Theory and Ontology in Building Cumulative Behavioural Science

Janna Hastings[1,*], Susan Michie[1] and Marie Johnston[2]

Author affiliations

[1] Department of Clinical, Educational and Health Psychology, and Centre for Behaviour Change, University College London, 1-19 Torrington Place, London WC1E 7HB

[2] University of Aberdeen; Aberdeen Health Psychology Group, Institute of Applied Health Sciences, College of Life Sciences and Medicine, 2nd floor, Health Sciences Building, Foresterhill, Aberdeen, AB25 2ZD

Correspondence

* Correspondence to Janna Hastings, email: j.hastings@ucl.ac.uk

Abstract

The robust use of theory as a driver for research and evidence synthesis has the potential to mitigate the reproducibility crisis and contribute to the accumulation of knowledge and progress in the field of behavioural science. However, agreement on a single theory or theoretical framework is highly unlikely and arguably undesirable. We suggest that an alternative approach is grounded in the use of ontologies: formal computational structures for clearly defining entities and relations that enable theoretical integration via a network of relations between entities described in different theories. Ontologies are already widely adopted in the life sciences, but as yet have seen little adoption in the behavioural sciences. They have the potential both for comparison between theories, and for aggregation and evidence synthesis regardless of the theoretical framework that led to the generation of findings, leading to genuine cumulative progress.

This article is an expanded version of a Correspondence submitted to Nature Human Behaviour.

Introduction

In a recent Perspective article (Muthukrishna & Henrich, 2019), Muthukrishna and Henrich (hereafter, MH) argue that an important, and thus far largely overlooked, driver for the replication crisis in the social and behavioural sciences is 'the lack of a cumulative theoretical framework or frameworks'. Understood broadly, we are in agreement with this point, and indeed some of us have written at length about the importance of theory for human behaviour research previously (e.g. Davis et al., 2015; West et al., 2019). In particular, we agree with MH that theories enable cumulative science by coordinating evidence, providing a rationale for predictions and giving a basis for interpreting new findings. Theories are essential to enable different findings to be aggregated and compared, and thereby for a body of agreed knowledge to build up in the domain that goes beyond isolated high-impact findings that are not mutually comparable.

However, it will be difficult for researchers across the behavioural sciences to agree on any one specific overarching theory, as theories inevitably vary in their perspective and scope, depending on their purpose. A plurality of theories is characteristic of fields such as behavioural science that bring together thinking and practices from a range of disciplines and domains of research. Moreover, theoretical innovation is often one of the drivers for scientific progress. Thus, in the domain of human behaviour, it is very likely that multiple theories (sets of specific propositions hypothesised to be true), which may differ in scope or focus, will continue to exist. Insofar as they contain conflicting statements, these can be compared by reference to empirical findings, and may even lead to new investigations and evidence. The challenge, as we see it, is not to identify a single specific theory (or theoretical framework) that can unify all research, but rather how best to integrate findings arising from different theoretical approaches in order to develop as comprehensive a view as possible about what is known.

Here, we propose that what is needed to further progress in the behavioural sciences is a framework that is able to integrate and synthesise between results arising from different theoretical perspectives in order to aggregate evidence across different scientific perspectives, including psychological, behavioural and those of other sciences. First, we examine how and why an overarching framework might be successful in enhancing replicability. Then, we introduce the idea of an ontology as a type of framework that is able to integrate and connect between different theories. We show

how ontologies are able to integrate across both theories and evidence, and contrast this approach to that of a single theoretical framework as suggested by MH.

The role of theoretical integration in facilitating progress and overcoming the replication crisis in behavioural science.

The replication crisis is a well-known methodological challenge facing scientific research: the results of many scientific studies have proven difficult to replicate on subsequent investigation. It affects multiple fields, including behavioural science (Camerer et al., 2018), psychology (Open Science Collaboration, 2015), biomedicine (Ioannidis, 2005, 2011), and neuroscience (Poldrack et al., 2017). Variation in the effects of interventions that are reportedly the same can in some cases be an important source of information that advances understanding, as the context tends to vary from one intervention to another, and this contextual variation can be the reason for observed differences in effect. However, if the variation in effect reflects unreported variation in methods – or questionable research practices such as hypothesising after the results are known, amplifying the risk of false-positive findings – then it indeed poses a problem for scientific progress, referred to as the problem of reproducibility (Wellcome, 2015). It is widely recognised that it will be necessary to harness a plurality of different strategies to mitigate this challenge (Munafò et al., 2017), including improved statistical and methodological procedures, mandatory replication of novel findings, shifting incentives and practices in scientific research, and appropriate use of theory.

Theory has been defined by a cross-disciplinary consensus as "a set of concepts and/or statements which specify how phenomena relate to each other, providing an organising description of a system that accounts for what is known, and explains and predicts phenomena" (Davis et al., 2015). A theory should account for what is known, and be able to explain and predict phenomena on the basis of the specification of how phenomena relate to each other.

Theories enhance reproducibility and replicability in a number of ways. First, by offering definitions of the concepts or entities that the theory addresses, investigation of these entities can become more structured and targeted than investigations that lack any theoretical underpinning. Second, by specifying how these entities relate to each other, predictions can be made which influence the design of studies to test, confirm or falsify the prediction. So, if the theory predicts that X influences Y, observational studies seek evidence that X precedes Y rather than vice versa, and experimental studies manipulate X and examine effects on Y. By increasing agreement about the entities and predicted relationships, theory enables a much larger body of evidence to be accumulated, and evidence to be synthesised across studies. This produces findings at the scale and generality necessary to advance science and its applicability to real world problems (Freedland 2019). The more diverse the nature of the empirical verification that supports the same theoretical conclusion, the more confident we can be that it is true, which is another important aspect of addressing the replication crisis (e.g. Munafò, 2018).

One of the challenges in behavioural science is not in fact the lack of theory – as MH imply in their discussion of psychological textbooks – but rather a lack of formal specification of the different theories that have been proposed that would allow comparisons between theories, and synthesis of the available evidence in a way that allows comparison of which theories are better supported by the evidence, thus leading towards cumulative progress. There is a lack of a comprehensive database linking evidence to theory in a way that would allow determination of which theoretical propositions are better supported by the available evidence, and which are not as well supported by evidence, across the multiple domains, fields or disciplines from which evidence may arise. In order for a synthesis of this type to occur, behavioural science needs to clearly define its entities and their relationships (Michie et al., 2017).

MH argue that what is needed to address the replication crisis is an 'overarching integrative theoretical framework' for the field of human behaviour. It is worth elucidating what they appear to mean by this. MH define theoretical frameworks as 'a broad body of connected theories', although they do not say in what way the theories are to be connected, nor do they offer a definition for 'theory'. Given the multitude of behavioural science theories, the nature of the integrative framework may critically determine what is able to be included, and resulting exclusions will simply continue the pattern of multiple theories. In particular, their desideratum for this overarching integrative theoretical framework is too restrictive for general applicability to a wide range of theories in behavioural science. They express their desideratum as follows: 'a general theory of human behaviour would be evolutionarily plausible… utilize formal models, and provide us with an ultimate framework that delivers proximate predictions'. This desideratum has not been generated by a multidisciplinary consensus and both explicitly and implicitly points towards a particular theoretical framework – dual inheritance theory. The suggestion that the framework should be 'evolutionarily plausible' points to dual inheritance theory explicitly because using evolutionary theory to explain behaviour is the foundation of the proposed 'dual inheritance' theory. The suggestion to use 'formal models,' which elsewhere they define as mathematical models, implicitly points to the mathematical models commonly used in evolutionary theory as it is applied to behavioural research. And their definition of 'ultimate' frameworks, as those that provide 'ultimate' explanations, i.e. those that can explain 'why the phenomenon exists in the first place,' strictly constrains the scope such that only evolutionary frameworks can be ultimate. They illustrate this with the example of sex: we enjoy sex not because it is pleasurable nor because of the release of neurochemicals (denigrated as merely 'proximate'), but because of the 'ultimate' explanation that those of our ancestors who had more sex also had more children. This example is extended to the case of religions: the fact that religions tend to be pro-fertility is explained in terms of survival of the religion through procreation. By this explication of what an 'ultimate framework' is, only evolutionary frameworks can be ultimate.

But what type of theoretical framework would instead be fit for the purpose of integrating across different theories and findings in behavioural science? Each theory in behavioural science addresses its own set of entities and specifies relationships between them. As a result it is difficult to integrate findings into an overarching framework able to integrate across the different theories, unless there is agreement about the entities to be included. In addition, it is often unclear how a theory in one discipline e.g. behavioural science, relates to a theory in other disciplines. The 'dual inheritance' theory, for example, includes entities such as genes that are essential to other sciences such as those dealing with sub-cellular biological processes, but it is unclear how these entities as defined in the dual inheritance theory would relate to how entities such as genes are defined in theories in these other disciplines, which in turn are having their own theoretical debates (e.g. Gerstein et al., 2007). On the other hand, it would be quite feasible to make explicit connections and comparisons via the entities involved, insofar as the same entities – 'genes' in this example – are referred to in both theories.

Ontology as an integrative framework

Theories have a specific scope and subject matter. Moreover, tenets within theories may be in agreement or in opposition to one another. Theories, either implicitly or explicitly, consist of claims about the nature of the entities that are taken to exist in the world, as well as claims about how those entities are related. Claims about the nature of the entities that exist in the world are, in a philosophical sense, *ontological*. Ontology in this sense is the study of the basic nature and classification of the entities that exist. However, ontologies in recent years have come to have an incarnation outside of philosophy: as structured, computational representations of the entities and relationships that form the subject matter of a given domain (Hastings, 2017; Smith, 2003).

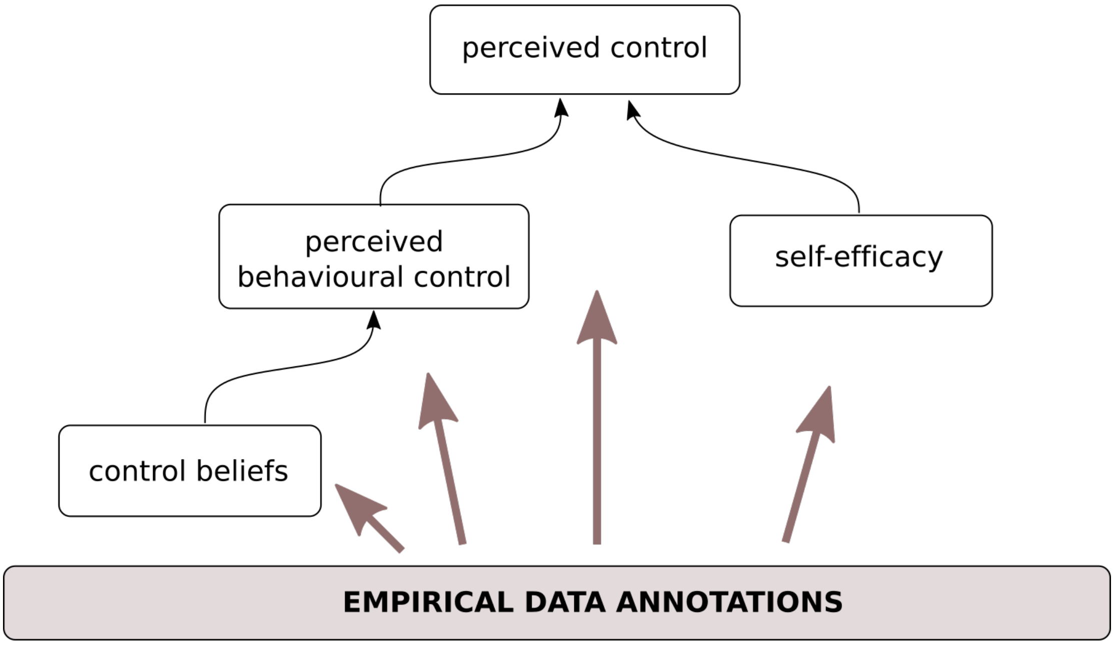

Figure 1: Illustration of an ontology representing entities from different theories

As computable representations of knowledge and part of the "data science" family of semantic technologies, ontologies serve multiple purposes: they enable sophisticated computational applications, and they serve as hubs around which evidence can be aggregated and theoretical debates can be resolved. In this sense, an ontology transcends individual theories by offering a framework for structuring the different entities that different theories take as their subject matter. For example, an entity such as 'perceived control' might encompass the entities 'perceived behavioural control' (from the Theory of Planned Behaviour, Ajzen, 1991), 'self-efficacy' (from Social

Cognitive Theory, Bandura, 1986) and ‘control representation’ (from the Common Sense Model, Leventhal, 1970). Each of these theory-specific terms, or entities, is in turn then related to the other entities in its respective parent theory. Explicitly listing entities from across different theories enables those theories to be connected explicitly, as the entities that they define or describe are connected.

Theories entail commitments to the existence of various entities: a theory which explains human behaviour in terms of social norms has a commitment to the existence of such a thing as a social norm, a human and human behaviour. A theory about electron orbital arrangements in atoms has a commitment to the existence of atoms and electrons. It is seldom made explicit what the commitments of a particular theory are, and it can be complex to elucidate these in full in particular cases, but nevertheless they are there.

We propose that the nature of the overarching integration effort that is needed in order to unify behavioural research is an ontological effort that is mindful of the commitments of distinct theories in the following way.

Two theories are only *comparable* – and may therefore be congruent or contradictory – if they are *about the same entities*. Thus, the Rutherford-Bohr model of the atom was comparable to the Thomson model of the atom because they were both models of the atom; the Rutherford-Bohr model was a better model because it offered a better agreement with the empirical evidence. In the field of behaviour, an important element of working towards theoretical integration is identifying the entities that the different theories are committed to, so as to be able to determine when different theories are addressing the same, overlapping, distinct – or poorly specified entities. A project that aims to achieve this objective has already been initiated (West et al., 2019), which thus far has catalogued the entities and relationships of 70+ different theories in behavioural science using a formal ontology-based modelling system. This formal representation of the content of theories allows theories to be automatically integrated and compared, as well as associated with evidence in a consistent, systematic fashion.

Comparison of our approach with the MH proposal of integration based on a single theory

MH propose that the role of an overarching integrative theoretical framework is to allow researchers to derive specific predictions from more general premises, and in the absence of such frameworks, results 'are neither expected nor unexpected based on how they fit into the general theory' and moreover 'have no implications for what we expect in other domains'. In support of this being the situation in psychology, they claim that textbooks in psychology are a 'potpourri of disconnected empirical findings' while outside of psychology, textbooks are theoretically structured and tell scientists what to expect and what not to expect, showing 'the interconnections between theories'.

In a theoretically integrated field, they say, 'each empirical result reverberates through the interconnected web of our understanding'. They give examples of what they say are fit for purpose overarching theories from physics and the natural sciences – Einstein's theory of special relativity, the periodic table in chemistry, and Darwin's theory of the evolution of species, each of which makes specific predictions that can be tested, such as that nothing can travel faster than light. However, it is difficult to understand in what sense these examples are indeed overarching theories. They are theories that have very specific domains and scopes – special relativity relating to the mass and movement of objects in space, the periodic table relating to the composition and structure of fundamental types of matter. Neither special relativity nor the periodic table has any direct bearing on human behaviour. And their explanatory success relates only to the nature of the laws and regularities in the parts of nature that they are about.

MH discuss examples intended to illustrate how difficult it is to extrapolate from specific theoretical fragments in the discipline of human behaviour. 'Without an underlying theoretical framework from which to draw hypotheses and tune our intuitions, it is difficult to distinguish results that are unusual and interesting from results that are unusual and probably wrong.' The theoretical framework that they offer to fill this gap, 'dual inheritance theory', is a variant of cultural inheritance theory, an application of evolutionary theory to learning in order to explain behaviour in terms of both genetic and cultural inheritance. Dual inheritance theory plays on ambiguity in what 'inheritance' means – the sense in which things are 'inherited' is very different in the cultural domain to the genetic one – and has been critiqued elsewhere (e.g. Fracchia & Lewontin, 1999). MH describe dual inheritance theory as 'capable of explaining the immense global psychological variation that's recently been documented'. However, the examples they give meet their criterion for an overarching integrative framework of being ‘evolutionarily plausible,’ but not of delivering ‘proximate predictions’ in advance of the evidence.

In contrast, the integrative approach we propose, based on ontologies, does not depend on a specific theory. It requires theory authors to become more explicit about the tenets of their theories and to define the entities and relations within those theories in a way that allows direct comparisons between theories. Furthermore, it provides a direct link between theories and evidence, in the process of ontology annotations that connect entities to evidence regardless of the theoretical background that led to the generation of the evidence.

Conclusion

We agree with the potential of an overarching framework to advance behavioural science. We argue that this is likely to increase replicability by increasing clarity about the entities investigated and ensuring that hypothesised relationships specified within theory relate to the same entities. A shared ontological framework at the level of entities and their relationships is likely to be more achievable than any specific theory as the latter may be challenged and may isolate behavioural sciences from other sciences. The resulting overarching network of entities and their relationships can facilitate the development of theory and generation of evidence that communicates effectively with theorising and evidence in other domains of science.

References

Ajzen, I. (1991) “The theory of planned behaviour”. *Organisational Behavior and Human Decision Processes.* 50 (2): 179-211.

Bandura, A. (1986) Social foundations of thought and action: a social cognitive theory. Englewood Cliffs, N.J.: Prentice-Hall.

Camerer, C. F.*;* Dreber, A.; et al. (27 August 2018). “Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015”. *Nature Human Behaviour*. **2** (9): 637–644.

Davis, R., Campbell, R., Hildon, Z., Hobbs, L. and Michie, S. “Theories of behaviour and behaviour change across the social and behavioural sciences: a scoping review”. *Health Psychology Review*. 2015;9(3):323-44.

Fracchia, J. and Lewontin, R.C. (1999). "Does culture evolve?". *History and Theory*. **38** (4): 52–78.

Freedland, K.E. (2019). “The Behavioral Medicine Research Council: Its origins, mission, and methods”. *Health Psychology*, 38(4), 277.

Gerstein, M.B., Bruce, C., Rozowsky, J.S., Zheng, D., Du, J., Korbel, J.O., Emanuelsson, O., Zhang, Z. D., Weissman, S. and Snyder, M. (2007) “What is a gene, post-ENCODE? History and updated definition.” *Genome Research*, 17: 669-681

Hastings, J. (2017) “Primer on ontologies”. *Methods in Molecular Biology* 1446:3-13.

Ioannidis, J.P.A. (2005). “Why most published research findings are false”. *PLoS Medicine* 2 (8): e124.

Ioannidis, J.P.A. (2011). “An epidemic of false claims: Competition and conflicts of interest distort too many medical findings”. *Scientific American, 304,* 16.

Leventhal, H. (1970) “Findings and theory in the study of fear communications” *Advances in Experimental Social Psychology* 5:119-186.

Michie, S., Thomas, J., Johnston, M., Aonghusa, P.M., Shawe-Taylor, J., Kelly, M.P., Deleris, L.A., Finnerty, A.N., Marques, M.M., Norris, E., O’Mara-Eves, A., and West, R. (2017) “The Human Behaviour-Change Project: harnessing the power of artificial intelligence and machine learning for evidence synthesis and interpretation.” *Implementation Science* 12:121.

Muthukrishna, M. and Henrich, J. (2019). “A problem in theory”. *Nature Human Behaviour*. 3, 221-229.

Munafò, M.R. and Smith, G.D. (2018). "Robust research needs many lines of evidence". *Nature*. **553** (7689): 399–401.

Munafò, M.R., Nosek, B.A., Bishop, D.V.M., Button, K.S., Chambers, C.D., Percie du Sert, N., Simonsohn, U., Wagenmakers, E.-J., Ware, J.J., and Ioannidis, J.P.A. (2017) “A manifesto for reproducible science”, *Nature Human Behaviour* 1:0021.

Open Science Collaboration (2015). “Estimating the reproducibility of psychological science”. *Science* 349 (6251), aac4716.

Poldrack, R.A., Baker C.I., Durnez, J., Gorgolewski, K.J., Matthews, P.M., Munafò, M.R., Nichols, T.E., Poline J.B. Vul, E. and Yarkoni, T. (2017) “Scanning the horizon: towards transparent and reproducible neuroimaging research”. *Nature Reviews Neuroscience* 18 (2): 115-126.

Smith, B. (2003) “Ontology” In Luciano Floridi (ed.), *The Blackwell Guide to the Philosophy of Computing and Information*. Oxford: Blackwell. 153-166.

West, R., Godhino, C.A., Bohlen, L.C., Carey, R.N., Hastings, J., Lefevre, C.E. and Michie, S. (2019), “Development of a formal system for representing behaviour-change theories”. *Nature Human Behaviour.*

Wellcome Trust, MRC, BBSRC and Academy of Medical Sciences Joint Symposium Report (2015), “Reproducibility and reliability of biomedical research: improving research practice.” Accessed online at https://acmedsci.ac.uk/file-download/38189-56531416e2949.pdf. Last accessed in June 2019.