Chapter 8
Describing Human Behaviour Through Computational Social Science

Giuseppe A. Veltri

Abstract The possibilities offered by digital and Computational Social Science can improve our understanding of human behaviour as never before. The availability of behavioural data in a society where the digital has been widely adopted is because of two reasons: first, the vast amount of digital traces produced by people in their daily lives and related behaviours and, second, the possibility of running online experiments that can cover a large segment of a target population (we have seen online experiments with hundreds of thousands of participants). This chapter will discuss the opportunity offered by online large behavioural experiments. The implications for policymakers of this shift are the possibility of having behavioural insights both across different societies and better understanding and capturing within a country heterogeneity. In other words, large-scale online experiments combined with computational methods allow for unprecedented cognitive and behavioural based segmentation.

8.1   Introduction

This chapter describes the role of Computational Social Science in enhancing our understanding of human behaviour. We will highlight the importance of behavioural data compared to the more common self-reported ones and how the increased availability of digital traces of human behaviour is crucial in the new potential analysis.

The digital revolution that has affected the social sciences in the past decade or so (e.g. Salganik, 2018; Veltri, 2020) created the context for three possible forms of studies about human behaviour:

1. The use of large online behavioural experiments, which will be the focus of this chapter

G. A. Veltri (✉)
Department of Sociology and Social Research, University of Trento, Trento, Italy
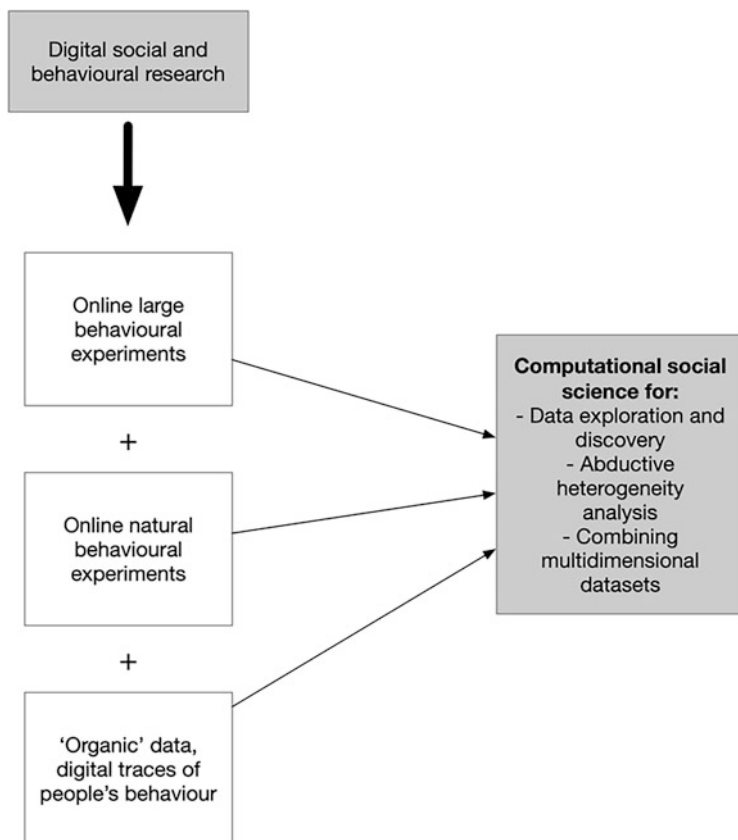e-mail: giuseppe.veltri@unitn.it

Fig. 8.1 Different forms of behavioural data available due to the 'digital revolution' in the social sciences and the role of CSS in it

2. Similarly to point one, the online has become a large field in which natural experiments occur
3. Many digital platforms collect behavioural data of their users and these can be repurposed for social scientific purposes

The amount, type, and complexity of data generated from the three approaches above require innovation from the analytical point of view. This is where Computational Social Science methods are being applied (Fig. 8.1). Finally, we will discuss one promising form of modelling that we believe is particularly relevant for studies of human behaviour and decision-making.

Large-scale online experiments can test the behavioural response to intervention and explore the heterogeneity of treatment effects across different social strata helping to tailor differentiated options. The use of computational models to identify subgroups in large datasets is a growing area of interest. Of particular interest is the form of online experiments that combine the lessons learned from online

surveys. The so-called population-based survey experiments or PBSEs aim to address this problem through research design rather than analyses, combining the best aspects of both approaches, capitalizing on their strengths, and eliminating many weaknesses. We will discuss the potential of unstructured data such as text and the use of text mining techniques that, combined with other types of data, can further enrich our understanding of social behaviour. The contribution of Computational Social Science to our understanding of social behaviour is not limited to data availability. However, it includes the opening to analytical approaches developed in computer science, particularly in machine learning, which brings a new 'culture' of statistical modelling that bears considerable potential for the social scientist and informs policy harnessing heterogeneity. Segmentation is widely used in decision-making. It is usually based on sociodemographic factors (e.g. age, gender, income, geographical location). However, cognitive and behavioural differences within the population are an important source of variability that needs to be considered in policy design and implementation because how one reacts to a public policy is conditioned by what cognitive and cultural patterns are used by those targeted. The target population is behaviourally and cognitively plural: people vary in how they feel, think, and act. Moreover, each citizen interprets reality according to different cognitive and cultural schemas specific to the individual and active in the cultural environment. Therefore, public policies need to be designed in ways that allow for the flexibility to take into account differences in the target population and other dimensions (e.g. sociodemographic, linguistic, and so on).

8.2 Data in the Digital World

To appreciate the transformative nature of digital data and computational methods in the social sciences, we need to draw some fundamental distinctions between the types of data that social scientists have been dealing with. A great deal of social science research has been produced based on self-reported data. Self-reported data stands for the accounts and reporting people do about their views, psychological states, and behaviours. However, the biggest challenge to self-reported data has come from a shift in the model of human behaviour in the wider social sciences except for psychology. Since the late 1990s, psychologists have distinguished between two systems of thought with different capacities and processes (Kahneman, 2012; Kahneman & Frederick, 2002; Lichtenstein & Slovic, 2006; Metcalfe & Mischel, 1999; Sloman, 1996; Smith & DeCoster, 2000), which have been referred to as System 1 and System 2 (Evans & Stanovich, 2013). System 1 (S1) is made up of intuitive thoughts of great capacity, is based on associations acquired through experience and quickly and automatically calculates information.

On the other hand, System 2 (S2) involves low-capacity reflective thinking based on rules acquired through culture or formal learning and calculates information in a relatively slow and controlled manner. The processes associated with these systems have been defined as Type 1 (fast, automatic, unconscious) and Type 2

(slow, conscious, controlled), respectively. The so-called dual model of the mind is now the most supported way of understanding human behaviour at the individual level and in continuous evolution (De Neys, 2018). The model has also been applied outside psychology, for example, in sociology (Lizardo et al., 2016; Moore, 2017) and political science (Achen & Bartels, 2017). The implications of Kahneman and Tversky's work have led to the research programme labelled behavioural economics, which has dramatically impacted traditional microeconomics theory.

A more precise human behaviour and decision-making model has implications for social science research methodology, particularly for the distinction between self-reported and observational/behavioural data. The dual mode of thinking brings back the importance of unconscious thought processes and contextual and environmental influences on the latter in the broader context of the social sciences, which is highly problematic in studies only using self-reported measures and instruments. Traditionally, collecting behavioural data has been very difficult and expensive for social scientists. Keeping track of people's actual behaviour could be done only for small groups of people and for a minimal amount of time. However, the availability of digital data has brought us a significant increase in behavioural data; we now have digital traces of people's actual behaviour that were never available before.

The combined effect of a relatively new and powerful foundational model of human behaviour and decision-making offered by the dual model and the availability of behavioural data thanks to the digital traces recorded by a multitude of services and tools is very promising for social scientists. Before continuing this line of argument, let's clarify one point that might be the object of criticism—considering human behaviour as the outcome of mutual influences of conscious acting and unconscious heuristics, biases, and environmental influences are not a return to reductionism. People's opinions count for nothing. Self-reported data will remain an essential source of information for social scientists, but, at the same time, the availability of behavioural data will function as complementary data to understand complex social phenomena.

8.3   Behavioural Digital Data

The distinction between self-reported and behavioural data is no longer mainly theoretical because the new opportunities for collecting the latter are unprecedented. Such an option opens new research opportunities and the possibility of reviewing current theories and existing models. However, the increased availability of collecting data about people's behaviour does not free us from biases generated by the design and aims of digital platforms. People's behaviour is constrained by the platform they use; for example, it is impossible to write an essay on Twitter unless we decide to write it using many individual tweets. There are, therefore, several potential sources of confounding factors, as we will further elaborate in the section below on construct validity.

Another distinction is relevant about the different levels of analysis: the one between static and dynamic data. The large majority of data collected in the social sciences have been 'static'—that is, data collection has been carried out at a given time. This is because longitudinal data collection, data collected over some time, was challenging and expensive. Digital data introduce a much-increased capacity for recording and using longitudinal data for social scientific purposes. Digital data have not been historically around for many decades, but future researchers might have at their disposal longitudinal datasets that were absent in the past.

Behavioural digital data are the object of attention of a new generation of social scientists who believe in their potential to regenerate the current theories and framework that were developed in a condition of data scarcity, with different models of human behaviour and using only self-reported data. It is too early to say what changes will bring increased data availability, but this is the most exciting aspect of the use of digital data for social scientific research. However, the nature of the data collected from the digital world is not without problems, and it poses specific challenges to researchers.

The distinction between self-reported and behavioural data touched briefly on one feature of digital data: their nonreactive nature in terms of data collection. We can distinguish digital data as the outcome of unobtrusive or obtrusive data collection methods (Webb et al., 1966). The distinction between these two data collection modalities is essential in the social sciences because people 'react' to researchers' measurements and can figure out what a researcher's goals are. Two of the most common problems are people's reactions to measurements, the Hawthorne effect and the social desirability effect. The Hawthorne effect, as mentioned before, refers to the fact that individuals modify their behaviour in response to their awareness of being observed. Recent scandals related to social media and privacy, in which users' data have been harvested for commercial or political campaigning purposes, have made people more conscious that their online behaviour is observed and recorded. Social desirability is the tendency of some respondents to report an answer in a way they deem to be more socially acceptable if they believe they are under observation than would be their 'true' answer, where true means aligned to current dominant social norms. They do this to project a favourable image of themselves and to avoid receiving negative evaluations. The outcome of the strategy results in the over-reporting of socially desirable behaviours or attitudes and the under-reporting of socially undesirable behaviours or attitudes (Nederhof, 1985). Social media are particularly affected by social desirability bias because people manage their presence online to generate a positive self-image. This process leads to a positivity bias in the content present on social media (Spottswood & Hancock, 2016).

The availability of behavioural data in a society where the digital has been widely adopted is because of two reasons: first, the vast amount of digital traces produced by people in their daily lives and related behaviours and, second, the possibility of running online experiments that can cover a large segment of a target population (we have seen online experiments with hundreds of thousands of participants). Next, we

will discuss the opportunity offered by the online large behavioural experiments, particularly in the form of population-based survey experiments or PBSE.

8.4   Online Population-Based Survey Experiments

The limitations of laboratory experiments and the opportunities of the digital as a field in which to conduct research have prompted researchers to develop online experiments both in academia and in the private research world. Of particular interest is the form of online experiments that combine lessons learned from online surveys. The aim of so-called population-based survey experiments, or PBSEs (Mutz, 2011), is to address this problem through research design rather than analysis, combining the best aspects of both approaches, capitalizing on their strengths, and eliminating many of their weaknesses.

Defined in the most rudimentary terms, a population-based survey experiment is an experiment that is administered to a representative sample of the population. Another common term for this approach is simply 'survey experiment', but this abbreviated form can be misleading because it is not always clear what the term 'survey' means. The use of survey methods does not distinguish this approach from other combinations of survey and experimental methods. After all, many experiments already involve survey methods in the administration of pre-test and post-test questionnaires, but this is not what is meant here. Population-based survey experiments are not defined by their use of interview techniques, whether written or oral nor by their location in a setting other than a laboratory. Instead, a population-based experiment uses sampling methods to produce a set of experimental subjects that is representative of the target population of interest for a particular theory, whether that population is a country, a state, an ethnic group, or some other subgroup. The population represented by the sample should be representative of the population to which the researcher intends to extend his results. In population-based survey experiments, experimental subjects are randomly assigned to conditions by the researcher, and treatments are administered as in any other experiment. Nevertheless, participants are generally not required to show up in a laboratory to participate. Theoretically, they could, but population-based experiments are infinitely more practical when representative samples are not required to appear in one place.

Strictly speaking, population-based survey experiments are more experiments than surveys. By design, population-based experiments are experimental studies that draw on the power of random assignment to establish unbiased causal inferences. They are also administered to randomly selected representative samples of the target population of interest, just as a survey would be. However, population-based experiments do not need (and often have not relied on) nationally representative population samples. The population of interest could be members of a particular ethnic group, parents of children under 18, people who watch television news, or

others. Still, the key is that convenience samples are abandoned in favour of samples representing the target population of interest.

The advantage of population-based survey experiments is that theories can be tested on samples that are representative of the populations to which they are said to apply. The downside of this trade-off is that most researchers have little experience administering experimental treatments outside of a laboratory setting, so new techniques and considerations come into play, as (Veltri, 2020) described extensively. In a sense, population-based survey experiments are by no means new; simplified versions of them have existed since at least the early years of research. However, technological developments in survey research, combined with the development of innovative techniques in experimental design, have made highly complex and methodologically sophisticated population-based experiments increasingly accessible to social scientists from many disciplines.

The development of the digital has made implementable the possibilities of population-based experiments. With the diffusion of pre-recruited online panels that are built according to the golden standards of sampling, the ability to exploit such dynamic data collection tools has expanded social scientists' methodological repertoire and inferential range in many fields (e.g. Veltri et al., 2020). The many advances in interview technology offer social science researchers the potential to introduce some of its most important hypotheses into virtual laboratories scattered across countries. Whether evaluating theoretical hypotheses, examining the robustness of laboratory results, or testing empirical hypotheses of other varieties, the ability of scientists to experiment on large and diverse groups of subjects allows them to address critical social and behavioural phenomena more effectively and efficiently.

Population-based experiments can be used by social scientists in sociology, political science, psychology, economics, cognitive science, law, public health, communication, and public policy, to name just a few of the main fields that find this approach appealing. Although most social scientists recognize the enormous benefits of experimentation, the traditional laboratory setting is unsuitable for all important research questions. Experiments have always been more prevalent in some social science fields than in others. To a large extent, the emphasis on experimental versus investigative methods reflects a field's emphasis on internal versus external validity, with fields such as psychology more oriented towards the former and fields such as political science and sociology more oriented towards the latter. For researchers, population-based experiments provide a means of establishing causality that is unmatched by any large-scale data collection effort, no matter how extensive.

Conducting online population-based survey experiments can benefit from the latest development of survey design and, in particular, adaptive survey design or ASD. ASD (Wagner, 2010) is based on the premise that samples are heterogeneous, and the optimal survey protocol may not be the same for each individual. For example, a particular survey design feature such as incentives may appeal to some individuals but not to others (Groves et al., 2000; Groves & Heeringa, 2006), leading to design-specific response propensity for each individual. Similarly, relative to

interviewer-administration, a self-administered mode of data collection may elicit less measurement error bias for some individuals but more measurement error bias for others. The general objective in ASD is to tailor the protocol to sample members to improve targeted survey outcomes. The basic premise of adaptive interventions is shared by ASDs—tailoring methods to individuals based on interim outcomes. We label these *dynamic* adaptive designs to reflect the dynamic nature of the optimization and *static* adaptive designs when they are based solely on information available prior to the start of data collection. A tailoring variable is used to inform the decision to change treatments, such as the type of concerns the sample member may have raised at the contact moment. Decision rules would include the matching of information from the tailoring variables (concerns about time, not worth their effort) to interventions (a shorter version of the task, a larger incentive). Finally, the decision points need to be defined, such as whether to apply the rules and intervene at the time of the interaction or at a given point in the data collection period.

What is noteworthy is that either of these approaches and much more complex experimental designs are easily implemented in the context of use of online platforms. The ability to make strong causal inferences has little to do with the laboratory environment itself and much to do with the ability to control the random assignment of people to different experimental treatments. By moving the possibilities of experimentation out of the laboratory in this way, population-based experiments strengthen the internal validity of social science research and provide the potential to interest a much wider group of social scientists in the possibilities of experimentation. Of course, the fact that it can be done outside the laboratory is not itself a good reason to do so. Therefore, we will review some of the key advantages of online population-based experiments, starting with four advantages over traditional laboratory experiments and then ending with some of their more general benefits for accumulating valuable social scientific knowledge.

The main strategic advantage of an online experiment over a laboratory experiment is the greater possibility of generalization (external validity), the greater statistical power and possibly the quality of the data produced. Web-based studies, having larger samples, usually have greater statistical power than laboratory studies. Data quality can be defined by variable error, constant error, reliability, or validity. Comparisons of power and some quality measures have found cases where web data are of higher quality for one or other of these definitions than comparable laboratory data, although not always (Birnbaum, 2004). Many web researchers are convinced that data obtained via the web can be 'better' than data obtained from students (Reips, 2002), despite the laboratory's obvious advantage for control. The main disadvantage of an online experiment compared to a laboratory experiment is the lack of complete environmental control. Participants in online experiments may answer questions and perform behavioural tasks in very different environments (a room with light and silence, versus their own desk at work with less light and surrounded by much noise) and with different equipment (a participant may use a browser that does not display visual stimuli correctly or may have a slow connection, thus delaying task completion and increasing fatigue, frustration and 'noisy' responses). Most importantly, as lab assistants do not monitor participants,

there is more chance that they will engage in automatic responses and task completion, which introduces noise into the data. This can be controlled with control questions and is less of a problem for between-subjects design with randomization of treatments and control conditions.

Other technical/tactical issues can be controlled for in the online experiment (multiple submissions, drop-outs, self-selection). Still, the main trade-off between online experiments and laboratory is to trade off greater generalizability and power of data for less experimental control. Therefore, it is not surprising that experiments are often repeated with the same outcome measures both online and in the laboratory to check the quality and validity of the data.

8.5   Heterogeneity Analysis and Computational Methods

Extending experiments to large samples, both national and international, increases the potential heterogeneity present in response to our treatments. Therefore, identifying and studying such heterogeneity is a crucial step in the world of online behavioural experiments. New analytical techniques have emerged in computational and computer sciences that are very promising to achieve this goal. One of the best examples of how social science can benefit from analytical approaches developed in computational methods is the development of model-based recursive partitioning. This approach improves the use of classification and regression trees. The latter also is a method from the 'algorithmic culture' of modelling that has valuable applications in the social sciences but is essentially data-driven (Berk, 2006; Hand & Vinciotti, 2003).

In summary, classification and regression trees are based on a purely data-driven paradigm. Without using a predefined statistical model, such algorithmic methods recursively search for groups of observations with similar response variable values by constructing a tree structure. Thus, they are instrumental in data exploration and express their best utility in the context of very complex and large datasets. However, such techniques make no use of theory in describing a pattern of how the data was generated and are purely descriptive, although far superior to the 'traditional' descriptive statistics used in the social sciences when dealing with large datasets.

Model-based recursive partitioning (Zeileis et al., 2008) represents a synthesis of a theoretical approach and a set of data-driven constraints for theory validation and further development. In summary, this approach works through the following steps. Firstly, a parametric model is defined to express a set of theoretical assumptions (e.g. through a linear regression). Second, this model is evaluated according to the recursive partitioning algorithm, which checks whether other important covariates that would alter the parameters of the initial model have been omitted. Third, the same regression or classification tree structure is produced. This time, instead of partitioning by different patterns of the response variable, model-based recursive partitioning finds different patterns of associations between the response variable and other covariates that have been pre-specified in the parametric model. In other

words, it creates different versions of the parametric model in terms of beta (β) estimation, depending on the different important values of the covariates (for the technical aspects of how this is done, see Zeileis & Hornik, 2007). In other words, the presence of splits indicates that the parameters of the initial theory-driven definition are unstable and that the data are too heterogeneous to be explained by a single global model. The model does not describe the entire dataset.

Classification trees look for different patterns in the response variable based on the available covariates. Since the sample is divided into rectangular partitions defined by the values of the covariates and since the same covariate can be selected for several partitions, classification trees can also evaluate complex interactions, non-linear and non-monotonic patterns. Furthermore, the structure of the underlying data generation process is not specified in advance but is determined in an entirely data-driven way. These are the key distinctions between classification and regression trees and classical regression models.

Model-based recursive partitioning was developed as an advancement of classification and regression trees. Both methods originate from machine learning, which is influenced by both statistics and computer science. Classification and regression trees are purely data-driven and exploratory—and thus mark the complete opposite of the model specification theory approach prevalent in the empirical social sciences. However, the advanced model-based recursive partitioning method combines the advantages of both approaches: at first, a parametric model is formulated to represent a theory-driven research hypothesis. Then this parametric model is handed over to the model-based recursive partitioning algorithm, which checks whether other relevant covariates have been omitted that would alter the model parameters of interest.

Technically, the tree structure obtained from the classification and regression trees remains the same for model-based recursive partitioning. However, the application of model-based recursive partitioning offers new impulses for research in the social, educational, and behavioural sciences. For the interpretation of model-based recursive partitioning, we would like to emphasize the connection to the principle of parsimony: following the fundamental research paradigm that theories developed in the social sciences must produce falsifiable hypotheses, these are translated into statistical models. The aim of model building is thus to simplify complex reality. What is the advantage of having such information? The answer to this question relates to the initial distinction that was introduced about the two modelling cultures. In the predominant (in the social sciences) data modelling culture, comparing different models has always been complex and problematic. The hybrid approach of model-based recursive partitioning modelling can help review models that work for the whole dataset and do not neglect such information that imposes on the models as 'global' straitjackets. Furthermore, suppose the researcher in question values the 'Ockham's razor' rule (that a model should not be more complex than necessary but must be complex enough to describe the empirical data). In that case, model-based recursive partitioning can be used to evaluate different models.

Another valuable piece of information generated by this approach is that the recursive model-based method allows for identifying particular segments of the

sample under investigation that might merit further investigation. That is, the possibility of identifying segments of our sample (and, therefore, presumably segments of the population if our sample is representative) that have a different version of the general theoretical model we have employed, in the form of statistical regression, to explain a given phenomenon Y. This possibility of identifying 'local' models of the population is not just a matter of chance. When applied to independent variables involving the measurement of attitudes and preferences, this possibility of identifying' local' models as defined above allows us to identify subgroups characterized by a particular cognitive pattern shared by that group. Such a group could very well be transversal to traditional sociodemographic categories (the young, the old, the middle class, etc.). Applied to experiments, it represents an advanced form of heterogeneity of treatment effects analysis that, with sufficient cases, can be very informative about the presence of general and local effects of a treatment.

This approach is very promising but has a 'cost' in methodological terms. To work well, it needs large samples and, even better, samples collected in several countries. Only with a sufficient number of cases, we can identify noteworthy subgroups. In contrast, if we have a few hundred cases, we cannot be sure of the statistical validity of the partitioning, besides the fact that we are talking about subgroups consisting of a few tens of cases are uninteresting as results.

This brief overview of model-based recursive partitioning illustrates the general point discussed in the previous sections: the complexity, quantity, and availability of digital data have highlighted the need to use analytical approaches other than those considered conventional in the social sciences. Therefore, Computational Social Science is, among other things, an attempt to adapt these new computational techniques and their associated 'modelling culture' to the research goals and questions of social scientists (Veltri, 2017). In other words, it is not only a matter of having more data of different types, which is important but also of innovating modelling techniques that can bring about transformative changes in the social sciences. Of course, there will also be methodological problems. Still, the ability to answer old questions with alternative approaches and ask new questions is the most attractive feature of Computational Social Science.

8.6 Conclusions

The possibilities offered by the new turn of digital and Computational Social Science can improve our understanding of human behaviour as never before. We move from data scarcity and local studies to potential large-scale, complex, and international ones. The implications for policymakers of this shift are the possibility of having behavioural insights both across different societies and better understanding and capturing within a country heterogeneity. In other words, large-scale online experiments combined with computational methods like the one discussed do allow for unprecedented

cognitive and behavioural based segmentation (see recent example Steinert et al., 2022).

Consequently, such differences can be used to differentiate the population to identify subsets of people, each characterized by a particular cognitive style. Segmentation is usually associated with profiling—the description of the relevant characteristics of the identified segment—sociodemographic characteristics, occupational status, geographical and spatial location, health status, attitudes towards essential aspects of the public policy in question. It is clear that cognitive and cultural segmentation also interacts with classical forms of classification resulting from affiliations such as occupations, generations, social classes, and status groups. Still, it cannot be taken for granted that they coincide. An example of such cultural segmentation is the analysis of the Brexit vote in the UK and how different cognitive-cultural styles are predictive of that vote (Veltri et al., 2019).

Behavioural segmentation is a potential tool for policy development. It is particularly suitable for the ex ante phase because it refers to a segmentation strategy of the target population and during the monitoring of the intervention *in progress* because it allows identifying the mismatch between the policy objectives and the citizen's interpretation of the policy. Similarly, cognitive-behavioural segmentation helps both the effectiveness and efficiency of policy interventions. In the first case, it helps to tailor instruments to the cognitive and cultural variability within the target population. An analogy here is precision medicine, an emerging approach for the treatment and prevention of diseases that considers individual variability in genes, environment, and lifestyle. In the context of public policy, the unit cannot be the individual but subgroups of the target population that will respond differently to the same public policy intervention. Thus, cognitive and behavioural segmentation plays an important role in improving efficiency. It can warn against implementing policy interventions that are likely to be ineffective with specific subgroups and thus help develop solutions that take cognitive and behavioural specificities into account. The other great opportunity comes from the use of digital traces and unstructured data. The sheer amount of this type of data provides insights into people's behaviour. However, because we are repurposing existing data collected for other purposes, some challenges are present. The first is entirely methodological: the criterion validity of these data types is still unclear (McDonald, 2005). The second concerns the ethical and privacy dimension of covert research, meaning that people are not often fully aware of the extend of their digital traces and how third-parties use them. Computational Social Science is no longer a complementary addition to or an embellishment in the social scientific study of society. Instead, it is changing the nature of social research because the digital has changed our societies. This is the starting point, we believe, that should accompany social scientists from now on.

References

Achen, C. H., & Bartels, L. M. (2017). *Democracy for realists: Why elections do not produce responsive government*. Princeton University Press. https://doi.org/10.2307/j.ctvc7770q

Berk, R. A. (2006). An introduction to ensemble methods for data analysis. *Sociological Methods & Research, 34*(3), 263–295.

Birnbaum, M. H. (2004). Human research and data collection via the internet. *Annual Review of Psychology, 55*(1), 803–832.

De Neys, W. (2018). *Dual process theory 2.0* (1st ed.). Routledge.

Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science, 8*(3), 223–241. https://doi.org/10.1177/1745691612460685

Groves, R. M., & Heeringa, S. G. (2006). Responsive design for household surveys: Tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 169*(3), 439–457. https://doi.org/10.1111/j.1467-985X.2006.00423.x

Groves, R. M., Singer, E., & Corning, A. (2000). Leverage-saliency theory of survey participation: Description and an illustration. *The Public Opinion Quarterly, 64*(3), 299–308.

Hand, D. J., & Vinciotti, V. (2003). Local versus global models for classification problems. *The American Statistician, 57*(2), 124–131.

Kahneman, D. (2012). *Thinking, fast and slow*. Penguin Books.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. *Heuristics and Biases: The Psychology of Intuitive Judgment, 49*, 81.

Lichtenstein, S., & Slovic, P. (Eds.). (2006). *The construction of preference*. Cambridge University Press. https://doi.org/10.1017/CBO9780511618031

Lizardo, O., Mowry, R., Sepulvado, B., Stoltz, D. S., Taylor, M. A., Van Ness, J., & Wood, M. (2016). What are dual process models? Implications for cultural analysis in sociology. *Sociological Theory, 34*(4), 287–310.

McDonald, M. P. (2005). Validity, data sources. In K. Kempf-Leonard (Ed.), *Encyclopedia of social measurement* (pp. 939–948). Elsevier. https://doi.org/10.1016/B0-12-369398-5/00046-3

Metcalfe, J., & Mischel, W. (1999). A hot/cool-system analysis of delay of gratification: Dynamics of willpower. *Psychological Review, 106*(1), 3–19. https://doi.org/10.1037/0033-295X.106.1.3

Moore, R. (2017). Fast or slow: Sociological implications of measuring dual-process cognition. *Sociological Science, 4*, 196–223. https://doi.org/10.15195/v4.a9

Mutz, D. C. (2011). *Population-based survey experiments*. Princeton University Press.

Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology, 15*(3), 263–280. https://doi.org/10.1002/ejsp.2420150303

Reips, U. D. (2002). Standards for internet-based experimenting. *Experimental Psychology (Formerly Zeitschrift Für Experimentelle Psychologie), 49*(4), 243–256.

Salganik, M. J. (2018). *Bit by bit: Social research in the digital age*. Princeton University Press.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*(1), 3–22. https://doi.org/10.1037/0033-2909.119.1.3

Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review, 4*(2), 108–131. https://doi.org/10.1207/S15327957PSPR0402_01

Spottswood, E. L., & Hancock, J. T. (2016). The positivity bias and prosocial deception on facebook. *Computers in Human Behavior, 65*, 252–259. https://doi.org/10.1016/j.chb.2016.08.019

Steinert, J. I., Sternberg, H., Prince, H., Fasolo, B., Galizzi, M. M., Büthe, T., & Veltri, G. A. (2022). COVID-19 vaccine hesitancy in eight European countries: Prevalence, determinants, and heterogeneity. *Science Advances, 8*(17), eabm 9825. https://doi.org/10.1126/sciadv.abm9825

Veltri, G. A. (2017). Big data is not only about data: The two cultures of modelling. *Big Data & Society, 4*(1), 205395171770399. https://doi.org/10.1177/2053951717703997

Veltri, G. A. (2020). *Digital social research*. Polity Press.

Veltri, G. A., Lupiáñez-Villanueva, F., Folkvord, F., Theben, A., & Gaskell, G. (2020). The impact of online platform transparency of information on consumer's choices. *Behavioural Public Policy*, 1–28. https://doi.org/10.1017/bpp.2020.11

Veltri, G. A., Redd, R., Mannarini, T., & Salvatore, S. (2019). The identity of Brexit: A cultural psychology analysis. *Journal of Community & Applied Social Psychology, 29*(1), 18–31. https://doi.org/10.1002/casp.2378

Wagner, J. (2010). The fraction of missing information as a tool for monitoring the quality of survey data. *Public Opinion Quarterly, 74*(2), 223–243. https://doi.org/10.1093/poq/nfq007

Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1966). *Unobtrusive measures: Nonreactive research in the social sciences.* (pp. xii, 225). Rand Mcnally.

Zeileis, A., & Hornik, K. (2007). Generalized M-fluctuation tests for parameter instability. *Statistica Neerlandica, 61*(4), 488–508. https://doi.org/10.1111/j.1467-9574.2007.00371.x

Zeileis, A., Hothorn, T., & Hornik, K. (2008). Model-based recursive partitioning. *Journal of Computational and Graphical Statistics, 17*(2), 492–514.