be used as features for other machine learning algorithms to improve their performance.

<sup>15</sup> That is not to say that trace data don't bring their own set of problems. For example, it is difficult to collect representative samples of such data. For the discussion of such issues and the question of how to integrate trace and survey data, see Stier et al. (2020).

in the social sciences include ontological assumptions in at least two ways, via the input and via the architecture.<sup>16</sup>

#### 4.1 Input

Compared to most machine learning and statistical methods, the data fed into neural networks are raw, but the data still need to be selected and put into a format that can be processed by neural networks. This selection and processing offers an opening for ontological assumptions. It would be incoherent to bemoan the way a statistical model codes family relations so as to include only the nuclear family and then feed an ANN exclusively data about the nuclear family.

Of course, avoiding the manual selection of relevant features was supposed to be the advantage of representation-learning techniques such as neural networks. The problem is that an ANN can learn its representations only from the data it is given. One cannot feed the social directly into the neural network; one has to select an input. A linear regression requires the specification of features, but both a linear regression and an ANN require data.

The situation is even worse when ANNs are trained for a prediction task with closed-question survey data.<sup>17</sup> For example, Davidson (2019) participated in the *Fragile Families Challenge* and therefore the predictions of his neural model are based on the pre-selected survey data supplied by the organisers. Such surveys select features for their closed-questions and, thus, reintroduce all the problems of non-representation-learning-based methods. The data encodes interviews with mother, father, child, and “primary caregiver”.<sup>18</sup> This selection already suggests that the features in the survey were guided by ontological assumptions about families and the ANN can only draw on them.

The partial neutrality and power of ANNs are much better served by raw behavioural data, for example, unfiltered text messages sent by members of the family. An ANN might then pick up on parenting style and effort or on any other set of features that are included in the messages. Of course, the selection of such trace data can also be influenced by ontological assumptions.

Assume that an ANN is supposed to predict a child’s educational outcomes based on family interactions. For this purpose, the ANN can be trained on raw text messages between family members. While the use of such behavioural traces mitigates the issue of selection compared to survey data, such traces still need to be collected by someone, often research assistants who bring their own assumptions. For example, an assistant might make sure to collect the trace data for parents and grandparents, but not friends of the parents, because they assume that the family of the child is constituted by

---

<sup>16</sup> Ontological assumptions can also guide the structure of the output, e.g. when the ANN is used to create labelled trees. So far, however, the use of ANNs in the social sciences has been largely restricted to predicting scalar values, e.g. GPA scores. Therefore, I neglect this potential role of ontological assumptions.

<sup>17</sup> That is not to say that there are no interesting applications of ANNs to survey data, e.g. Khan and Kulkarni (2013).

<sup>18</sup> For more information on the data see <https://fragilefamilies.princeton.edu/documentation>.

parents and grandparent. In this case, the ANN would approximate a function based on ontologically problematic data.

As can be seen, ontological commitments can be reduced but hardly avoided in the selection of input. If available, the use of raw trace data reduces the ontological burden of ANNs relative to common statistical methods in the social sciences. But their availability in sufficient quantity is a major challenge and even with trace data the selection can introduce ontological assumptions. In sum, being a representation-learning technique does not remove all openings to ontological assumptions through the input.

## 4.2 Architecture

While neural networks are sometimes sold as a one-size-fits-all solution, they often need to be adapted to the problem at hands. Not only the input needs to be selected, but the architecture of the neural network. I will argue that architectural choices can come with a considerable ontological burden, contrary to the following passage on ANNs by Polhill and Salt:<sup>19</sup>

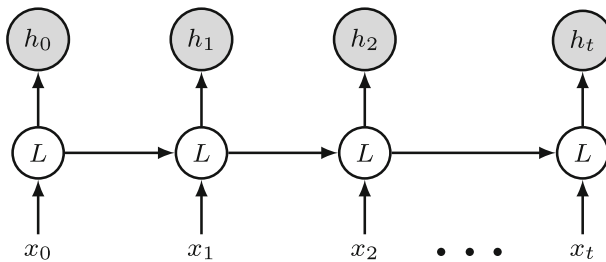
“[A]ssumptions about functional form are embedded in the structure of the network itself—how the nodes are arranged into layer and/or connected to each other. This structure, however, only reflects the flexibility will have to achieve certain combinations of outputs on all the inputs it might be given (its ‘wiggleness’). This is a rather weak ontological commitment to make to a set of data.” (Polhill and Salt 2017, p. 153)

Polhill and Salt suggest that architecture primarily affects how easy it is to make a neural network approximate a function without creating large ontological commitments. Their suggestion is correct for some architectural choices, which are relatively innocent by any measure. For example, Davidson (2019) explored different activation functions for the *Fragile Families Challenge* and found tangible effects on the performance of the models, but it is hard to see how this architectural choice could bear on any ontological controversies. That being said, other choices bring greater commitment.

Some architectures assume dependence patterns. For example, convolutional neural networks (CNNs), which have been widely used for image recognition, including the social scientific work of Jean et al. (2016), are suitable for recognising local patterns (LeCun et al. 1990; see also Buckner 2019 for a philosophical discussion). The LSTM-architecture (Hochreiter and Schmidhuber 1997), by contrast, has enjoyed prominence in natural language processing, because it is better suited for long term dependencies in a string of input, such as text. The original LSTM-architecture comes with the assumption that the dependence is one-directional from left to right (see Fig. 4), although a bidirectional version exists as well.

---

<sup>19</sup> It might seem that I am picking on these two particular authors, but that is not because the book chapter I quote is particularly deficient. The focus of their chapter is on agent-based models and offers valuable discussion of validating the ontology of these models. Their use of ANNs as comparison points, however, has led them to make especially straightforward assertions about the relation between ontology and neural networks.



**Fig. 4** Unfolded Recurrent Neural Network (RNN) encoder, of which the LSTM encoder is a special form. In the case of the LSTM, the  $L$  units are constructed to ensure that the network can maintain a memory of the previous input

As can be seen, the architecture encodes important assumptions about how the input data are related. While the dependence is not necessarily ontological, ontological assumptions are likely to make a difference. In the case of encoding the meaning of words, the dependence assumed by an LSTM-encoder might very well be ontological. Assuming a form of contextualism, the meanings of the encoded word tokens are partially constituted by the meaning of the surrounding tokens. Using a one-directional LSTM for this purpose assumes that the meaning of word tokens *only* depends on previous tokens, not the subsequent ones. This directionality assumption is not just a minor factor of wiggleness, as suggested by Polhill and Salt, but a strict commitment to a direction of dependence.

In the case of the social sciences, one might also choose an architecture because it fits one's assumption about the dependence between input data. In some cases, this might be defensible, for example when one chooses a one-directional LSTM for modelling time-series data from financial markets. Presumably, the earlier data do not depend on the later one. Nonetheless, the choice of the architecture is one place where ontological assumptions can creep in.

Assume that the decision of an organisation is to be encoded using a neural network. For this purpose, indicators of the current decision state of various individuals, groups, and departments might be passed into a one-directional LSTM.<sup>20</sup> The ordering of the input might then make the assumption that the decision state of the marketing department depends on that of the departmental managers which in its turn is dependent on that of the leading manager, but not vice versa. This assumed hierarchy of dependence, which could be ontological or causal, would be enforced by the architecture of the LSTM.

In virtue of the architecture, the encoding is unable to capture that the decision state of the leading manager might constitutively depend on that of the department. For example, if the decision of the leading manager is the result of deferring to the attitude of the group of managers, this would not be appropriately reflected. Hence, the assumptions underlying the LSTM create an ontological commitment, undermining the neutrality of ANNs.

<sup>20</sup> One way to do this would be to introduce some finite vocabulary of decision states and treat it just analogous to a natural language processing task. The selection of this vocabulary would be another instance where selection of input creates an opening to ontological assumptions.

Mitigating such ontological troubles is a trend towards broad-purpose architectures in neural networks. The transformer-architecture (Vaswani et al. 2017), which has overtaken the natural language processing world, largely displacing LSTMs, is one example of such a more generic architecture. Via a so-called attention-mechanism they can account for dependence relations and learn their relative strength from the data. The researcher can withhold judgement and let the data do more of the work.<sup>21</sup>

As can be seen, architectural choices influence the ontological commitment of neural models. In addition, they affect the interpretability of ANNs. For example, the attention scores can be extracted and used to interpret the functioning of the neural network (e.g. Mullenbach et al. 2018). A classical example for this can be found in neural machine translation. If the translation of an English sentence into a French one is undertaken using a transformer-architecture one can identify and visualise to which word tokens in the English sentence the mechanism attended when creating a French token (Bahdanau et al. 2014). Translating “This concerns all of us.” to “Cela nous concerne tous.”, the model puts more attention to “all” when outputting the token “tous”.

By studying attention scores, neural networks can help to uncover dependence relations, although one ought not overinterpret them. Standard ANNs with attention have no sense of the difference between ontological dependence and mere correlation. They will track whatever helps them to locally reduce the loss-function.<sup>22</sup> Nonetheless, the case of attention-mechanisms shows that the choice of architecture also influences the interpretability of ANNs, in addition to creating or avoiding ontological commitments.

## 5 The impact of mistaken assumptions

I have established that while ANNs are ontologically less committed than other statistical and machine learning models, they are not entirely neutral either. But do the ontological assumptions underlying a neural network matter?

Of course, if ontological assumptions affect the performance of the network negatively, for example because a researcher chooses a CNN rather than an LSTM-architecture despite a long-distance linear dependence, then this is an issue worth addressing. The model might make worse predictions or classifications in virtue of mistaken assumptions.<sup>23</sup> However, that incorrect ontological assumptions lead to a performance penalty is far from given. It might very well be that the best-performing model a social scientist trains neglects some ontological dependence relations. Hence, the question arises whether mistaken assumptions can pose a problem when the performance on the existing benchmarks is not negatively affected.

For other types of statistical models, this is certainly the case. In the case of linear regression, social scientists are interested in the estimation of the parameters for a

---

<sup>21</sup> This description simplifies the situation, because the question of when to apply attention-mechanisms typically remains.

<sup>22</sup> It is important to note the local restriction of ANNs with given learning techniques. They are not guaranteed to find a global optimum in approximating a function.

<sup>23</sup> The performance issues might not only be realised in the accuracy value for the existing test data, but for example also in how broadly the model generalises beyond the originally available data.

specific feature, be it favourability rating in Gallup polls or family structure. If the feature coding for family structure is based on mistaken ontological assumptions, then this will lead to a misinterpretation of these parameters.

But the typical use of an ANN in the social sciences has no such ambitions. The use does not aim at interpreting the weights and biases of the neural networks directly. Nonetheless, a mistaken ontology can have at least two problematic consequences in the absence of a performance penalty.

First, the ANN might in fact address another predictive task than the one intended. This case is illustrated by the previous examples of trace data being selected to detect whether the features of a child's family allow to predict its future educational outcomes. If in the collection of the trace data, support networks are assumed to be exclusively constituted by nuclear families, then the ANN might effectively address another research goal than intended. It would only show whether the educational outcomes could be predicted on the basis of data about the parents, rather than the extended family. The selected input data might hide ontological assumptions undermining the research goal.

Second, even if an ANN addresses the correct task, incorrect ontological assumptions can limit the interpretation of the network. To see this issue, reconsider the case of a social scientist encoding the decision process in an organisation using a one-directional LSTM-architecture. The use of the LSTM assumes that that the decision state of the marketing department depends on that of the group of managers in the department which in its turn is dependent on that of the leading manager. With this architectural choice, the researcher never gave the ANN a chance to also learn dependence in the other direction. Hence, the architecture and its ontological commitments limit the possible interpretations of the model. Without comparing it to another architectural choice, we cannot conclude that the LSTM correctly captured the dependence relations.

In sum, the ontological assumptions of ANNs matter in at least three ways, for the performance, for meeting the task specification, and for interpretation. Neural networks are not ontologically neutral and it makes a difference for the purposes of the social sciences.

## 6 Conclusion

I compared ANNs to other statistical and machine learning models with a special focus on whether they can avoid problematic ontological assumptions in the social sciences. The result is that they can avoid more assumptions than other models, but are not free of them. Choices regarding the input and the architecture of neural networks will reflect ontological assumptions. There are, however, comparatively easy ways to mitigate these problems.

Common statistical models cannot deal with relatively raw data and require ontologically-laden feature selection, since they are not a form of representation-learning. For the time being, ANNs stand out in virtue their ontological flexibility.

Although it was not the focus of the present investigation, the issue of interpretability is on the mind of everyone interested in the use of neural networks for the social sci-



ences. The recent and quickly expanding literature on interpreting ANNs ameliorates this situation considerably, but neural networks are not well-suited for interpreting parameters for selected features of data. For this purpose, other statistical approaches are preferable.

While the limitations of ANNs should make social scientists hesitate to throw out their established tools in favour of the shiny new technology, which in fact has a decades long history, neural networks make a tempting offer. They are universal function-approximators with considerably fewer ontological assumptions than other approaches. While awareness of where the controversies of social ontology afflict neural networks is required, ANNs are the modelling tool best placed to avoid them.

**Acknowledgements** I thank Aris Spanos, Michael Messerli, Cameron Buckner, and the anonymous referees for their comments. This paper reports on research supported by Cambridge Assessment, University of Cambridge.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Angwin, J., & Lawson, J. (2016). Machine bias. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.
- Attewell, P. A., & Monaghan, D. B. (2015). *Data mining for the social sciences: An introduction* (1st ed.). Oakland, CA: University of California Press.
- Bagautdinov, T., Alahi, A., Fleuret, F., Fua, P., & Savarese, S. (2017). Social scene understanding: End-to-end multi-person action localization and collective activity recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4315–4324. [https://openaccess.thecvf.com/content\\_cvpr\\_2017/html/Bagautdinov\\_Social\\_Scene\\_Understanding\\_CVPR\\_2017\\_paper.html](https://openaccess.thecvf.com/content_cvpr_2017/html/Bagautdinov_Social_Scene_Understanding_CVPR_2017_paper.html).
- Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. [arXiv:1409.0473](https://arxiv.org/abs/1409.0473)
- Bainbridge, W. S. (1995). Neural network models of religious belief. *Sociological Perspectives*, 38(4), 483–495. <https://doi.org/10.2307/1389269>.
- Beck, N., King, G., & Zeng, L. (2000). Improving quantitative studies of international conflict: A conjecture. *The American Political Science Review*, 94(1), 21. <https://doi.org/10.2307/2586378>.
- Behr, D. (2015). Translating answers to open-ended survey questions in cross-cultural research: A case study on the interplay between translation, coding, and analysis. *Field Methods*, 27(3), 284–299. <https://doi.org/10.1177/1525822X14553175>.
- Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8), 1798–1828. <https://doi.org/10.1109/TPAMI.2013.50>.
- Buckner, C. (2019). Deep learning: A philosophical introduction. *Philosophy Compass*, 14(10), e12625. <https://doi.org/10.1111/phc3.12625>.



- Buckner, C., & Garson, J. (2019). Connectionism. In E.N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*, fall 2019 edn, Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/fall2019/entries/connectionism/>.
- Cichy, R. M., & Kaiser, D. (2019). Deep neural networks as scientific models. *Trends in Cognitive Sciences*, 23(4), 305–317. <https://doi.org/10.1016/j.tics.2019.01.009>.
- Cooper, J. C. B. (1999). Artificial neural networks versus multivariate statistics: An application from economics. *Journal of Applied Statistics*, 26(8), 909–921. <https://doi.org/10.1080/02664769921927>.
- Corbett-Davies, S., Pierson, E., Feller, A., Goel, S., & Huq, A. (2017). Algorithmic decision making and the cost of fairness. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, Association for Computing Machinery, pp. 797–806. <https://doi.org/10.1145/3097983.3098095>.
- Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems*, 2(4), 303–314. <https://doi.org/10.1007/BF02551274>.
- Davidson, T. (2019). Black-box models and sociological explanations: Predicting high school grade point average using neural networks. *Socius*, 5, 1–11. <https://doi.org/10.1177/2378023118817702>.
- de Marchi, S., Gelpi, C., & Grynawski, J. D. (2004). Untangling neural nets. *American Political Science Review*, 98(2), 371–378. <https://doi.org/10.1017/S0003055404001200>.
- Epstein, B. (2015). *The ant trap: Rebuilding the foundations of the social sciences*. Oxford studies in philosophy of science, Oxford University Press, New York, NY.
- Epstein, B. (2019). What are social groups? Their metaphysics and how to classify them. *Synthese*, 196(12), 4899–4932. <https://doi.org/10.1007/s11229-017-1387-y>.
- Falat, L., & Pancikova, L. (2015). Quantitative modelling in economics with advanced artificial neural networks. *Procedia Economics and Finance*, 34, 194–201. [https://doi.org/10.1016/S2212-5671\(15\)01619-6](https://doi.org/10.1016/S2212-5671(15)01619-6).
- Flores, A. W., Bechtel, K., & Lowenkamp, C. T. (2016). False positives, false negatives, and false analyses: A rejoinder to machine bias: There's software used across the country to predict future criminals. And it's biased against blacks. *Federal Probation*, 80(2), 38–46.
- Gambäck, B., & Sikdar, UK. (2017). Using convolutional neural networks to classify hate-speech. In *Proceedings of the first workshop on abusive language online*, Association for Computational Linguistics, Vancouver, BC, Canada, pp. 85–90. <https://doi.org/10.18653/v1/W17-3013>, <http://aclweb.org/anthology/W17-3013>.
- Garson, G. D. (1998). *Neural networks: An introductory guide for social scientists. New technologies for social research*. London: Sage.
- Glymour, B., & Herington, J. (2019). Measuring the biases that matter: The ethical and casual foundations for measures of fairness in algorithms. In *Proceedings of the conference on fairness, accountability, and transparency*, FAT\* '19, pp 269–278. <https://doi.org/10.1145/3287560.3287573>.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning. Adaptive computation and machine learning*. Cambridge, MA: The MIT Press.
- Guest, D., Cranmer, K., & Whiteson, D. (2018). Deep learning and its application to LHC physics. *Annual Review of Nuclear and Particle Science*, 68(1), 161–181. <https://doi.org/10.1146/annurev-nucl-101917-021019>.
- Hawley, K. (2017). Social mereology. *Journal of the American Philosophical Association*, 3(4), 395–411. <https://doi.org/10.1017/apa.2017.33>.
- Herbrich, R., Keilbach, M., Graepel, T., Bollmann-Sdorra, P., & Obermayer, K. (1999). Neural networks in economics. In T. Brenner (Ed.), *Computational techniques for modelling learning in economics, advances in computational economics* (pp. 169–196). Boston, MA: Springer. [https://doi.org/10.1007/978-1-4615-5029-7\\_7](https://doi.org/10.1007/978-1-4615-5029-7_7).
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>.
- Hofman, J. M., Sharma, A., & Watts, D. J. (2017). Prediction and explanation in social systems. *Science*, 355(6324), 486–488. <https://doi.org/10.1126/science.aal3856>.
- Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5), 359–366. [https://doi.org/10.1016/0893-6080\(89\)90020-8](https://doi.org/10.1016/0893-6080(89)90020-8).
- Howison, J., Wiggins, A., & Crowston, K. (2011). Validity issues in the use of social network analysis with digital trace data. *Journal of the Association for Information Systems*, 12(12), 767–797. [10.17705/1jais.00282](https://doi.org/10.17705/1jais.00282).

- Huebner, B. (2014). *Macro cognition: A theory of distributed minds and collective intentionality*. New York: Oxford University Press.
- Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301), 790–794. <https://doi.org/10.1126/science.aaf7894>.
- Kane, L. W. (2019). What is a family? Considerations on purpose, biology, and sociality. *Public Affairs Quarterly*, 33(1), 65–88.
- Khan, I., & Kulkarni, A. (2013). Knowledge extraction from survey data using neural networks. *Procedia Computer Science*, 20, 433–438. <https://doi.org/10.1016/j.procs.2013.09.299>.
- Lakhani, P., & Sundaram, B. (2017). Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology*, 284(2), 574–582. <https://doi.org/10.1148/radiol.2017162326>.
- LeCun, Y., Boser, B. E., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. E., et al. (1990). Handwritten digit recognition with a back-propagation network. In D. S. Touretzky (Ed.), *Advances in neural information processing systems 2* (pp. 396–404). Burlington: Morgan-Kaufmann.
- Lee, H. K. H. (2004). Bayesian nonparametrics via neural networks. *Society for Industrial and Applied Mathematics*, 10(1137/1), 9780898718423.
- Lewis-Beck, M. S., & Rice, T. W. (1982). Presidential popularity and presidential vote. *Public Opinion Quarterly*, 46(4), 534–537. <https://doi.org/10.1086/268750>.
- Li, Y., & Ma, W. (2010). Applications of artificial neural networks in financial economics: A survey. In *2010 international symposium on computational intelligence and design*, Vol. 1, pp. 211–214. <https://doi.org/10.1109/ISCID.2010.70>
- Liao, WH., Huang, YT., Yang, TH., & Wu, YC. (2019). Analyzing social network data using deep neural networks: A case study using Twitter posts. In *2019 IEEE international symposium on multimedia (ISM)*, pp. 237–2371. <https://doi.org/10.1109/ISM46123.2019.00053>.
- Lipton, Z. C. (2018). The mythos of model interpretability. *Communications of the ACM*, 61(10), 36–43. <https://doi.org/10.1145/3233231>.
- List, C., & Pettit, P. (2011). *Group agency: The possibility, design, and status of corporate agents*. Oxford: Oxford University Press.
- López-Iturriaga, F. J., & Sanz, I. P. (2018). Predicting public corruption with neural networks: An analysis of Spanish provinces. *Social Indicators Research*, 140(3), 975–998. <https://doi.org/10.1007/s11205-017-1802-2>.
- Mancosu, M., & Bobba, G. (2019). Using deep-learning algorithms to derive basic characteristics of social media users: The Brexit campaign as a case study. *PLoS ONE*, 14(1), e0211013. <https://doi.org/10.1371/journal.pone.0211013>.
- McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4), 115–133. <https://doi.org/10.1007/BF02478259>.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Proceedings of the 26th international conference on neural information processing systems - Volume 2*, Curran Associates Inc., USA, NIPS'13, pp. 3111–3119. <http://dl.acm.org/citation.cfm?id=2999792.2999959>.
- Minsky, M. L., & Papert, S. (1972). *Perceptrons: An introduction to computational geometry*. Cambridge, MA: MIT Press.
- Mullenbach, J., Wiegrefe, S., Duke, J., Sun, J., & Eisenstein, J. (2018). Explainable prediction of medical codes from clinical text. In *Proceedings of the 2018 conference of the North American chapter of the association for computational linguistics: Human language technologies*, Volume 1 (Long Papers), Association for Computational Linguistics, pp. 1101–1111. <https://doi.org/10.18653/v1/N18-1100>, <http://aclweb.org/anthology/N18-1100>.
- Obermeyer, Z., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm that guides health decisions for 70 million people. In *Proceedings of the conference on fairness, accountability, and transparency—FAT\* '19*, pp. 89–89. <https://doi.org/10.1145/3287560.3287593>.
- Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on empirical methods in natural language processing—EMNLP '02*, Association for Computational Linguistics, vol 10, pp. 79–86. <https://doi.org/10.3115/1118693.1118704>.

- Polhill, G., & Salt, D. (2017). The importance of ontological structure: Why validation by ‘Fit-to-Data’ is Insufficient. In B. Edmonds & R. Meyer (Eds.), *Simulating social complexity* (pp. 141–172). Berlin: Springer.
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why Should I Trust You?”: Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1135–1144. <https://doi.org/10.1145/2939672.2939778>.
- Ritchie, K. (2013). What are groups? *Philosophical Studies*, 166(2), 257–272. <https://doi.org/10.1007/s11098-012-0030-5>.
- Ritchie, K. (2015). The metaphysics of social groups. *Philosophy Compass*, 10(5), 310–321. <https://doi.org/10.1111/phc3.12213>.
- Ritchie, K. (2020). Social structures and the ontology of social groups. *Philosophy and Phenomenological Research*, 100(2), 402–424. <https://doi.org/10.1111/phpr.12555>.
- Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386–408. <https://doi.org/10.1037/h0042519>.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533–536. <https://doi.org/10.1038/323533a0>.
- Salganik, M. J., Lundberg, I., Kindel, A. T., Ahearn, C. E., Al-Ghoneim, K., Almaatouq, A., et al. (2020). Measuring the predictability of life outcomes with a scientific mass collaboration. *Proceedings of the National Academy of Sciences*, 117(15), 8398–8403. <https://doi.org/10.1073/pnas.1915006117>.
- Satz, D. (2017). Feminist perspectives on reproduction and the family. In E.N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy*, summer 2017 edn, Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/sum2017/entries/feminism-family/>.
- Schmidt, J., Marques, M. R. G., Botti, S., & Marques, M. A. L. (2019). Recent advances and applications of machine learning in solid-state materials science. *NPJ Computational Materials*, 5(1), 1–36. <https://doi.org/10.1038/s41524-019-0221-0>.
- Sheehy, P. (2006). *The reality of social groups*. Aldershot: Ashgate.
- Siegelmann, H. T., & Sontag, E. D. (1992). On the computational power of neural nets. In *Proceedings of the fifth annual workshop on Computational learning theory*, pp. 440–449. <https://doi.org/10.1145/130385.130432>.
- Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., et al. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587), 484–489. <https://doi.org/10.1038/nature16961>.
- Spanos, A. (2006). Where do statistical models come from? Revisiting the problem of specification. In *Optimality: The Second Erich L. Lehmann symposium*, Institute of Mathematical Statistics, pp. 98–119. <http://projecteuclid.org/euclid.lnms/1196283957>.
- Spanos, A., & Mayo, D. G. (2015). Error statistical modeling and inference: Where methodology meets ontology. *Synthese*, 192(11), 3533–3555. <https://doi.org/10.1007/s11229-015-0744-y>.
- Stier, S., Breuer, J., Siegers, P., & Thorson, K. (2020). Integrating survey data and digital trace data: Key issues in developing an emerging field. *Social Science Computer Review*, 38(5), 503–516. <https://doi.org/10.1177/0894439319843669>.
- Strohmaier, D. (2018). Group membership and parthood. *Journal of Social Ontology*, 4(2), 121–135. <https://doi.org/10.1515/jso-2018-0016>.
- Strohmaier, D. (2020). Two theories of group agency. *Philosophical Studies*, 177(7), 1901–1918. <https://doi.org/10.1007/s11098-019-01290-4>.
- Thomasson, A. L. (2019). The ontology of social groups. *Synthese*, 196(12), 4829–4845. <https://doi.org/10.1007/s11229-016-1185-y>.
- Tollefsen, D. (2015). *Groups as agents*. Malden, MA: Polity.
- Uzquiano, G. (2004). The supreme court and the supreme court justices: A metaphysical puzzle. *Noûs*, 38(1), 135–153.
- Uzquiano, G. (2018). Groups: Toward a theory of plural embodiment. *The Journal of Philosophy*, 115(8), 423–452. <https://doi.org/10.5840/jphil2018115825>.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, AN., Kaiser, u., & Polosukhin, I. (2017). Attention is all you need. In I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in neural information processing systems 30*, pp. 5998–6008. <http://papers.nips.cc/paper/7181-attention-is-all-you-need.pdf>.
- Waldfogel, J., Craigie, T. A., & Brooks-Gunn, J. (2010). Fragile families and child wellbeing. *The Future of Children*, 20(2), 87–112. <https://doi.org/10.1353/foc.2010.0002>.

- Watts, D. J., Beck, E. D., Bienenstock, E. J., Bowers, J., Frank, A., Grubestic, A., et al. (2018). Explanation, prediction, and causality: Three sides of the same coin? *OSF Preprints*. <https://doi.org/10.31219/osf.io/u6vz5>.
- Zeng, L. (1999). Prediction and classification with neural network models. *Sociological Methods & Research*, 27(4), 499–524. <https://doi.org/10.1177/0049124199027004002>.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.