PERSPECTIVE
https://doi.org/10.1038/s41562-020-00989-3

nature human behaviour

Personality beyond taxonomy

Julien Dubois[1]✉, Frederick Eberhardt[1], Lynn K. Paul[1] and Ralph Adolphs[1,2,3]✉

Human and animal behaviour exhibits complex but regular patterns over time, often referred to as expressions of personality. Yet it remains unclear what personality really is: is it just the behavioural patterns themselves, something in the brain, in the genes or perhaps all of these? Here we offer a set of causal hypotheses about the role of personality, integrating psychological and neuroscientific approaches to personality in a testable framework. These hypotheses clarify the causal and constitutive relations that personality has with genes, environment, brain, mind and behaviour, and we suggest specific experiments that can adjudicate amongst the different hypotheses. We focus on a set of models that propose that personality is instantiated in the brain, distally caused by genes and environment and, in turn, causing the overt behaviours from which it is often inferred. We argue that articulating and testing such models will be essential in a mature science of personality.

Personality generally refers to "relatively permanent traits […] that give both consistency and individuality to a person's behavior"[1]. Despite strong consensus on some of the necessary features of personality (Box 1), many cases quickly reveal deeper problems. Does a completely paralyzed person have a personality? If so, then overt behaviour cannot be constitutive of personality, which must instead reside elsewhere. Does being blind-folded change your personality? What about losing your memory? What about having a stroke that alters your motivation or emotions? As we think through these cases, we are quickly confronted with a first ambiguity: whether personality is literally defined by categorizing behaviour (much like a Linnaean taxonomy) or whether it is something causally antecedent that explains the behaviour. Here we will discuss models that argue that personality is more than taxonomy; namely, it is a feature of the brain that causes the regularities in complex behaviour (and that is typically inferred from the behaviour). While there may be general agreement with this view, there is a spectrum of logically distinct, empirically testable models that describe how such a concept of personality would be related to other psychological variables.

There are many different ways of describing and quantifying personality, with published frameworks spanning three[2], five[3–9], ten[10], sixteen[11], thirty[12] or even a hundred dimensions[13] (where each 'dimension' is a relatively distinct axis of variability in behavioural tendencies in the population). Studies in mice, using dense phenotyping, have yielded perhaps the most comprehensive data-driven inventory of personality traits to date in a species, suggesting four dimensions that concisely capture variability in behaviourally inferred personality traits in that species[14]. For the sake of illustrating general conceptual questions in this Perspective, we take the example of the 'Big Five', which is the most widely used framework applied to humans. The Big Five personality traits (openness to experience, conscientiousness, extraversion, agreeableness and neuroticism) describe personality on continuous dimensions that were initially derived from a corpus of words that people use to describe other people[11]. Even if descriptively accurate, this popular modern account of personality leaves the scientific role of personality completely opaque: is Big Five personality a summary statistic, like the gross domestic product, which provides a convenient shorthand for a complex system of interactions but has no real independent existence? Or is it more like El Niño, the macro-level climate phenomenon whose causal effects on the weather in different parts of the world can be characterized independently of a detailed understanding of the Pacific sea surface temperature (a view that may be similar to network theories in personality psychology[15])? In the former case, personality simply is a descriptive summary of a set of behaviours that someone displays in specific contexts; there is no further underlying reality to personality. In the latter case, the behavioural measures in the Big Five questionnaires (or any other behavioural measures used to infer the Big Five) are actually indicators of an underlying personality that causes the behaviours; in other words, the behavioural measures are not themselves constitutive of personality, but are used to infer personality.

It is important at the outset to clarify two important points about how behaviour relates to personality, to avoid any confusion. First, throughout this Perspective we will use 'behaviour' to refer to effects of the somatic or autonomic motor systems: that is, movements of or in the body due to muscles, blood vessels or glands. While broader than volitional actions, this usage of the term 'behaviour' is narrower than the way many personality psychologists use the term, which additionally includes psychological states like thinking (for example, ref. [15]). Thus, behavioural expressions of personality, according to our usage, could include speaking, moving your body or blushing, but do not include thinking, feeling, remembering or attending (if these are not accompanied by behaviours). The second important point to reiterate is that the aspect of behaviour relevant to personality is not a specific instance of behaviour, but rather relatively temporally stable, context-dependent patterns of behaviour over time.

In this Perspective we do not provide a review of the history or the variety of extant accounts of personality. Instead, we advocate here for a view that takes personality to be an objective scientific quantity whose causal effects can be tested and characterized, a view already stressed by Eysenck[16]. Such a view must explain the causal role that personality plays in the chain from genes to brain and to behaviour, and it must delineate where it is instantiated. We do not think that currently available evidence from the study of personality in humans and animals offers conclusive evidence for any one specific causal hypothesis on personality. Instead we offer a small set of experimentally distinguishable causal hypotheses that we believe can be tested in a combination of human and animal studies.

[1]Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA, USA. [2]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CaliCAfornia, USA. [3]Chen Neuroscience Institute, California Institute of Technology, Pasadena, CA, USA. ✉e-mail: jcrdubois@gmail.com; radolphs@hss.caltech.edu

Box 1 | Accepted and speculative features of personality

Allport was one of the first to enumerate a modern list of features that could be used to characterize a personality trait[67]. Other work has stressed features that any psychological variable should aim for, such as good inter-rater reliability, face validity and construct validity[68]. Most who work in the field implicitly endorse a set of features that may provide a working definition of personality, but there are also features on which there is disagreement. Here we provide our own provisional list of personality features, dividing them into those that are well accepted and those that are speculative.

Accepted features:

1. **Stability**. Stability over time refers to the fact that personality is closely related to our concept of an enduring personal identity. Under normal circumstances, personality does not change from one moment to the next, unlike emotions, thoughts and behaviours.
2. **Individual differences**. Personality forms part of the explanation of why each individual behaves somewhat differently in a given context, and it can serve to identify groups of individuals distinguished by different personality types[69]. Taken together, features 1 and 2 suggest one common metric for personality traits: they maximize between-individual variability but minimize within-individual variability.
3. **Multicomponent**. Personality involves regularities across emotion, cognition, mood and behaviour. This feature also serves to exclude behavioural regularities that are unrelated to personality, such as having a limp or tremor in one's movements.
4. **Subject to pathology**. Personality disorders can be diagnosed[70] and generally show personality structure quite similar to the norm.

Speculative features:

1. **Low-dimensional**. Most individual differences in personality can be described in a relatively low-dimensional space: as few as five personality dimensions capture most of the variance, according to the leading view in human personality research. This view would simplify the study of personality, but it may also omit meaningful variation.
2. **Evolutionary continuity**. Evolutionary precursors to some human personality traits may be found in other species[71]. We argue that the construct of personality applies to all animals with complex behaviour. This view unifies studies in animals and in humans and generates opportunities for experimental studies of personality that are impossible or unethical in humans.
3. **Independent of conscious experience**. Personality is often taken to include conscious experience. To the contrary, we hold that conscious experience is not a necessary part of personality. We thus take the question of whether a person or animal (or robot) is capable of having experience at all to be logically distinct from the question of whether it has personality. This simplifies the study of personality, especially in humans or animals who cannot report on their experience.
4. **Neurological**. This is the view we defend here: personality is constituted by features of the brain (see Fig. 1 and main text).

The causal role of personality

At the core of our proposal lies the distinction between causal and constitutive relations, which we already briefly alluded to in the introduction. This distinction is critical to formulate testable models of personality. If two variables are causally related, one can manipulate them independently; if their relation is constitutive, one cannot (since in that case they are different descriptions of the same thing). As our set of models below tries to capture, personality is obviously related to many other things: genes, environment, the brain, the mind and behaviour. Which of these relations are constitutive and which are causal?

It is useful also to make a brief comment on the mind–brain relation in general, which has a long history of debate: some people believe the relation is causal ('dualism'), whereas others (including us) think it is constitutive ('monism'). These commitments have consequences for how one thinks also about personality: if one believes (as we do) that personality is, first and foremost, a mental (psychological) variable, then monism would commit one to saying that personality is constituted by neurobiological events, whereas dualism would entail that it is caused by neurobiological events (depending on the exact variety of dualism). Here we adopt the monist position that if personality is a psychological (mental) variable of some sort, then it must be constituted by neurobiological events of some sort.

A causal description of any system requires a level of abstraction from the details to delineate quantities that stand in causal relations. These quantities are identified because they lend themselves particularly well as points of intervention where the system can be changed and manipulated or because they constitute bottlenecks in the processes that make up the system[17,18]. In the case of personality, the broad strokes of the underlying total causal system are relatively clear: genetics and environment are the distal causes of an individual's behaviour, and brain and body are the proximal causes. Of course, causes can have different strengths and show probabilistic outcomes (in which case the arrows in the models shown in Fig. 1 might have different weights).

Our review of possible causal models of personality starts from the most comprehensive view of personality, according to which personality is constituted by all the types of variables that are normally studied in relation to personality, from genes to minds to behaviour (Fig. 1a). One could imagine an argument for such a model that claims that, insofar as personality requires long-term patterns of behaviour over time, these behaviours are a direct product of the expression of genes over that long time period, or perhaps they emerge from gene expression during development. It is even conceivable that the right kind of gene therapy would change your personality. Eysenck himself may have held such an all-encompassing view (for example, Fig. 2 in ref. [16], although the meaning of the arrows in that figure is unclear). The main drawback of this model A, however, is its lack of specificity. Personality would be constituted by vastly different types of processes (gene expression, brain activity, behaviour, etc.) encompassing extremely complex causal interactions. There is little that is left 'outside' of personality, and it would look more like a summary statistic, like gross domestic product. We think such generality misses many of the important causes and effects of personality: changes in genes do not instantaneously change personality, but require a long causal chain from gene expression through brain changes. In contrast, personality can be changed quite directly without changing genes, for example, in the case of localized brain damage. Thus, it seems clear that while genes have causal relations to personality (arrows in all models in Fig. 1) they cannot be constitutive of personality (contra model A).

One could also imagine a proposal according to which personality is constituted only by the genes (and/or the environment), but we know of nobody who has ever held that view and so do not present

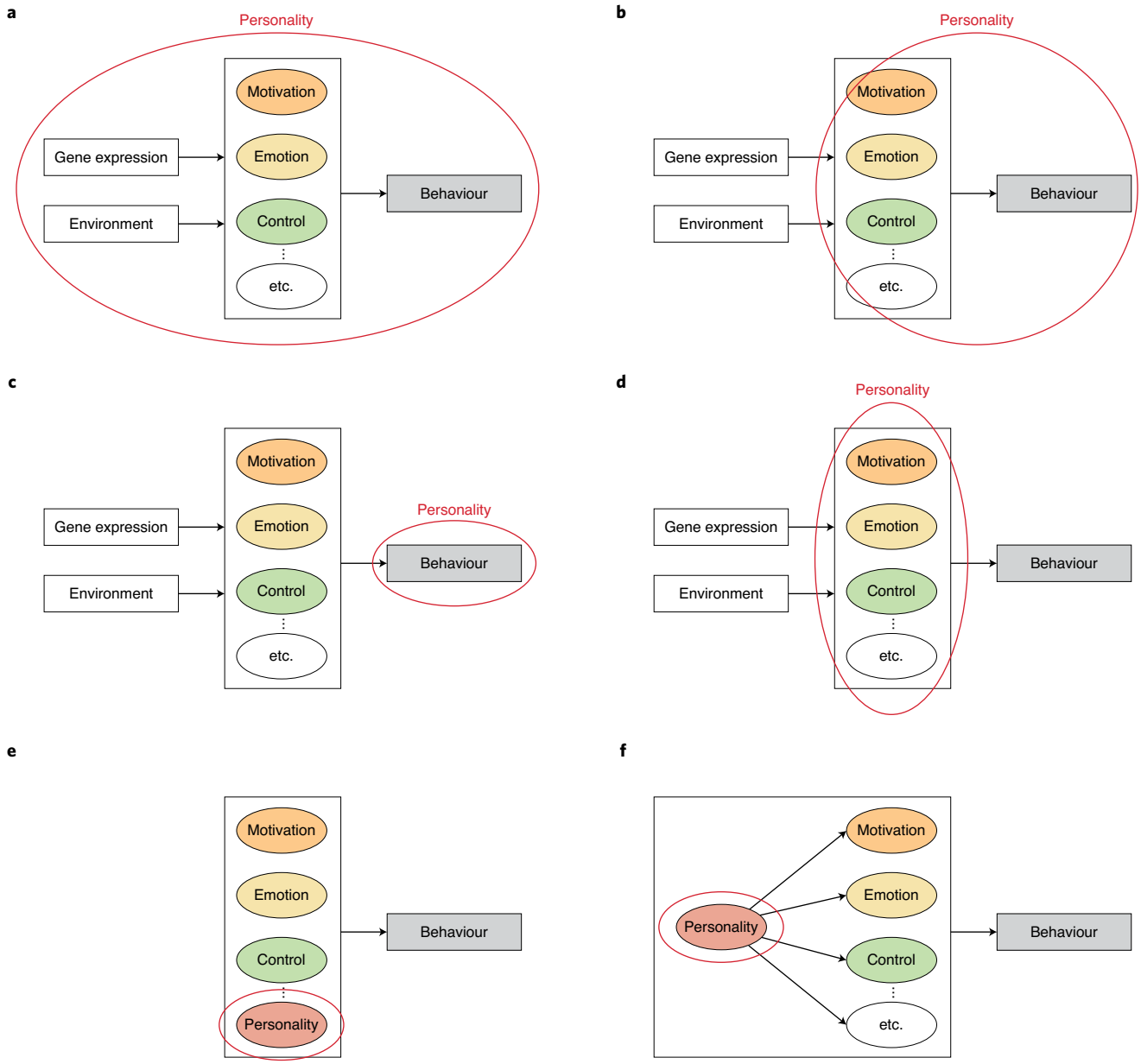

Fig. 1 | Causal models of personality. These schematics provide a survey of conceivable models that relate personality to other variables. The red circles denote constitutive relations; arrows denote causal relations. Only the simplest schemes are indicated. Any viable candidate model would need to be elaborated, possibly including additional variables of the types shown (for example, other psychological variables, like memory or attention), denser causal relations (for example, bidirectional ones) and causal relations with different weights or probabilities. **a**, The most comprehensive scheme, according to which personality is constituted by genes, environment, psychology and observed behavioural patterns. **b**, A view according to which personality is constituted by patterns of "broad behavior"[15], encompassing patterns in both psychological variables as well as in overt behaviour. **c**, A scheme according to which personality is just identified with patterns of observed behaviour, eliminating its causal role in behaviour. **d**, Personality as constituted by patterns across a variable set of psychological states (perhaps only a specific subset of them). **e**, Personality as a psychological variable distinct from others and not strongly causally influencing other psychological states. **f**, Personality as a psychological variable strongly causally influencing other psychological states and acting through them to produce patterns of behaviour. Our own view finds **d**–**f** the most interesting versions for further development. Genes and environment are not shown in **e** and **f** only for clarity.

it here. None of this is to deny the rich and interesting effects of genes and environment on personality—for instance, evidence that epigenetic effects such as DNA methylation can contribute to personality disorders[19]—but in all these cases we take it that the relation is causal (typically, during development) rather than constitutive. We take it that most psychologists do not think that personality is literally 'in the genes', partly or entirely, but instead that genes and environment are distal causes of personality, which is located somewhere more proximal to behavior[16].

This brings us to model B, which may be a common view in psychology, especially if 'behaviour' is defined broadly to encompass both overt behaviour and psychological variables (see our clarification above)[15,20]. Since we define behaviour more narrowly, as distinct from psychological variables, we argue that personality causes

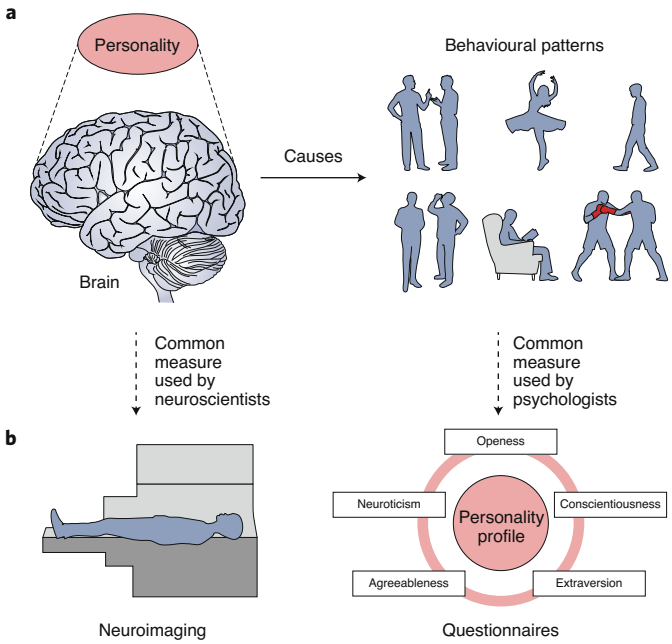

Fig. 2 | Challenges for personality neuroscience. a, In this Perspective we have delineated a science of personality where personality is a feature of the brain and a cause of behaviour. **b**, The nascent field of personality neuroscience[66] often uses metacognitive assessments of behaviour (questionnaires filled by participants or by informants) and measurements of brain activity (for example, resting-state functional MRI or task functional MRI). While some progress has been made using these proxies, measurements that better capture the processes of interest at both ends are actively being pursued (dense phenotyping of behaviour, for instance from video recording; and high-resolution recordings of electrophysiological activity, for instance from electrocorticograms).

behaviour (and thus, can be measured and inferred from behaviour), but is not constituted by behaviour. To make this case, we will next discuss a series of models that provide a more specific role for personality than model B and that make testable predictions about dissociations and interventions that can help to adjudicate between the different models. In Fig. 1, models C and D are refinements of model B, placing personality either in overt behaviour or in the mind, but not both. Model D in turn is further refined in models E and F, by specifying more precisely where in the mind to place personality in relation to other psychological variables.

Model C anchors one end of our spectrum of models by describing personality merely as patterns of overt behaviour (as in a Linnaean taxonomy). We do not attribute this extremely behaviourist view to anyone, but include it because it highlights a specific set of problems that show that overt behaviour must instead be a causal effect of personality. Model C does not provide the basis for a scientific account of pathologies or for a study of the relation between personality and behaviour. We take it that most readers would agree that a person without any patterns of behaviour (for example, a patient with locked-in syndrome) can nonetheless have a well-defined personality. Similarly, there would be agreement that an actor, even one who acts consistently over their entire life, can have a personality at odds with the behavioural evidence. These cases show us that we can be wrong about somebody's personality if all we have to go by is their behaviour: they may not be able to express their personality or they can fake it. Identifying personality directly with a feature of behaviour, no matter how abstract, confuses (one possible) measurement of personality with its metaphysics, i.e., with what personality actually is. Behaviour provides evidence of personality, and we indeed regularly infer personality from behaviour, but that is quite different from the claim that behaviour literally constitutes personality. One might still propose that personality is the disposition to behave in certain patterns. This, however, suggests that personality is constituted by something that has the causal power to bring about behaviour. In which case, personality is causally antecedent to behaviour, fitting with one of models D, E or F instead.

The second concern with a behaviour-based account of personality as shown in model C is methodological: like the view that conceives of general intelligence as the common factor resulting from a factor analysis of scores on a battery of cognitive tasks (Spearman's *g*)[21], the mere presence of shared positive or negative correlations among aspects of behaviour measured over time would necessarily result in a shared personality factor, yet not ensure any meaningful reality (for example, as an overarching common cause) to the resulting concept[22]. Like an account of biological species based solely on morphology, an account of personality based solely on patterns of behaviour would become a useful bookkeeping tool, but could not provide a basis for the scientific explanation of those behavioural regularities. This point raises the intriguing possibility that, once we know enough about what personality really is (namely, a certain feature of the brain), this may also permit a revision of what personality factors there should be, a point we return to at the end.

The above considerations already help to eliminate both genes and behaviour as constitutive of personality and suggest that personality is located in the brain. We consider this possibility in models D, E and F. All of these models propose that personality is 'in the mind', and hence 'in the brain', but differ in where exactly in the mind they put personality in relation to other psychological variables. As such, this class of models proposes particular mental architectures, and we take them up in more detail below.

Personality is in the brain

We posit that personality is a feature of the brain (Fig. 2a): it is not simply a summary of behaviour (contrary to Fig. 1c), nor partly residing in events that are causally antecedent to the brain (contrary to model A). While we take this fundamental position itself to be consonant with much of personality psychology[20,23], we now look to examine open scientific questions about exactly how personality causes behaviour and how personality is related to other psychological variables (like memory, attention, emotion, etc.). In particular, we have to address the question of how we can think of personality as playing a causal role when it is instantiated in brain processes that have their own low-level causal descriptions. And if personality is to play a scientific role as a causal variable with influences on behaviour, we have to explain why it is more than just a re-naming of the causal mediators between genetics and the environment on the one hand and behaviour on the other. In the following we outline three distinct causal hypotheses that relate personality, behaviour and other psychological constructs (Fig. 1d–f).

We begin by considering model F. Here personality corresponds to a specific feature of the brain that influences behaviour by affecting other psychological constructs (such as motivation[24], emotion[25], or control or regulation[26,27]; but many others are of course possible). Such a feature could be instantiated locally in the function of some specific circuit or neuroanatomical module of the brain, or it could be realized in a more global property of the brain's neural networks, such as the strength of connectivity or levels of particular neurotransmitters. In this view, the neural features of personality (whatever they may be) would influence behaviour only indirectly through other psychological constructs, such as attention or memory. Personality, while instantiated by features of the brain, would act as a high-level modulator of (some subset of) other psychological constructs, which in turn influence behaviour. Personality would be not merely an aggregate re-description of the other psychological constructs, but an independent feature of the brain. For analogy,

one could think of personality as the degree of connectivity in a power grid, which otherwise maintains the same power generation and storage capacities. The effect on the consumer is only indirectly through a more reliable or erratic energy supply.

Model F also corresponds best to how laypeople usually make sense of human behaviour. According to one common account, they do so by inferring states and traits. States are usually thought of as more transient and as causally more proximal to observed actions, while traits are more temporally durable and have causal effects on states. This tripartite view of the relations between traits, states and actions has them situated in a causal hierarchy, in that order[28]. In fact, there may be a range of psychological variables we regularly infer in order to make sense of observed behaviour, varying in degrees of abstraction from the actions they explain[29]. For instance, seeing somebody smile while helping another person might lead to the following sequence of psychological attributions, from states to traits: why are they smiling? Because they feel happy (state). Why? Because they are pleased at having helped another person. Why? Because they have a generous and empathic personality (trait). It remains an important empirical question whether an objective science of personality should also consider such a continuum between states and traits, or whether the distinction is more categorical.

The possibility that personality affects behaviour by acting through a set of other psychological variables, as suggested by model F, is an empirical question, and it needs to be tested against the plausibility of alternative models. So, model E considers the possibility that personality may reside at the same level as other psychological constructs and is thus able to influence behaviour directly. The key question distinguishing the role of personality in models E and F is the extent to which differing personalities can retain an influence on behaviour (and hence be detectable from differences in behaviour) for individuals who otherwise are indistinguishable in other psychological constructs. This is not a question that can be addressed by merely studying normal behaviour, but requires an independent understanding of personality as a brain feature whose pathways of influence on behaviour can then be studied as described below.

The two causal hypotheses expressed in the previous paragraph (models E and F) take personality to be at least partly distinct from other psychological constructs. But there is a third alternative: personality may not be distinct from all other psychological constructs, but constituted by (a subset of) them (model D). According to this view, psychological constructs such as emotion and motivation are not causally influenced by personality but, taken together, constitute personality. Personality then provides a higher-level description of the psychological constructs, instead of itself being some separate cause (or effect). The role of personality here can be thought of as analogous to the temperature of a gas: the temperature is a function of the kinetic energy of the particles it is made of, but we find the notion of temperature scientifically useful because for various physical effects it is sufficient to describe the cause in terms of the temperature, rather than in terms of the particle movements that define the temperature. In contrast to treating personality as a summary statistic of either broad behaviour (model B) or overt behaviour (model C), the proposal of model D ensures that the resulting notion of personality has clear causal effects on behaviour. The causal hypothesis in model D requires any meaningful variation in personality to be accompanied by variation in other psychological constructs (just as any change in temperature requires a change in the kinetic energy of at least some particle).

All three causal models (models D, E and F) treat personality as a causal variable instantiated in the brain, and they all benefit from a scientific study of personality that includes both neuroscience and psychology. Despite the simplified representation in Fig. 1, none of these models is intended to deny that there can be feedback from behaviour back to personality. In fact, an understanding of the causal pathways from behaviour back to personality (or the psychological constructs that constitute it) would be crucial to understanding and developing behavioural therapies for personality disorders; we take this to be the view also held by theories of psychotherapy[30]. The causal models pick out personality as a scientific variable in each case, because it is considered to be a relevant target for possible intervention. The nature and effect of interventions will be different for each of the models. Of course, one may be able to influence personality by changing the genes during development, but—according to models E and F—one could in principle also target personality more precisely by interventions on the particular brain features that constitute personality or by targeting, for example with therapy or training, those psychological constructs that constitute personality in model D. Indeed, for model D there may be a variety of different interventions on the underlying psychological constructs that would result in the same change in personality. Just like there are many ways of arranging the kinetic energy of the particles in a room to achieve an ambient temperature of 80 °F (27 °C), with regard to personality, there may be various patterns of change in emotions and cognitive control that result in equivalent personality changes.

Each of the three causal hypotheses about the role of personality can be investigated experimentally. For instance, for the neuroscientist, models E and F leave open the possibility that there may be distinct anatomical regions or patterns of brain activation that underlie personality; model E could be further distinguished from model F by finding people with lesions in whom personality is impaired but other cognitive or emotional functions are not. By contrast, model D would be supported by finding that all such individuals invariably show concomitant deficits in both personality and some other psychological functions (for example, cognitive control, emotion, etc.).

Testing the different causal models

The three causal models that we find the most plausible (models D, E and F) are difficult to distinguish solely with data from healthy humans, but may be examined more effectively in studies of human pathological conditions and of personality in animals. In humans, ever since the classic case of Phineas Gage, who sustained damage to the prefrontal cortex, it has been clear that focal brain lesions can alter personality[31,32]. More evidence has accumulated since Gage's case that lesions to ventromedial prefrontal cortex (vmPFC) can lead to dramatic personality changes[33,34]. For example, findings from nearly 200 people with lesions whose personality changes were rated by relatives suggest that pathologies of personality are more likely to result from damage to the prefrontal cortex than from damage to other brain regions[35]. Within this sample, there appear to be clear cases of personality changes following lesions (for example, of the vmPFC) that are typically not accompanied by corresponding impairments on other neurocognitive tests, suggesting that personality and cognition (for example, memory, perception, attention) are to some extent distinct (in line with Fig. 1e,f). Interventional tests of the models sketched in Fig. 1 can thus be revealed, at least to some degree, through neuropsychological dissociations in humans, long a workhorse for revealing cognitive architectures[36,37].

In contrast to studies with humans, research on animal personality affords greater experimental control, greater ability to measure physiological parameters, greater opportunities for naturalistic observations and, typically, greater potential for longitudinal studies. A cross-species integration of the study of personality will require a species-independent delineation of the concept of personality; we think there are relevant criteria for personality that permit such an integration, just as there are such criteria for other psychological constructs, for example, intelligence[38] (Box 1). Research on animal personality has relied on several sources of measures to infer personality[39]: explicit ratings of personality traits provided by observers familiar with the animal (trait rating[40,41]); detailed observations made by trained ethologists (behavioural coding;

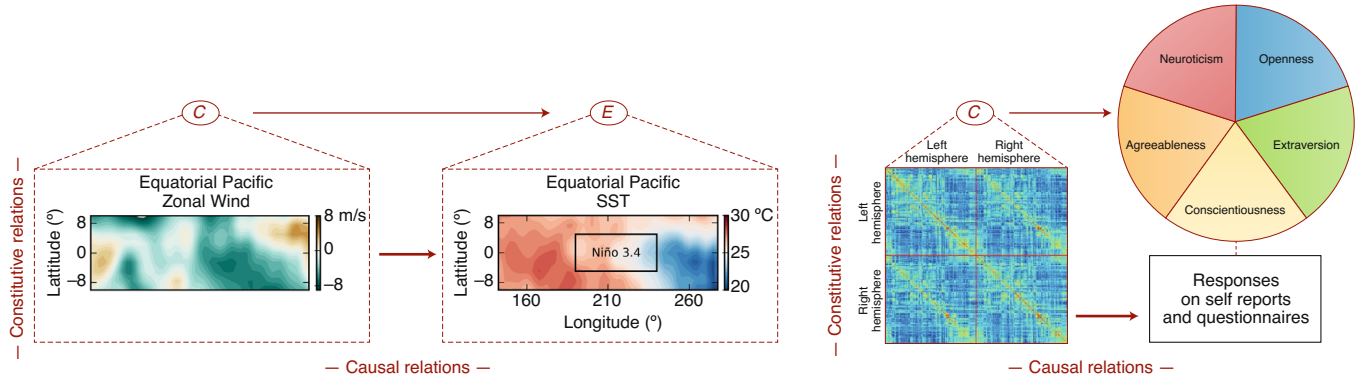

Fig. 3 | Causal feature learning (CFL). Left: CFL was applied to zonal wind and sea surface temperature maps over the equatorial Pacific to automatically identify macro-variables *C* (cause) and *E* (effect) that describe the causal relations between wind and temperature, respectively. One of the states of *E* had many features characteristic of El Niño, which is normally defined by a divergence of the average temperature within the rectangular geographical area marked by 'Niño 3.4'. In other words, CFL was able to identify the macro-level climate phenomenon of El Niño as constituted by particular micro-level sea surface temperature patterns and plausibly caused by specific macro-level wind patterns (westerly equatorial winds), without any input about when El Niño had occurred in the past or that the broad class of westerly equatorial winds is relevant. Right: proposed application of CFL to the study of personality, which could provide a revision of the presently used personality dimensions and the discovery of their neural causes. Each panel schematizes the vertical components as constitutively related and the horizontal components as causally related: thus, the macro-variable *C* is constituted by broad patterns of neural activity that are causes of the behavioral responses from which personality is often inferred, which is thus a discovered coarse representation of the measured behavioural responses. The correspondence between neural and behavioural aggregates discovered by CFL would provide a data driven characterization of personality.

for example, https://mousebehavior.org/); and, most recently, use of extensive video data in controlled environments (for example, in flies[42–44], mice[14,45] and cuttlefish[46]). These methods have largely produced findings that confirm that stable personality traits exist in animals and that animal personality traits have a dimensional structure showing some resemblance to what has been found in humans[47]. An exciting application for large-scale behavioural data is the further development of data-driven techniques that can help refine models of personality structure, a future direction (see below) that utilizes dense video data from animals as well as the corpus of human data now available through our use of social media[48,49].

In addition to empirical studies of human clinical populations and animals, personality neuroscience will continue to benefit from non-invasive neuroimaging in healthy individuals. Most neuroimaging research to date has attempted to find neural correlates of personality traits from leading psychometric models, usually derived from self-report questionnaires, such as the 60-item NEO Five-Factor Inventory (NEO-FFI[50,51]; Fig. 2b). While brain connectivity is presumed to be highly relevant to personality, other neural properties are candidates as well. It is well known that specific neurotransmitter receptor subtypes, or combinations of them, can produce fairly specific psychological and behavioural effects. For example, the neurotransmitters serotonin and dopamine, and their receptors, fundamental components of signalling between neurons, have been implicated in neurobiological theories of personality disorders[52], and there is even a 'neurotransmitter attributes questionnaire' of personality[53]. This raises the possibility of using psychopharmacological studies to gain insights about personality. Distinguishing between more complex versions of models D, E and F in Fig. 1d–f may require the specific manipulation of certain neurotransmitter systems, perhaps in specific brain regions. Just like brain lesions can give us some initial data by the pattern of dissociations that they produce (for example, impairing memory independently of personality), so too could targeted drug studies. Ultimately, knowledge about the specific neural features that comprise personality traits is not only of theoretical interest and useful for distinguishing the different models we outlined in Fig. 1, but may also provide critical information for clinical interventions (for example, what aspects of personality can be changed most effectively by which methods).

Future directions

As we move towards a more interdisciplinary science of personality, some specific challenges arise. It is unlikely that the personality traits from existing psychometric models will have a 1:1 relationship with brain features. One of the most interesting future directions is to update currently used personality frameworks (such as the Big Five) which have been described as "pragmatic abstractions with no special biological value"[54] and use measures from the brain and from behaviour concomitantly to revise personality space to more accurately match its neurological substrate. This is analogous to the recent use of data from functional neuroimaging studies to classify psychiatric disorders based on their similarity in patterns of brain activation[55]. More generally, large-sample neuroimaging and behavioural data are now becoming available in databases that provide an opportunity to use data-driven methods to discover or revise our models of psychological architecture in humans, including personality[56,57].

New techniques will be required to fully distinguish between causal and constitutive relationships, and it may be fruitful to look to other disciplines for analogies. For instance, causal feature learning (CFL; Fig. 3)[58] is a tool that has been successfully applied to climate data[59]. In this approach, micro-level causal features of one type (for example, wind speed measured on a grid over the ocean) are aggregated so as to predict causal features of another type (say, sea surface temperature, also measured on a geographical grid). Instead of clustering wind measurements and temperature measurements separately, the aggregation algorithm aggregates the relation between the two sets of micro-variables with the aim of respecting macro-level boundaries that make a causal difference. This approach has successfully identified El Niño as a macro-level climate phenomenon without including any information about past occurrences of El Niño in the aggregation procedure. It also yields results distinct from standard clustering algorithms.

Similarly, CFL could, in principle, be used to aggregate neurobiological and behavioural data in the same individuals to derive

macroscopic variables (for example, brain networks and behaviour patterns) that describe personality traits. To use such approaches, we would need large datasets with comprehensive personality assessments (ideally based on observed behaviour, as has been done for deriving other data-driven human behaviours such as posture[60]) and comprehensive neuropsychological profiling of individuals, to be able to build more complete models (including other psychological variables) that can begin to distinguish between the causal architectures that we discussed in the previous section (Fig. 1). Regardless of what approaches are used to derive new personality traits, the traits so derived will of course need to be tested with respect to their utility. Obtaining El Niño through CFL should have predictive value for its effects on rainfall, the economy, etc. The same applies to obtaining personality variables through CFL. These personality indices should allow for predictions of real-life outcomes, such as social network size, health or income (just like the current ones already do[61,62]).

Perhaps the main limitation for such a data-driven approach at present is the unavailability of dense, large-sample behavioural data. Data for inferring personality in humans comes either from rather brief self-report or informant-report questionnaires or from (also sparse) objective behavioural observations (as made by human raters or as quantified from video recordings, for instance). However, this situation is changing. There are now behavioural data, or proxies of them, that may indeed be quite dense. A person's year-long email or twitter behavior[63], their movements recorded with a webcam[64] and/or other technology-enabled observations, could begin to provide a much more comprehensive inventory of behavioural patterns for data-driven analyses. This approach has been successful in some animal studies; for instance, dense recording of mouse behaviours over long time epochs, in response to specific environmental contexts, has provided a derivation of mouse personality traits at a level that permits identification of one mouse from others based on context-dependent behaviors[14]. That study used as its mathematical definition of personality essentially the first two criteria that we list in Box 1: personality can be inferred from that aspect of behaviour that minimizes within-subject variability and maximizes between-subject variability. Of course, objective behavioural analysis would likely be more complicated in humans, especially when accounting for the contexts in which these behaviours occur.

Theoretical approaches such as CFL, and empirical opportunities such as dense sampling of emails or webcams, provide exciting new possibilities for quantifying and revising personality traits in humans. Rather than continuing to stick with entrenched notions of personality, we think that now is the time to be bold. We can build a science of personality that emphasizes utility across disciplines and species and that builds a foundation amenable to the next generation of questions—questions about the genetic basis of personality, about the influence of environmental factors across the lifespan and about how personality depends on specific brain functions[16,65]. A critical ingredient in such progress will be the delineation of specific, testable causal models, for which we hope our sketches (Fig. 1) have provided initial discussion.

Received: 18 November 2018; Accepted: 28 September 2020;
Published online: 9 November 2020

References

1. Feist, J. & Feist, G.J. *Theories of Personality* (McGraw-Hill Education, 2008).
2. Eysenck, H.J. Biological dimensions of personality. in *Handbook of Personality: Theory and Research* (ed. Pervin, L. A.) 244–276 (Guilford Press, 1990).
3. Borgatta, E. F. The structure of personality characteristics. *Behav. Sci.* **61**, 8–17 (1964).
4. Fiske, D. W. Consistency of the factorial structures of personality ratings from different sour sources. *J. Abnorm. Psychol.* **44**, 329–344 (1949).
5. Norman, W. T. Toward an adequate taxonomy of personality attributes: replicated factors structure in peer nomination personality ratings. *J. Abnorm. Soc. Psychol.* **66**, 574–583 (1963).
6. Smith, G. M. Usefulness of peer ratings of personality in educational research. *Educ. Psychol. Meas.* **27**, 967–984 (1967).
7. Tupes, E.C. & Christal, R.E. Recurrent personality factors based on trait ratings. Technical Report ASD-TR-61-97 http://www.dtic.mil/dtic/tr/fulltext/u2/267778.pdf (USAF, 1961).
8. Goldberg, L. R. Language and individual differences: The search for universals in personality lexicons. *Rev. Pers. Soc. Psychol.* **2**, 141–165 (1981).
9. McCrae, R. R. & John, O. P. An introduction to the five-factor model and its applications. *J. Pers.* **60**, 175–215 (1992).
10. DeYoung, C. G., Quilty, L. C. & Peterson, J. B. Between facets and domains: 10 aspects of the Big Five. *J. Pers. Soc. Psychol.* **93**, 880–896 (2007).
11. Cattell, R. B. The description of personality: principles and findings in a factor analysis. *Am. J. Psychol.* **58**, 69–90 (1945).
12. Costa, P. T. & McCrae, R. R. Six approaches to the explication of facet-level traits: examples from conscientiousness. *Eur. J. Pers.* **12**, 117–134 (1998).
13. Mõttus, R., Kandler, C., Bleidorn, W., Riemann, R. & McCrae, R. R. Personality traits below facets: The consensual validity, longitudinal stability, heritability, and utility of personality nuances. *J. Pers. Soc. Psychol.* **112**, 474–490 (2017).
14. Forkosh, O. et al. Identity domains capture individual differences from across the behavioral repertoire. *Nat. Neurosci.* **22**, 2023–2028 (2019).
15. Baumert, A. et al. Integrating personality structure, personality process, and personality development. *Eur. J. Pers.* **31**, 503–528 (2017).
16. Eysenck, H. J. Personality and experimental psychology: the unification of psychology and the possibility of a paradigm. *J. Pers. Soc. Psychol.* **73**, 1224–1237 (1997).
17. Shalizi, C. R. & Crutchfield, J. P. Information bottlenecks, causal states, and statistical relevance bases: how to represent relevant information in memoryless transduction. *Adv. Complex Syst.* **5**, 91–95 (2002).
18. Woodward, J. *Making Things Happen: A Theory of Causal Explanation* (Oxford Univ. Press, 2005).
19. Gescher, D. M. et al. Epigenetics in personality disorders: today's insights. *Front. Psychiatry* **9**, 579 (2018).
20. DeYoung, C. G. Cybernetic big five theory. *J Res Pers* **56**, 33–58 (2015).
21. Spearman, C. "General intelligence," objectively determined and measured. *Am. J. Psychol.* **15**, 201–293 (1904).
22. Shalizi, C. Review: the domestication of the savage mind. *Am. Sci.* **97**, 244–247 (2009).
23. McNaughton, N. & Smillie, L. D. Some metatheoretical principles for personality neuroscience. *Personal Neurosci.* **1**, e11 (2018).
24. Parks, L. & Guay, R. P. Personality, values, and motivation. *Pers. Individ. Dif.* **47**, 675–684 (2009).
25. Reisenzein, R. & Weber, H. Personality and emotion. in *The Cambridge Handbook of Personality Psychology* (eds Corr, P.J. & Matthews, G.) 54–71 (Cambridge Univ. Press, 2009).
26. Prabhakaran, R., Kraemer, D. J. M. & Thompson-Schill, S. L. Approach, avoidance, and inhibition: personality traits predict cognitive control abilities. *Pers. Individ. Dif.* **51**, 439–444 (2011).
27. Gray, J. R. & Burgess, G. C. Personality differences in cognitive control? BAS, processing efficiency, and the prefrontal cortex. *J. Res. Pers.* **38**, 35–36 (2004).
28. Tamir, D. I. & Thornton, M. A. Modeling the predictive social mind. *Trends Cogn. Sci.* **22**, 201–212 (2018).
29. Vallacher, R. R. & Wegner, D. M. What do people think they're doing? Action identification and human behavior. *Psychol. Rev.* **94**, 3–15 (1987).
30. Roberts, B. W. et al. A systematic review of personality trait change through intervention. *Psychol. Bull.* **143**, 117–141 (2017).
31. Damasio, H., Grabowski, T., Frank, R., Galaburda, A. M. & Damasio, A. R. The return of Phineas Gage: clues about the brain from the skull of a famous patient. *Science* **264**, 1102–1105 (1994).
32. MacMillan, M. *An Odd Kind of Fame: Stories of Phineas Gage* (MIT Press, 2000).
33. Blumer, D. & Benson, D.F. Personality changes with frontal and temporal lobe lesions. in *Psychiatric Aspects of Neurologic Disease* 151–170 (Grune & Stratton, 1975).
34. Eslinger, P. J. & Damasio, A. R. Severe disturbance of higher cognition after bilateral frontal lobe ablation: patient EVR. *Neurology* **35**, 1731–1741 (1985).
35. Barrash, J. et al. "Frontal lobe syndrome"? Subtypes of acquired personality disturbances in patients with focal brain damage. *Cortex* **106**, 65–80 (2018).
36. Caramazza, A. On drawing inferences about the structure of normal cognitive systems from the analysis of patterns of impaired performance: the case for single-patient studies. *Brain Cogn.* **5**, 41–66 (1986).
37. Coltheart, M. Lessons from cognitive neuropsychology for cognitive science: a reply to Patterson and Plaut (2009). *Top. Cogn. Sci.* **2**, 3–11 (2010).
38. Arden, R. Cognitive abilities in other animals: An introduction to this special issue. *Intelligence* **74**, 1–2 (2019).

39. Vazire, S., Gosling, S.D., Dickey, A.S. & Schapiro, S.J. Measuring personality in nonhuman animals. in *Handbook of Research Methods in Personality Psychology* (ed. Robins, R. W.) 190–206 (The Guilford Press, 2007).
40. Bennett, P. C., Rutter, N. J., Woodhead, J. K. & Howell, T. J. Assessment of domestic cat personality, as perceived by 416 owners, suggests six dimensions. *Behav. Processes* **141**, 273–283 (2017).
41. Gosling, S. D., Kwan, V. S. Y. & John, O. P. A dog's got personality: a cross-species comparative approach to personality judgments in dogs and humans. *J. Pers. Soc. Psychol.* **85**, 1161–1169 (2003).
42. Dankert, H., Wang, L., Hoopfer, E. D., Anderson, D. J. & Perona, P. Automated monitoring and analysis of social behavior in *Drosophila*. *Nat. Methods* **6**, 297–303 (2009).
43. Berman, G. J., Choi, D. M., Bialek, W. & Shaevitz, J. W. Mapping the stereotyped behaviour of freely moving fruit flies. *J. R. Soc. Interface* **11**, 20140672 (2014).
44. Robie, A. A. et al. Mapping the neural substrates of behavior. *Cell* **170**, 393–406.e28 (2017).
45. Jhuang, H. et al. Automated home-cage behavioural phenotyping of mice. *Nat. Commun.* **1**, 68 (2010).
46. Reiter, S. et al. Elucidating the control and development of skin patterning in cuttlefish. *Nature* **562**, 361–366 (2018).
47. Gosling, S. D. From mice to men: what can we learn about personality from animal research? *Psychol. Bull.* **127**, 45–86 (2001).
48. Datta, S. R., Anderson, D. J., Branson, K., Perona, P. & Leifer, A. Computational neuroethology: a call to action. *Neuron* **104**, 11–24 (2019).
49. Anderson, D. J. & Perona, P. Toward a science of computational ethology. *Neuron* **84**, 18–31 (2014).
50. Costa, P.T. & McCrae, R.R. NEO Five-Factor Inventory (NEO-FFI). http://highriskdepression.org/forms/NEO%20Five%20Factor%20Personality%20Inventory.pdf (Psychological Assessment Resources, 1989).
51. Dubois, J., Galdi, P., Han, Y., Paul, L. K. & Adolphs, R. Resting-state functional brain connectivity best predicts the personality dimension of openness to experience. *Personal Neurosci.* **1**, 1–21 (2018).
52. Depue, R. A. & Iacono, W. G. Neurobehavioral aspects of affective disorders. *Annu. Rev. Psychol.* **40**, 457–492 (1989).
53. O'Connor, L.E., Berry, J.W., Lewis, T., Rangan, R.K. & Poursohrab, N. Personality, Psychopathology, and the Neurotransmitter Attributes Questionnaire (NAQ). *SAGE Open* https://doi.org/10.1177/2158244013492540 (2013).
54. Yarkoni, T. Neurobiological substrates of personality: a critical overview. in *APA Handbook of Personality and Social Psychology, Volume 4: Personality Processes and Individual Differences* (eds Mikulincer, M., Shaver, P.R., Cooper, M.L. & Larsen, R.J.) 61–83 (American Psychological Association, 2015).
55. Poldrack, R. A. et al. Discovering relations between mind, brain, and mental disorders using topic mapping. *PLOS Comput. Biol.* **8**, e1002707 (2012).
56. Varoquaux, G. et al. Atlases of cognition with large-scale human brain mapping. *PLOS Comput. Biol.* **14**, e1006565 (2018).
57. Eisenberg, I. W. et al. Uncovering the structure of self-regulation through data-driven ontology discovery. *Nat. Commun.* **10**, 2319 (2019).
58. Chalupka, K., Eberhardt, F. & Perona, P. Causal feature learning: an overview. *Behaviormetrika* **44**, 137–164 (2017).
59. Chalupka, K., Bischoff, T., Perona, P. & Eberhardt, F. Unsupervised discovery of El Nino using causal feature learning on microlevel climate data. in *UAI '16: Proceedings of the Thirty-Second Conference on Uncertainty in Artificial Intelligence* 72–81 (AUAI Press, 2016).
60. Mathis, A. et al. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat. Neurosci.* **21**, 1281–1289 (2018).
61. Ozer, D. J. & Benet-Martínez, V. Personality and the prediction of consequential outcomes. *Annu. Rev. Psychol.* **57**, 401–421 (2006).
62. Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A. & Goldberg, L. R. The Power of personality: the comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspect. Psychol. Sci.* **2**, 313–345 (2007).
63. Ahmad, N. & Siddique, J. Personality assessment using Twitter tweets. *Procedia Comput. Sci.* **112**, 1964–1973 (2017).
64. Silveira Jacques, J.C. Junior. *et al*. First impressions: a survey on vision-based apparent personality trait analysis. *IEEE Trans. Affect. Comput*. https://doi.org/10.1109/TAFFC.2019.2930058 (2019).
65. Allen, T.A. & DeYoung, C.G. Personality neuroscience and the five factor model. in *The Oxford Handbook of the Five Factor Model* (ed. Widiger, T.A.) (Oxford Univ. Press, 2017).
66. Deyoung, C.G. & Gray, J.R. Personality neuroscience: explaining individual differences in affect, behaviour and cognition. in *The Cambridge Handbook of Personality Psychology* (eds Corr, P.J. & Matthews, G.) 323–346 (Cambridge Univ. Press, 2009).
67. Allport, G. W. What is a trait of personality? *J. Abnorm. Soc. Psychol.* **25**, 368–372 (1931).
68. Gosling, S. D. & Vazire, S. Are we barking up the right tree? Evaluating a comparative approach to personality. *J. Res. Pers.* **36**, 607–614 (2002).
69. Gerlach, M., Farb, B., Revelle, W. & Nunes Amaral, L. A. A robust data-driven approach identifies four personality types across four large data sets. *Nat. Hum. Behav.* **2**, 735–742 (2018).
70. Millon, T. What is a personality disorder? *J. Pers. Disord.* **30**, 289–306 (2016).
71. Gosling, S. D. Personality in non-human animals. *Soc. Personal. Psychol. Compass* **2**, 985–1001 (2008).

Acknowledgements

We are grateful for feedback on this paper from the members of our laboratory and from T. Yarkoni, S. Vazire, W. Revelle and C. DeYoung. The authors are supported in part by NSF grants BCS-1840756 and BCS-1845958.

Competing interests

The authors have no competing interests.

Additional information

Correspondence should be addressed to J.D. or R.A.

Peer review information Primary Handling Editor: Marike Schiffer