

How can Data Predict our Behaviour?

Dr Joanne Hinds and Professor Adam Joinson

Human and computer-based prediction

Is it possible to judge someone accurately from their online activity? We often rely on our internet-based judgments of others to make decisions, such as who to socialise with, date or employ. Recently, researchers have turned to studying social media and digital devices in order to ask whether a person's digital traces can reveal aspects of their identity. Simultaneously, advances in 'big data' analytics have demonstrated that computer algorithms can predict an individual's personality traits or demographic attributes from their digital traces. In our research, we have conducted two systematic reviews to address:

- 1) What do we currently know about human- and computer based assessments?
- 2) How accurate are these assessments?

The 'Big Five' is the most popular approach currently used by researchers when measuring personality. Assessments involve evaluating how highly individuals score across five dimensions – Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. Figure 1. below highlights how humans formulate judgments of others using digital 'cues' (also known as the Brunswik lens model).

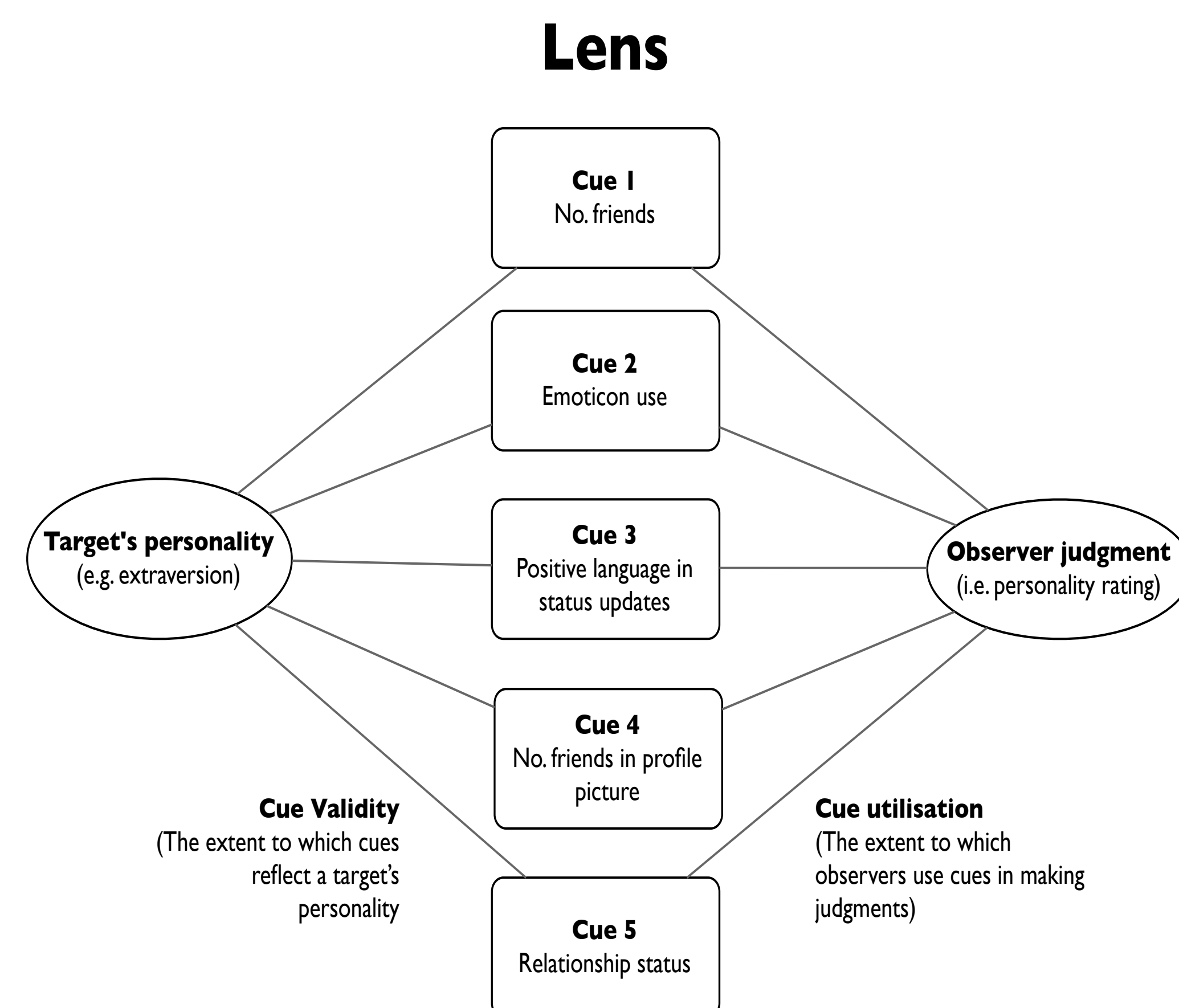


Figure 1. Brunswik lens model

Predicting individuals' demographic attributes is well established in areas such as computer forensics and computational linguistics which often use text-based sources to predict an individual's age and gender. Figure 2. shows how computer algorithms can 'learn' how to use digital data to make predictions.

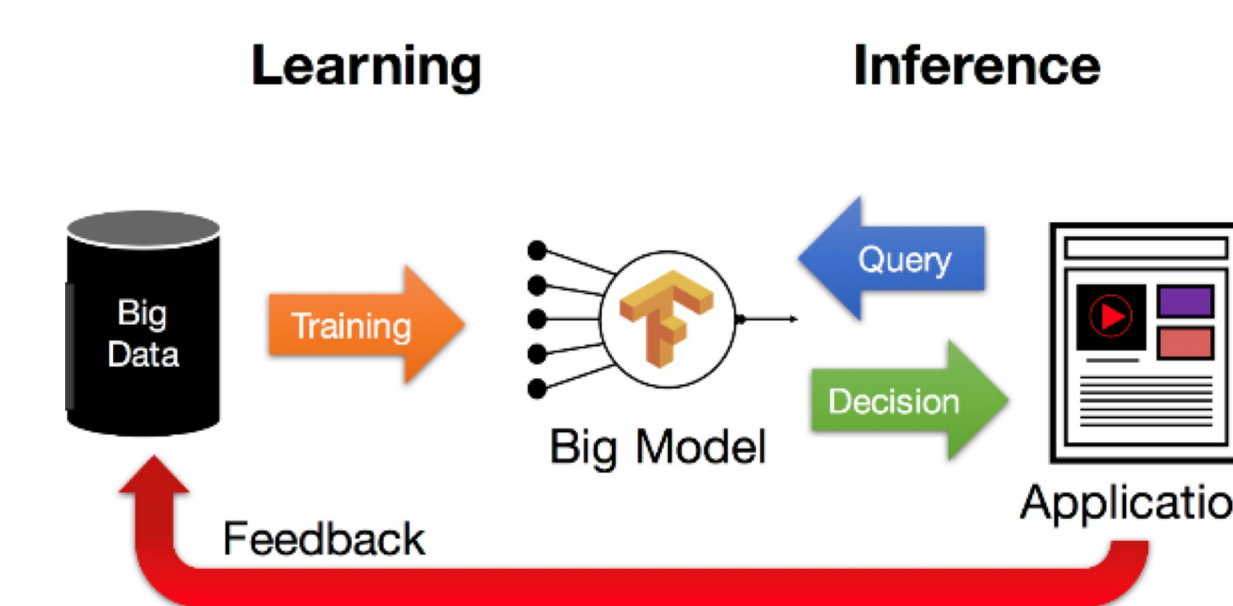


Figure 2. Machine learning for demographic prediction

Method and Results

We performed two systematic reviews (1 x personality prediction and 1 x demographics prediction) from digital traces. Through analysis of approx. 25,000 academic articles, we found that personality prediction has a mixture of research situated in both psychology and computer science, whereas demographic prediction is almost entirely conducted by machine learning (in computer science). Figure 3. presents findings from two meta-analyses, allowing us to comparing human and computer-based personality prediction.

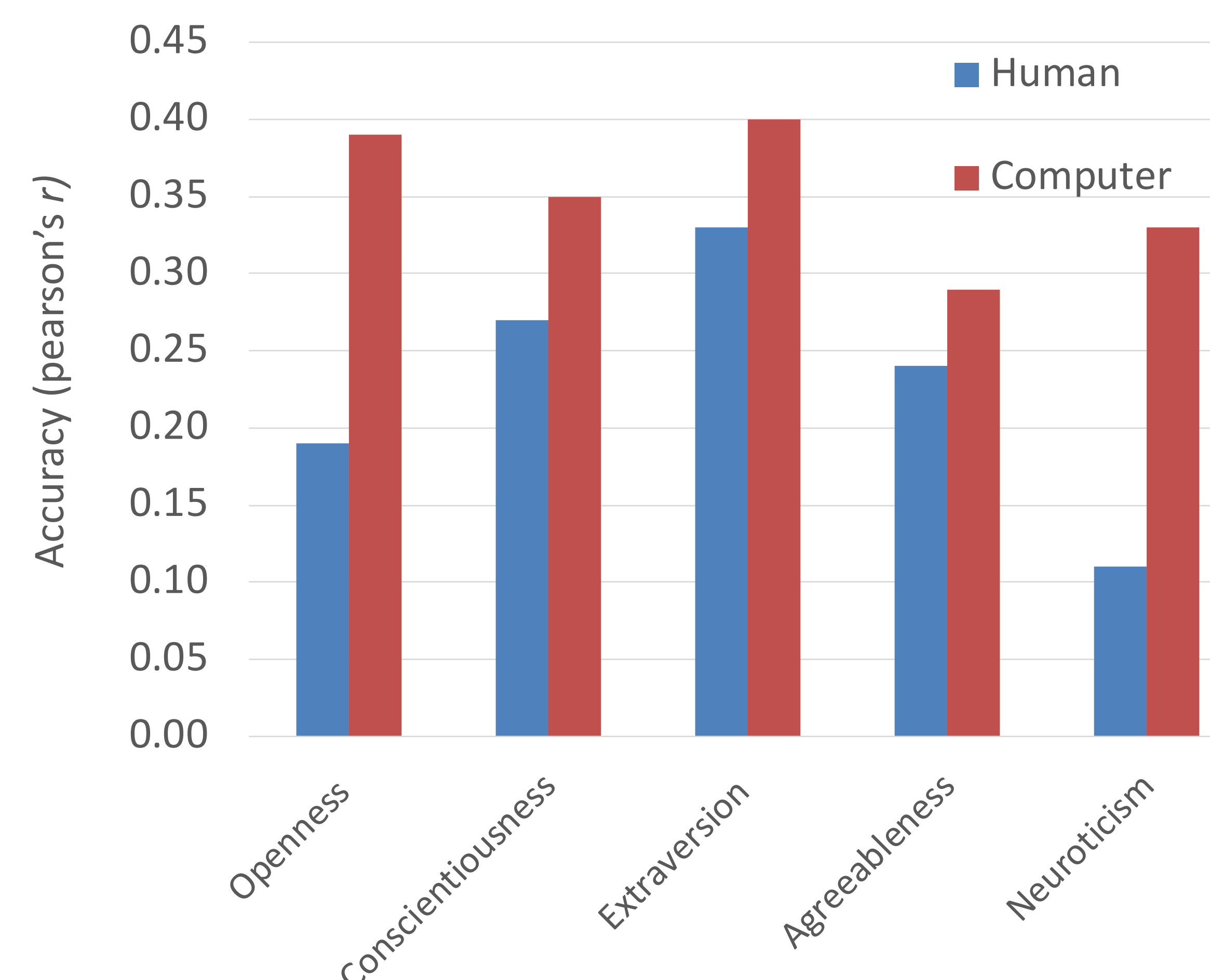


Figure 3. Meta-analyses of human and computer-based personality prediction

Figure 4 highlights the range of demographic attributes examined (by no. articles) up until January 2018.

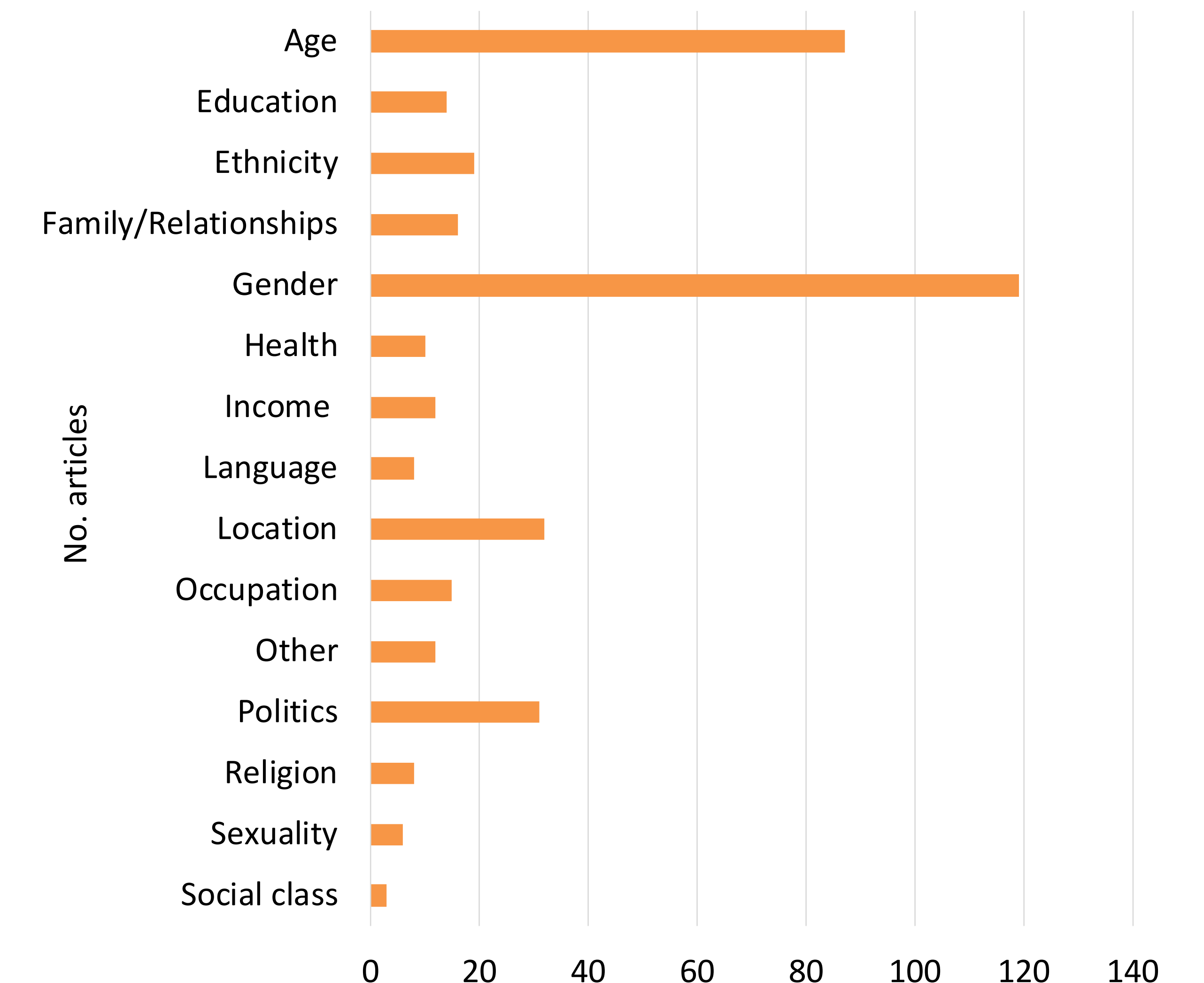


Figure 4. Trends in research for demographic prediction

Ethics and challenges

The ability to predict behaviour accurately provides numerous ethical issues and challenges. For instance, individuals, who are similar (in age, location interests, etc.) tend to be friends with, or connected with each other. The notion that 'birds of a feather flock together' often reflects in individuals' social networks online. These patterns in human networks online can create the potential for shadow profiling – where an individual's undisclosed or private information is revealed or inferred from data accessed through other people within their network.

Amidst concerns of how data is collected, used, shared and what true 'informed consent' really is, many people feel uncomfortable with the notion that their devices are 'listening' or that their behaviour is constantly being monitored or analysed. Whilst it is an exciting time for technological advancement and social science, organisations and cybersecurity practitioners face some complex challenges when it comes to handling data carefully and reinforcing trust in using technology.