Biology, society, or choice: How do non-experts interpret scientific explanations of behaviour?

Daniel Nettle[1,2]*
Willem E. Frankenhuis[3,4]
Karthik Panchanathan[5]

1. Institut Jean Nicod, Département d'études cognitives, Ecole Normale Supérieure, Université PSL, EHESS, CNRS, Paris, France
2. Population Health Sciences Institute, Newcastle University, Newcastle, UK
3. Department of Psychology, Utrecht University, Utrecht, The Netherlands
4. Max Planck Institute for the Study of Crime, Security and Law, Freiburg, Germany
5. Department of Anthropology, University of Missouri, USA

* To whom correspondence should be addressed: daniel.nettle@ens.psl.eu

Abstract

Scientists have developed different framings to explain human behaviour, from the structural context to the individual motivation down to the neurobiological implementation. We know comparatively little about how non-experts interpret these scientific framings, and what they infer when one framing rather than another is made salient. In four experiments, UK volunteers read vignettes describing the same behaviour, but using different explanatory framings. In study 1, we found that participants grouped explanatory framings into 'biological', 'psychological' and 'sociocultural' clusters. Different framings were often seen as incompatible with one another, especially when one belonged to the 'biological' cluster and the other did not. In study 2, we found that adopting a particular explanatory framing produced spontaneous inferences. Specifically, psychological framings led participants to assume the behaviour was malleable, and biological framings led them to assume it was not. In studies 3 and 3b, we found that the choice of explanatory framing can affect people's conclusions about effective interventions. For example, presenting a biological framing increased people's conviction that interventions like drugs would be effective, and decreased their conviction that psychological or socio-political interventions would be effective. These results illuminate the intuitive psychology of explanations, and also some pitfalls in scientific communication. Foregrounding a particular explanatory framing is likely to produce ideas in the audience – about what other factors are not causally important, how easy it is to change the behaviour, and what kinds of remedies are worth considering – that the communicator may not intend and that may not follow logically.

Introduction

The social and behavioural sciences have a number of different explanatory framings available for any given behaviour. These include the hermeneutic (it happens because of the meanings it has for people); the voluntaristic (the actors want or choose it); the dispositional (cognitive processes or traits of the actors' minds); the structural (social forces or roles); the cultural (cultural influences or norms); the adaptive (increasing survival or reproduction); the physiological (hormones or neurotransmitters); and the genetic (genetic propensities). Scientists agree that the explanations offered under these different framings need not be mutually exclusive. Indeed, influential frameworks such as David Marr's levels of analysis (Marr, 1982) and Niko Tinbergen's four questions (Tinbergen, 1963) exemplify how subsets of the framings complement one another. Thus, investigators choose the most useful framing for their current concerns, but without necessarily meaning to imply that other framings are not valid.

However, non-expert thought about explanations may not be so pluralistic. For example, respondents from the US public generally claim that anything involving conscious awareness and volition cannot be explained in terms of science (Gottlieb & Lombrozo, 2018). This suggests a non-expert distinction between events with psychological causes (the realm of subjectivity) and events with biological or physical ones (the realm of science). Even physicians divide mental disorders into those that have psychological causes and those that have biological causes (Ahn, Proctor, & Flanagan, 2009). Adopting a particular explanatory framing also has consequences for the inferences people make. People assume an intervention must be of a congruent type to the cause in order to be effective (e.g. drugs for biological causes, talking for psychological causes; Ahn et al., 2009; Iselin & Addis, 2003). This means that support for a certain type of intervention can be increased merely by controlling the way causality is framed. For example, highlighting social-structural causes of poverty as opposed to individual effort increases support for redistributive policies (Piff et al., 2020). Moreover, merely making the biological level of explanation salient makes people assume that a behaviour is more inherent, less malleable, and less under a person's control (Berent & Platt, 2021a, 2021b; Haslam & Kvaale, 2015). Thus, non-expert cognition seems to have the following features: there are several discrete types of possible cause of a human outcome; malleability and the locus of responsibility depend on which type of cause is operation; and to change the outcome, you have to intervene in way that matches the type of cause.

Previous research on explanatory framings offers or assumes typological distinctions, for example between 'biogenetic' and 'psychological' explanations (Haslam & Kvaale, 2015); 'individualistic' and 'societal' explanations (Cozzarelli, Wilkinson, & Tagler, 2001); or 'dispositional' and 'situational' explanations (Piff et al., 2020). Typically, these distinctions are established by examining responses to just a subset of the explanatory framings that are routinely used in the social and behavioural sciences. It is unclear how many types of explanatory framing there would be if respondents rated a fuller set. Moreover, in many cases, the internal homogeneity of these categories remains to be established: genes and hormones are both 'biogenetic', but they might have different assumed relationships to non-'biogenetic' causes. The aim of the current project was to map the landscape of explanations amongst British non-experts more clearly: how many types of explanation are there, which are seen as compatible and which incompatible, and what are the inferential patterns associated with each one?

Study 1 investigated the perceived similarity and perceived compatibility of 12 types of explanation for behaviour, with the aim of understanding how they cluster. Study 2 examined what inferences about malleability and other features participants made in response to the behaviour being framed by an explanation from one cluster versus another. Studies 3a and 3b examined the issue of congruence between explanatory framings and interventions: does changing the explanatory framing automatically produce different beliefs about what kinds of interventions would alter the behavioural outcome?

Study 1: Introduction

In study 1, UK general-population adults read vignettes describing a behaviour and were then asked to consider 12 possible explanations. They rated either how similar those explanations were to one another (half the participants), or how compatible they thought they were (the other half). This design allowed us to map the landscapes of perceived similarity and compatibility, and investigate how similarity and compatibility related to one another. We also used cluster analysis to determine how many types of explanation there were amongst the 12.

The study was exploratory rather than confirmatory. However, we were guided by one possible hypothesis: explanations would be seen as similar and compatible to the extent to which they

activate the same core cognitive system. Some psychologists argue that humans are endowed with a small number of separate core cognitive systems for dealing with different classes of entity (Spelke & Kinzler, 2007). From early childhood onwards, these give rise to several distinct systems of intuitive knowledge and spontaneous inference: intuitive physics, which deals with objects and their motion (Spelke, 1990); intuitive biology, which deals with species of plants and animals (Atran, 1998); intuitive psychology, which deals with the beliefs, desires and goals of human agents (Kamps, et al., 2017); and, perhaps, intuitive sociology, which deals with social groups, group membership and social norms (Shutts & Kalish, 2021). Each of these systems produces a characteristic, and different, type of intuitive reasoning. Human beings can potentially fall into the proper domain of any of these cognitive systems: they are, after all, physical objects, animals, individual agents, and members of social groups. Thus, in principle, any or all of the systems could be active in considering human behaviour, depending on which aspect of humans is foregrounded by the explanatory framing. We expected the explanations would cluster for similarity and compatibility according to which feature of humans – their animal-like embodiment, their psychological states, or their membership of social structures or networks – was made most salient.

Method

Preregistration. Although this study was exploratory, we pre-registered methods, materials and planned analyses at: https://osf.io/8ba5m.

Participants. Participants (200 UK resident adults with first language English) were recruited via online platform Prolific www.prolific.co. Mean age was 39.9 years (sd 13.0), with 104 women and 96 men. Participants were mostly non-students (145 non-students, 22 students, 33 not stated), in employment (124 employed full or part-time, 32 not currently employed, 34 not stated). Participants received £2.50 for taking part. Six participants were excluded for failure to fully complete the survey, leaving 194.

Design. Participants were allocated, in a between-subjects design, to one of four different vignettes, and one of two different response conditions (similarity or compatibility). Each vignette described a behavioural phenomenon that was common in a particular population, respectively: homicide, teenage motherhood, land diving (a risky display behaviour), and blood

blessing (a precautionary behaviour for dangers). At the end of the vignette, participants saw 12 'explanations that have been offered for the behaviour'. These explanations were in terms of: meaning, choice, a psychological trait, culture, social structure, social roles, evolutionary advantage, physiology, childhood experience, genetic propensity, motivation, and social pressure. From here on, we refer to the psychological-trait explanation as 'trait', because 'psychological' will be later used as a superordinate category. The precise wording of each explanation was slightly adapted to the vignette (see pre-registration for all materials).

Following the initial presentation of the vignettes and explanations, one explanation (the focal explanation) was placed at the top of the screen. The participant was then asked to rate, for each of the other 11 explanations, either how similar they thought that explanation was to the focal explanation (similarity condition), or how compatible each explanation was with the focal explanation (compatibility condition). In the compatibility condition, the initial instructions further specified that 'by compatible, we mean that both explanations could be true at the same time'.

The rating procedure was then repeated a further 11 times with each of the explanations in turn serving as the focal explanation. Thus, each participant completed a total of 132 ratings. Every possible pairing of explanations appeared twice (once with the first explanation of the pair as the focal, and once with the second explanation as the focal). As a robustness check prior to the main analysis, we examined the correlation between the rated similarity or compatibility the first time the participant rated the explanation pair, and the second time. In the pre-registration, we specified that we would exclude any participant for whom the correlation between their first set of ratings and their second was less than 0.7. Observed correlations for each participant were lower than anticipated (similarity condition: mean $r = 0.39$; compatibility condition, mean $r = 0.43$). However, the correlations between the across-participants *average* rating the first time and the second time were extremely high (similarity condition: $r = 0.93$; compatibility: $r = 0.79$). We therefore adjusted our pre-registered exclusion criterion and excluded any participant who lacked a significant positive correlation between their first set of ratings and their second. This excluded 58 participants, leaving 136. In the post-exclusion dataset, the correlation between across-participants average ratings the first and second time were $r = 0.96$ (similarity condition) and $r = 0.89$ (compatibility condition). All subsequent analyses use ratings averaged across participants and across the first and second time of rating.

Data Analysis. Data and code for all studies are available at: All data and code are available at: https://osf.io/wte2c/. Similarity ratings for pairs of explanations were well correlated across vignettes (inter-vignette correlations 0.68 – 0.88), as were compatibility ratings (inter-vignette correlations 0.67 – 0.84). All subsequent analyses were performed on the data pooled across vignettes. First, we used multidimensional scaling to create a map of rated similarity, using R package 'smacof' (see Jones, Mair, & McNally, 2018). We then used the configuration matrix from the multidimensional scaling output to perform k-means clustering analysis to detect clusters in the similarity space. We determined the optimal number of clusters to extract using the function fviz_nbclust() from the 'factoextra' R package (Kassambara & Mundt, 2020). We then repeated the multidimensional scaling and clustering for rated compatibility. Finally, we examined how rated compatibility related to rated similarity. We examined how rated compatibility differed by the respective similarity cluster membership of the two explanations in question.

Results

The multidimensional scaling results for similarity are shown in figure 1A. The stress metric was 0.21. This value is considerably less than the metric for randomly generated networks of this size, though there is no simple rule for what constitutes good fit of a multidimensional scaling solution (Mair, Borg, & Rusch, 2016). The optimal number of clusters was three. Cluster membership is indicated by colour-coding on figure 1A. One cluster consisted of explanations that we can label 'biological' (genetic, hormonal, evolutionary); the second, 'psychological' states or processes (trait, choice, motivation, meaning); and the third cluster, processes that would be described as 'sociocultural', though childhood experience was also included in this cluster (constituent explanations were culture, social role, social pressure, opportunity, and childhood experience).

The multidimensional scaling map for compatibility (figure 1B, stress metric 0.20) was very similar to that for similarity (coefficient of congruence after Procrustes rotation, 0.98). The optimal number of clusters for compatibility was two rather than three. The clusters extracted for compatibility were identical to the 'biological' similarity cluster (genetic, hormonal and evolutionary), plus the other two similarity clusters combined. Within the large 'non-biology' cluster for compatibility, the relative distances of the explanations were much as for similarity, with the more psychological explanations perceived as more compatible to one another than

they were to the more sociocultural explanations. However, culture and childhood in particular were rated as more compatible with ‘psychological’ explanations than they were similar to those explanations, explaining why, in the compatibility clustering, the ‘psychological’ and ‘sociocultural’ clusters did not separate.

The reason for the close congruence of the similarity and compatibility spaces shown in figure 1 is that compatibility was treated, in our sample, as a near-synonym for similarity (figure 2A). The correlation, across pairs of explanations, between their rated compatibility and their rated similarity was $r = 0.96$. Pairs of explanations belonging to different similarity clusters were perceived as much less compatible than pairs of explanations belonging to different similarity clusters (same clusters: mean compatibility 64.1, sd 11.8; different clusters: mean compatibility 42.7, sd 12.8; $t_{64} = -6.30$, $p < 0.001$). More specifically, a pair of explanations was perceived as much more incompatible if one member belonged to the ‘biological’ similarity cluster and the other did not than in any other scenario (figure 2B).

Our participants mostly did not view all the different explanations as compatible. Of the 12,672 compatibility ratings given, 5677 (45%) were lower than the mid-point of the scale (i.e. more towards incompatible than compatible). 1114 (9%) were zeroes (i.e. the totally incompatible endpoint of the scale). Of the 96 participants who rated for compatibility, 51 (53%) gave at least one rating of zero, and 28 (29%) gave at least 10 ratings of zero. All but 3 participants (97%) gave 10 or more compatibility ratings that were below the mid-point of the scale.

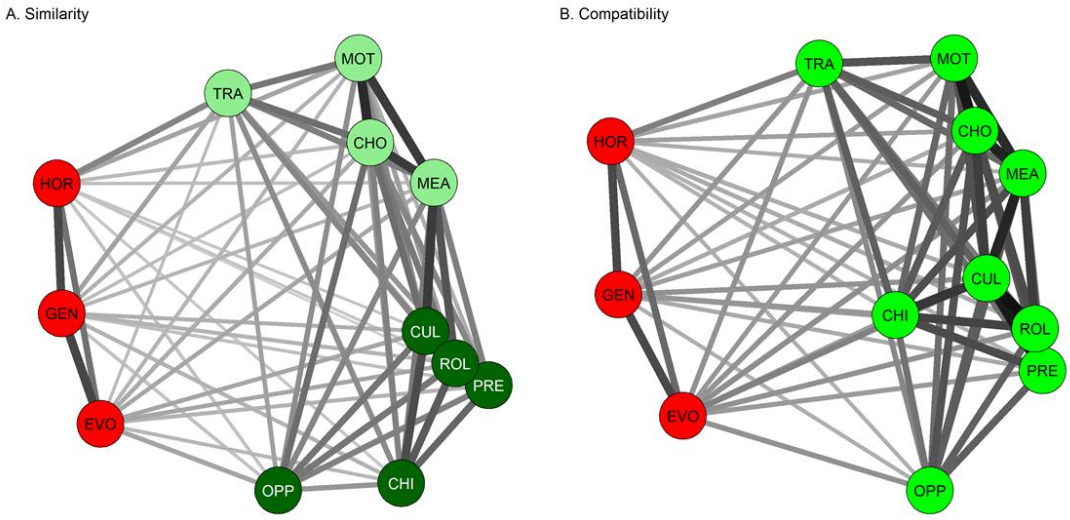

Figure 1. Visualization of multi-dimensional scaling and k-means clustering for (A) similarity and (B) compatibility of explanations, study 1. Euclidean distance and line darkness represent similarity or compatibility. Networks have been subject to Procrustes rotation for comparability. Node colour represents cluster membership in k-means clustering. Abbreviations: HOR: hormonal; GEN: genetic; EVO: evolutionary; TRA: (psychological) trait; MOT: motivational; CHO: choice; MEA: meaning; CUL: culture; ROL: social roles; PRE: social pressure; CHI: childhood experience; OPP: opportunity.

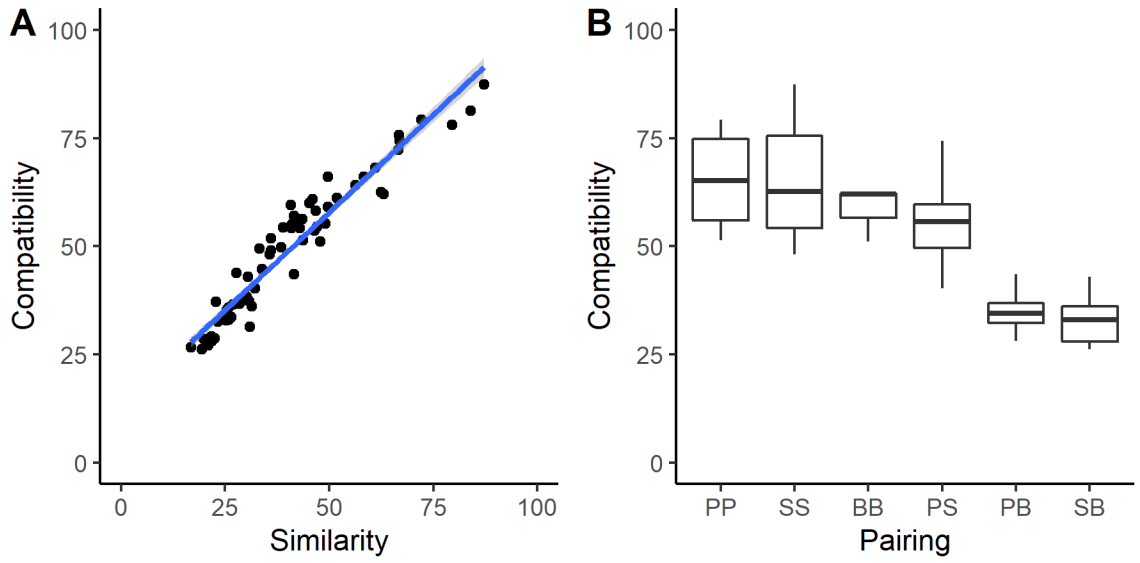

Figure 2. A. Relationship between rated compatibility and rated similarity, study 1. Each point represents a pair of explanations. B. Boxplot of rated compatibility of pairs of explanations, by the clusters they respectively belong to in the similarity space, study 1. PP: both 'psychological'; SS: both 'sociocultural'; BB: both 'biological'; PS: one 'psychological' and one 'sociocultural'; PB: one 'psychological' and one 'biological'; SB: one 'sociocultural' and one 'biological'.

Discussion

The main findings of study 1 can be summarised as follows. In terms of similarity to one another, explanations fell into three clusters, a 'biological' cluster (hormones, genes and evolutionary advantage); a 'psychological' cluster (choice, meaning, motivation and psychological traits); and a 'sociocultural' cluster (culture, social role, social pressure, opportunity and childhood experience). These three clusters plausibly correspond to the domains of intuitive biology, intuitive psychology and intuitive sociology respectively.

In terms of compatibility, our participants did not view all the different explanations as mutually compatible: around half of the ratings given were towards the incompatibility end of the scale, and a majority of participants found at least one pairing of explanations to be totally incompatible. Compatibility judgements were driven by similarity: explanations were perceived as more compatible with one another exactly when they were perceived as more similar to one

another. This is a non-obvious and possibly disturbing result from a scientific point of view. For example, many scientists would view a social role explanation and a hormonal explanation as dissimilar, but compatible (e.g. the hormonal changes provide the more proximal cause through which social expectations are internalised and expressed). It is therefore an important discovery that among these non-experts, dissimilar explanations are automatically presumed to be incompatible with one another. Since 'biological' explanations are considered the most dissimilar from other kinds of explanations, they are also considered to be the most incompatible with other kinds of explanations. This cognitive clash may represent the different core procedures of intuitive biology from those of intuitive psychology and sociology (although, this is not the only possible explanation, see General Discussion).

Study 2: Introduction

Study 1 described the mental map of explanation similarity in our study population. It suggested a three-way clustering of 'biological', 'psychological' and 'sociocultural' explanations. If these three clusters stem from different core cognitive domains (intuitive biology, intuitive psychology and intuitive sociology), then people should make different patterns of automatic inference from explanations in each of the clusters. For example, when intuitive biology is activated by referring to a bodily basis, psychological disorders are assumed to be immutable and transmissible to family members (Berent & Platt, 2021b, 2021a; Haslam & Kvaale, 2015). Presumably this is because intuitive biology embodies the automatic assumption that traits are due to an inner essence that is passed on through reproduction (Atran, 1998). Our general expectation is, therefore, that choosing one type of explanatory framing over another will automatically trigger different inferences about the behaviour under description, without any further explicit information needing to be given.

In study 2, we used the same four vignettes and the same 12 explanation as in study 1. Instead of asking participants to rate similarity and compatibility, we asked them to make judgements about what would be true of the behaviour if each of the 12 explanations were in fact correct. Specifically, we asked about *malleability* (how easily each behaviour could change); *externality* (the extent to which responsibility for the behaviour lies out in society rather than internally to individuals); and *simplicity* (whether the causes of the behaviour are simple or complex). As in study 1, our aims were exploratory. We expected different explanations to have different profiles of malleability, simplicity and externality. We tested this explanation both by comparing ratings of every explanation to every other, and also by grouping the explanations into the similarity

clusters discovered in study 1, and examining how malleability, simplicity and externality vary between clusters.

Methods

Pre-registration. We pre-registered methods, materials and planned analyses at: https://osf.io/a7s45. Study 2 was conducted concurrently with study 1 (although with different participants). Hence, the three-way similarity clustering of explanations observed in study 1 was not known. Analyses reported below that use the three clusters were therefore not pre-registered. There was also a pilot study for study 2 whose results informed the design and are presented in the pre-registration of the main study.

Participants. A total of 400 participants (200 male, 200 female; UK resident adults with first language English who did not participate in study 1) were recruited via online recruitment platform Prolific www.prolific.co. Mean age was 41.89 years (sd 13.28). Participants were mostly non-students (299 non-students, 25 students, 76 not stated), in employment (218 employed full or part-time, 83 not currently employed, 99 not stated). Participants received £2.20 for taking part.

For 10 surveys, the final submit button was not selected. The study contained a total of 12 attention checks (see below). 55 participants failed at least one of these and were excluded, leaving a final sample of 335.

Design. We used the same four vignettes as study 1, with approximately equal numbers of participants seeing each vignette. At the end of the vignette, respondents saw the list of 12 explanations that have been offered for the behaviour described in the vignette (the same list as in study 1). Participants were then taken through each explanation in turn, in random order. Assuming the given explanation to be true, they were asked to indicate their agreement with six statements. Instructions made clear that participants were not rating their degree of belief that the explanation was true, but rather what would follow if the explanation were true.

Dependent variables.

The six rating measures, all assessed on a 100-point slider anchored with ‘disagree’ and ‘agree’, were as follows.

1. This behaviour could change easily.
2. All of society is responsible for this behaviour.
3. This behaviour has a simple cause.
4. This behaviour is inevitable.
5. Responsibility for this behaviour lies within the individuals that do it.
6. The causes of this behaviour are complex.

These ratings were intended to give two-item measures of malleability (items 1 and 4), externality (2 and 5) and simplicity (3 and 6). We scaled each rating within subjects, as we were primarily interested in the within-subjects covariation of ratings across explanations, rather than the between-subjects covariation of ratings. Results were very similar without applying within-subjects scaling. We tested whether the six items covaried as intended by fitting a three-component PCA with oblimin rotation. The KMO statistic for the correlation matrix of the six items was 0.61, which is adequate for PCA but not excellent. The three components extracted were essentially uncorrelated with one another, and accounted respectively for 26%, 26% and 19% of the variation. The first component had loadings of 0.83 on item 1 and -0.82 on item 4. The second component had loadings of 0.90 on item 3 and -0.74 on item 6. The third component had loadings of 0.96 on item 2, but only -0.29 on item 5 (this item also loaded on both other components with 0.44 and 0.31 respectively). Thus, the pairs of items covary as intended for malleability and simplicity, but not well for externality, and the component for externality primarily reflects just one of the items. The scores from the PCA were used as the three dependent variables.

Attention checks

After completing the ratings for each explanation, respondents were shown one of the 12 explanations and asked (True/False) whether this was the explanation they just provided ratings about. Six of the attention checks show the right explanation (True), and six showed a different one (False).

Data Analysis. As in study 1, we pooled results across the four vignettes. Our first set of analyses used individual participant (within-subjects scaled) ratings as the unit of analysis. Using MANOVA and general linear models, we examined whether these differed by explanation. For our second set of analyses, we calculated the mean malleability, externality

and simplicity per explanation; hence, the explanation is the unit of analysis for these models. We grouped the explanations into the similarity clusters observed in study 1. We used MANOVA and linear mixed models (with a random effect of explanation) to examine whether the mean malleability, externality and simplicity differed by explanation cluster.

Results

The 12 explanations received significantly different ratings from one another across the set of three dependent variables ($F(33, 11550) = 108.64$, $p < 0.001$). Specifically, the explanations differed significantly from one another in terms of malleability ($F(11, 3850) = 160.62$, $p < 0.001$), externality ($F(11, 3850) = 148.26$, $p < 0.001$)) and simplicity ($F(11, 3850) = 61.93$, $p < 0.001$). Explanations belong to different clusters (from study 1) received different average ratings across the set of three dependent variables ($F(2, 9) = 9.06$, $p = 0.007$); specifically, they received significantly different average ratings for malleability ($F(2, 9) = 7.02$, $p = 0.015$), externality ($F(2, 9) = 14.29$, $p = 0.002$) and simplicity ($F(2, 9) = 5.25$, $p = 0.031$).

Figure 3 visualizes these results. The explanations belonging to the ‘biological’ cluster were rated as relatively unmalleable; those belonging to the ‘psychological’ cluster as relatively malleable; and those belonging to the ‘sociocultural’ cluster generally intermediate (though there was considerable variation, with culture very non-malleable, and opportunity fairly malleable). As for externality, the ‘sociocultural’ explanations were relatively external, whereas the ‘biological’ and ‘psychological’ explanations were relatively internal. In the case of simplicity, differences across clusters were the least marked, but ‘psychological’ explanations were the simplest, and ‘sociocultural’ the least simple.

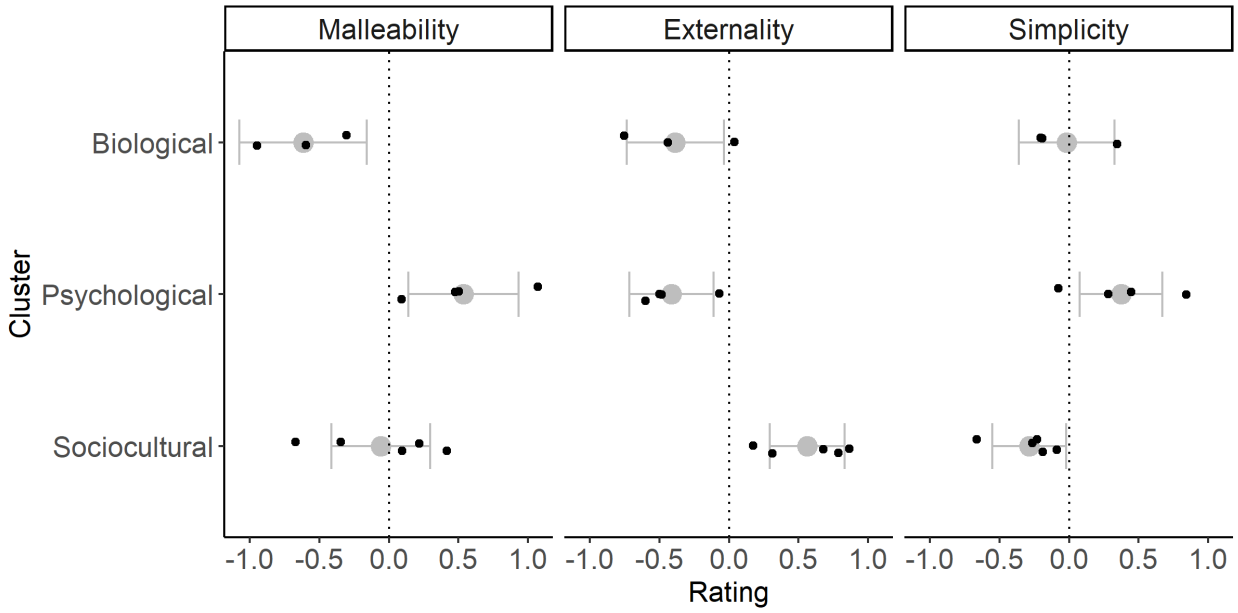

Figure 3. Ratings of malleability, externality and simplicity across explanation clusters, study 2. Grey points and whiskers show estimated marginal mean and 95% confidence interval at the cluster level. Superimposed black points represent means for individual explanations. Note that since variables are scaled, zero always represents the mean across all explanations.

Discussion

Different explanatory framings provoked different judgments in our participants. Sociocultural explanations were perceived as locating the causality out in society rather than within the individual, whereas both biological and psychological explanations were rated as locating causality within individuals. Partial exceptions to this generalization were evolutionary explanations and explanations in terms of meaning, which were perceived as intermediate between internal and external. These findings suggest our participants did attend to the material. There is an obvious sense in which psychological and biological explanations do concern intra-personal processes, and in which evolutionary and meaning explanations are atypical. Evolutionary explanations concern the fit between organism and environment, and explanations in terms of meaning concern processes that are internal to the mind but also

shared with others in social networks. We also found marked differences in how simple the explanations were perceived as being: psychological trait, motivational, choice and hormonal explanations were perceived as relatively simple. All other explanations were perceived as complex compared to these.

The finding with the most important implications is that different explanatory framings were perceived to imply different levels of malleability of the behaviour. Making reference to biology—through hormones, genes or evolution—automatically triggered the inference that the behaviour would be hard to change. There is no logical reason this should actually be true: hormone levels are highly dynamic, and can change rapidly in response to environmental inputs. Moreover, changes in the environmental context can completely change which behaviours are evolutionarily adaptive. Thus, merely demonstrating some 'biological' locus does not mean the behaviour could not change. However, this was not how our participants saw it. Possibly this is due to our 'biological' framings cueing intuitive biological cognition, under whose logic creatures follow inner, immutable, transmissible essences (Atran, 1998). The researcher interested in biological mechanisms or evolutionary adaptiveness must be alert to the possibility that their work will be taken to imply immutability of the behavior, even when no such immutability is intended.

By contrast, 'psychological' explanations (especially psychological traits, motivation and choice) produced the highest inferred malleability. This may stem from their engagement of intuitive psychology, whose function is to predict and intervene in individuals' behaviour as time and context changes. Thus, intuitive psychology sees behaviour as stemming from transient and reversible inner states such as beliefs and desires. It follows that the researcher who uses psychological explanatory framings may be unfairly accused of naivety about how difficult behaviour change really is (for example, because there are structural barriers to change).

The sociocultural explanations were markedly heterogeneous in terms of inferred malleability. Explanations in terms of opportunity produced fairly high ratings of malleability. Explanations in terms of culture, on the other hand, led to the inference that behaviour change was almost as difficult as in cases of genes and evolution. The difference between genes and culture in the minds of our participants was not that they implied different levels of malleability—these were about the same—but that genes locate the causality within the individual, and culture locates it out in society. These findings are interesting in light of the way culture is often described in

academic literature, as an inheritance system at least partly analogous to genes, with a fair degree of historical persistence. It also relates to critiques of the use of the culture concept (for example, in the case of 'the culture of poverty', Black and Dolgon (2021)). These critiques see the term as essentializing the current behaviours of people in poverty, obscuring the extent to which those behaviours may in some cases be dynamic, rational, responses to very specific structural factors or policy measures.

Study 3a

Studies 1 and 2 found evidence that people group explanations of behaviour into 'biological', 'psychological' and 'sociocultural' clusters. Further, people assume differential malleability according to which explanatory framing is invoked. Study 3 (a and b) extends studies 1 and 2 by focusing on how the choice of explanatory framing affects the preferred choice of intervention. That is, what kinds of measures are plausible candidates for making the behaviour change? There are close connections between explanations (why did X happen?), counterfactuals (what would have to have gone differently for X not to happen?), and intervention strategies (how can X be prevented from happening?) (Gerstenberg, Peterson, Goodman, Lagnado, & Tenenbaum, 2017; Halpern & Pearl, 2005; Quillien, 2020; Woodward, 2003). Thus, we would expect the choice of explanatory framing to guide consideration of relevant interventions. Physicians intuit that pharmacological treatment will be more effective for disorders they consider biogenetic in origin, and talking cures for disorders they consider of psychological origin (Ahn et al., 2009). If the tacit assumption that intervention must mirror cause is general, audiences may make inferences from researchers' choices of levels of explanation. For psychological disorders, for example, merely communicating about proximate neurobiological pathways could cause the audience to infer that political or socioeconomic reform would *not* make any difference, because such interventions seem to be of the wrong type. This would be an instance of the competitive nature of cause-and-effect attribution: mentally activating one set of causes and effects suppresses or inhibits consideration of others (Chater & Loewenstein, 2022).

In study 3a, we used two of the vignettes presented in studies 1 and 2, with slight wording modifications. In a between-subjects design, participants saw the description of the behaviour, plus one explanation, or no explanation. They were then asked how much difference could be made to the behaviour by changing biology (for example through drugs), changing psychology (for example through thinking training), and changing society (for example through political

reform). Since every participant was asked about the effectiveness of all three types of intervention, the design potentially allowed us to detect multiple effect patterns. For example, a certain explanatory framing might be perceived as making all interventions more effective (relative to no explanation), all interventions less effective, or some interventions more effective and some interventions less so.

We pre-registered three predictions (see preregistration link below), using the explanation cluster rather than the individual explanation as the explanatory variable:

> P1. The cluster of explanation offered will affect the rated effectiveness of the three types of interventions.
>
> P2. Offering an explanation belonging to a particular cluster will *increase* the rated effectiveness of interventions belonging to that cluster (relative to 'no explanation').
>
> P3. Offering an explanation belonging to a particular cluster will *decrease* the rated effectiveness of interventions belonging to the other two clusters (relative to 'no explanation').

We also performed exploratory analyses at the individual explanation level rather than the level of the explanation cluster. Participants of study 3a who were in the 'no explanation' group additionally completed study 3b (see below).

Method

Preregistration. We pre-registered methods, materials and planned analyses at: https://osf.io/8ba5m.

Participants. Participants were recruited via Prolific as per studies 1 and 2. We pre-registered a sample size of 320. However, due to an error, the participants in the 'no explanation' condition were not shown the study 3b questions (for study 3b, see below), leading us to add 80 more participants in the 'no explanation' condition. Preliminary inspection of the data after the 320+80 participants suggested we were under-powered to determine what was happening at the

individual explanation level. We therefore added another 320 participants as per the original design. Post-hoc increases in sample size inflate type-I error rates. We therefore recalculated the critical p-values required in the final sample to keep overall type-I error rates at 0.05, using the p_{crit} formula (Sagarin, Ambler, & Lee, 2014), for one additional round of data collection, and assuming that any *p*-value less than 0.2 would have led us to collect more data. This corrected α was 0.04, which should therefore be considered the appropriate α for the results presented below.

Design. We used vignettes from studies 1 and 2, homicide and teenage parenthood. Each participant saw one vignette. In study 1, there were unequal numbers of explanations in the three clusters: three 'biological', four 'psychological' and five 'sociocultural'. To equalize the representation of clusters, we added a biological explanation ('brain circuits'), and omitted the 'childhood' explanation (from 'sociocultural'), producing four explanations per cluster. The probability of being assigned to the 'no explanation' condition was four times that for every other condition. This is so that when explanations are aggregated to the cluster level, the numbers of observations in the 'no explanation', 'biological', 'psychological' and 'sociocultural' clusters were approximately equal (see table 1 for achieved sample sizes).

Table 1. Design summary, with realised numbers of participants in each cell.

Vignette	No explanation	Biological		Psychological		Sociocultural	
Homicide	122	Brain	27	Choice	15	Culture	20
		Evolutionary	23	Meaning	16	Opportunity	21
		Genetic	20	Motivation	14	Pressure	17
		Hormonal	18	Trait	25	Role	21
Teenage motherhood	115	Brain	22	Choice	23	Culture	16
		Evolutionary	17	Meaning	20	Opportunity	21
		Genetic	19	Motivation	23	Pressure	24
		Hormonal	20	Trait	22	Role	19

Vignettes. The vignettes describe the behaviour (homicide or teenage motherhood respectively) followed by the statement: 'scientists have found out that…' and one of the explanations. The vignettes and explanations had slight wording differences from studies 1 and 2, particularly to

make all options more grammatically similar and all contain an explicitly causal ‘because’ (see pre-registration for details). In the no-explanation condition, we omitted the ‘scientists have found out that…’ statement.

Dependent variables. After reading each vignette and any explanation, participants responded by answering three questions on a 100-point scale anchored with ‘none’ and ‘a great deal’.

- How much could the rate of homicide be decreased by changing biology, for example through tablets, injections, or brain stimulation?

- How much could the rate of homicide be decreased by changing psychology, for example through education, persuasion, or thinking training?

- How much could the rate of homicide be decreased by changing society, for example through better political representation, increasing incomes, or better jobs?

Predictions and analysis plan.
Pre-registered predictions were as listed in the study 3 introduction. In the pre-registration, we acknowledged the possibility that P2 and P3 might hold for some types of explanation but not others. For example, biological explanations might reduce the rated effectiveness of societal interventions, but not vice versa. We made no a priori predictions about this heterogeneity.

We fitted linear mixed models with the predictors: intervention type, explanation cluster, and their interaction; and rated effectiveness as the outcome. P1 implies effects of explanation cluster, either as main effects or interactions with intervention type. P2 and P3 imply significant interactions between intervention type and explanation cluster. We included random effects of individual explanation, and participant. We also ran the same analyses with individual explanation as the independent variable, and these results are superimposed on figure 4. Data from the two vignettes were pooled.

Results

In a mixed model with effectiveness as the outcome and intervention, explanation cluster, and their interaction as predictions, there was a significant effect of intervention ($F(2, 1431.77) = 207.35$, $p < 0.001$). This was driven by an overall endorsement of psychological and social interventions as more effective than biological ones (estimated marginal means: biological, 38.5 (se 1.55); psychological 60.9 (1.55); social 59.4 (1.55)). The main effect of explanation cluster was not significant ($F(3, 6.43) = 0.44$, $p = 0.73$), but there was a significant interaction between intervention and explanation cluster ($F(6, 1431.74) = 23.33$, $p < 0.001$).

To examine this interaction, we broke the data down into the three types of intervention. We ran separate models to examine the effect of each type of explanation (relative to the reference category of 'no explanation') on rated effectiveness for that type of intervention. Results are shown in figure 4. Providing a biological explanation both increased the rated effectiveness of biological interventions, relative to no explanation, and decreased the rated effectiveness of psychological and social interventions, relative to no explanation. By contrast, providing a psychological or social explanation had no systematic effect of the rated effectiveness of any type of intervention. Thus, we found support for P1 (choice of explanation cluster affects rated effectiveness of interventions) in general, but P2 (explanations enhance rated effectiveness of congruent interventions) and P3 (explanations suppress rated effectiveness of incongruent explanations) were only supported for 'biological' explanations.

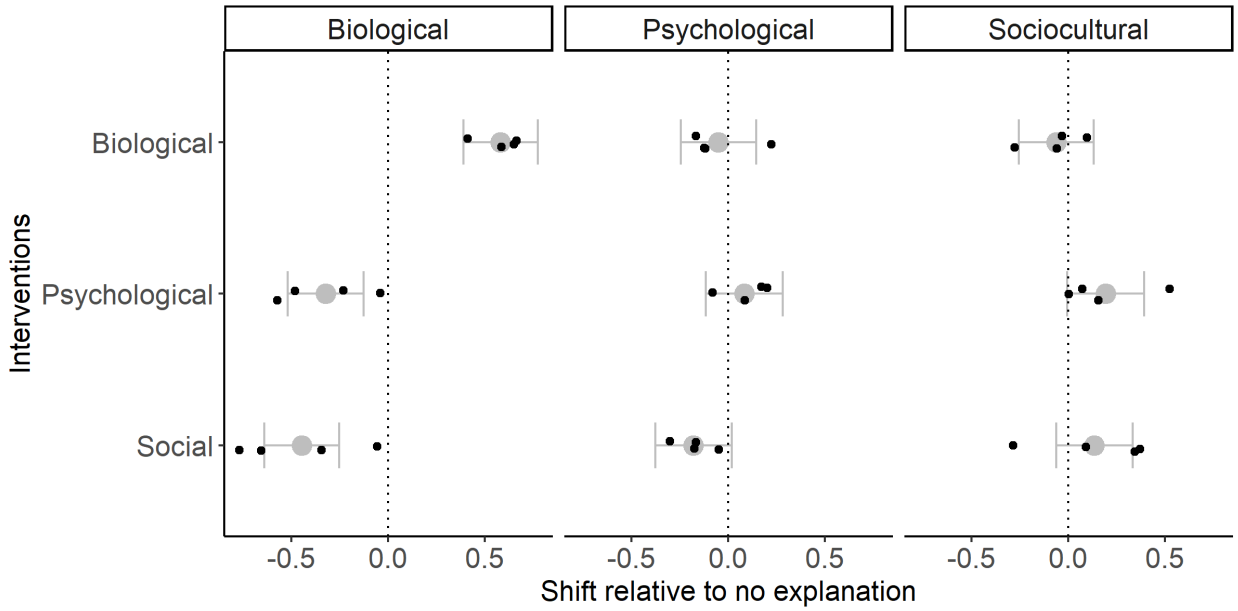

Figure 4. Effect of providing an explanation of a particular cluster (columns) on the rated effectiveness of a particular type of intervention (rows), relative to 'no explanation'. Grey points and whiskers show standardized effect sizes and their 95% confidence intervals at the explanation cluster level. Superimposed black points are the standardized effect sizes when the individual explanation is used as the unit of analysis.

Discussion

Study 3 confirmed that the choice of explanatory framing can matter for inferences about what kinds of interventions are potentially effective for changing the outcome. However, it was only biological explanations that had clear implications for intervention effectiveness: providing a biological-type explanation both *enhanced* the perceived effectiveness of biological interventions such as drugs; and, importantly, *suppressed* the perceived effectiveness of psychological and social interventions. These findings confirm the competitive nature of explanatory framing: adopting a framing both facilitates the inferences related to that framing, and inhibits the inferences that might follow from other framings (Chater & Loewenstein, 2022).

The lack of either enhancement or suppression effects when we provided psychological or sociocultural-type explanations is perhaps puzzling. One possibility is that these are our participants' default-type explanations for the kinds of behaviours in our vignettes. Thus, providing these types of explanation did not change anything for our participants compared to

providing no explanation. Compatible with this possibility, rated effectiveness of social and psychological interventions was higher overall than those of biological interventions, including in the ‘no explanation’ condition. Thus, providing a ‘biological’ explanation may have caused a reframing compared to no explanation, with consequences for assumptions about change, whereas providing a ‘psychological’ or ‘sociocultural’ explanation just confirmed assumptions people already held prior to our providing an explanation.

Study 3b

Study 3a showed there can be competition between explanatory framings, and that explanations are linked to assumptions about what interventions will make a difference: being given a ‘biological’ explanation both increased the perceived effectiveness of a biological intervention, and reduced the perceived effectiveness of psychological and social interventions. Study 3b was an ancillary study, added to the questions of the study 3a participants in the ‘no explanation’ condition, that tested a corollary of these findings. If ‘biological’ and ‘non-biological’ explanations are perceived to stand in competitive exclusion, then the same might be true for interventions. That is, participants who learn that a biological intervention is effective in changing the behaviour might infer that other kinds of intervention will have no effect, whilst those who learn that a ‘biological’ intervention is ineffective at changing the behaviour might infer the opposite, that another interventional approach is required.

In study 3b, after participants had read the vignette and rated biological, psychological and social interventions for changing the behaviour, we gave them additional information that one of the intervention types had been tried and either had a big effect on the behaviour (effective condition) or had no effect on the behaviour (ineffective condition, between subjects). We then asked participants to rate the extent to which now thought the other two intervention types would be less effective than they previously estimated, more effective, or the same. The principle of competitive exclusion between ‘biological’ and ‘non-biological’ explanations, suggested by studies 1 and 3, predicts that learning a biological intervention is effective will reduce rated effectiveness of other kinds of interventions, and learning that a biological intervention is ineffective will increase rated effectiveness of other kinds of interventions.

Methods

Preregistration. We pre-registered methods, materials and planned analyses along with study 3 at: https://osf.io/8ba5m.

Participants. Participants from the 'no explanation' condition of study 3 (n = 153).

Design. Study 3b a three (intervention type) by two (effective or ineffective) factorial design.

Stimulus. After completing the dependent measures of study 3a, study 3b participants were given the following information.

> Scientists have discovered that [changing biology, for example through drugs/changing psychology, for example through thinking training/changing society, for example through better political representation] [makes no difference/substantially reduces] [the rate of homicide/the rate of teenage parenthood].

Dependent measure.
Participants were then asked:

> Does the information given above make you think that [changing psychology/changing biology/changing society] would be less effective in changing [the rate of homicide/the rate of teenage parenthood] than you previously thought, more effective, or no change?

Two questions were displayed, for the two intervention types other than the one mentioned in the stimulus. Participants responded on a slider from 'less effective' to 'more effective', with 'no change' at the mid-point.

Predictions and analysis plan.
We pre-registered two predictions:

> P1. Learning that one class of intervention is ineffective will increase belief in the effectiveness of the other two classes of intervention.
>
> P2. Learning that one class of intervention is effective will decrease belief in the effectiveness of the other two classes of intervention.

We specified that P1 and P2 might hold for some combinations of intervention classes but not others. We tested P1 and P2 with linear mixed models with intervention mentioned, intervention asked about, and effectiveness condition (effective or ineffective) as the predictors. We also used one-sample t-tests to examine whether average ratings deviated from 'no change' for particular combinations of intervention mentioned and intervention asked about.

Results

We fitted a linear mixed model with change in perceived effectiveness as the outcome variable, and predictors of intervention in question, intervention mentioned, and whether the mentioned intervention was described as effective or ineffective. There was a significant interaction between intervention in question and intervention mentioned ($F(1, 147) = 4.30$, $p = 0.04$). All other interactions were non-significant (ps > 0.05). This suggests that learning about the effectiveness or ineffectiveness of some interventions affects the perceived effectiveness of some other interventions. To probe which ones affected which, we calculated the mean and 95% confidence interval of change in effectiveness for each combination of intervention in question, intervention mentioned, and effectiveness or ineffectiveness (figure 5). Where the 95% confidence interval does not include zero, we can conclude that this class of information significant changes perceived effectiveness of that class of intervention. As figure 5 shows, learning that a psychological or social intervention is effective reduced the perceived effectiveness of biological interventions. Curiously, learning that a psychological or social intervention was ineffective did not increase the perceived effectiveness of 'biological' interventions. Nor did learning that a biological intervention was effective reduced the perceived effectiveness of psychological or social interventions. The only other significant shift was that learning that a social intervention was effective increased the perceived effectiveness of a psychological intervention. The converse did not hold.

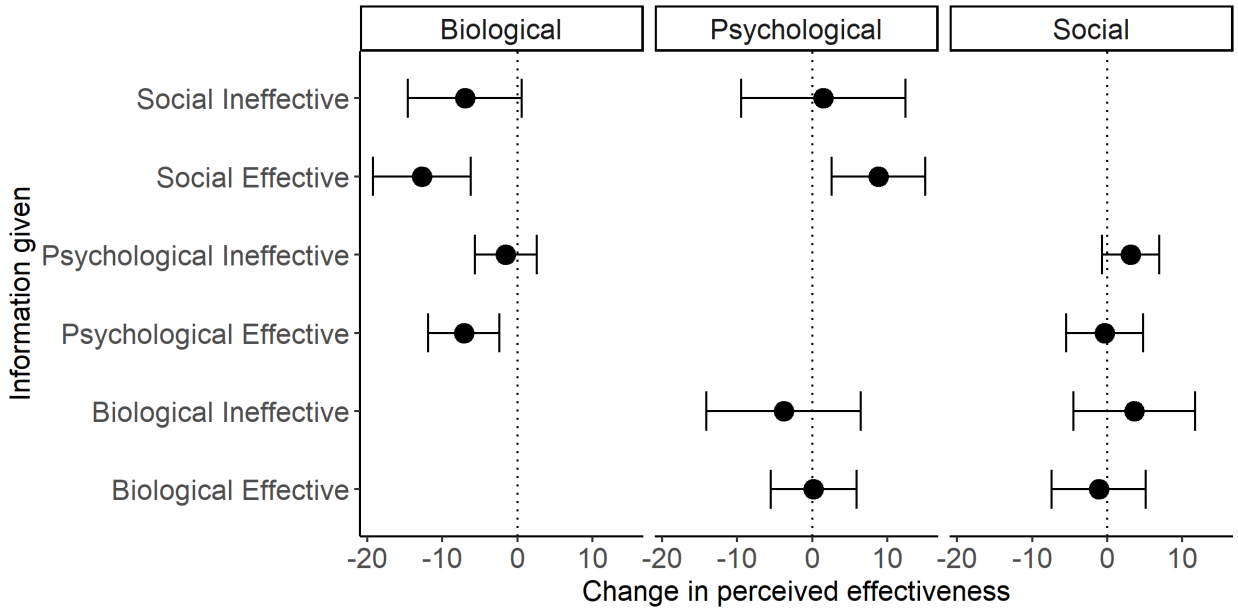

Figure 5. Study 3b results. Change in perceived effectiveness of an intervention (columns) by which other intervention had been mentioned, and whether it had been described as effective or ineffective (rows). Points show the mean shift and whiskers the 95% confidence interval. Where the 95% confidence interval does not include zero, that category of information significantly shifts average perceived effectiveness.

Discussion

If different explanatory framings suggest the effectiveness of different types of intervention, and different explanatory framings competitively exclude one another, we reasoned that learning that one type of intervention does (or does not) work might produce the automatic inference that another types of intervention will not (or will) do so. We found only partial evidence for such effects. Specifically, learning that a social or psychological intervention is effective led participants to reduce their perceived effectiveness of a biological intervention. The converse did not hold: learning that a biological intervention was effective did not reduce the perceived effectiveness of psychological or social interventions. On the face of it, this is puzzling: a widespread concern in the literature on psychological disorders, for example, is the concern that publicity around pharmaceutical treatments leads sufferers, clinicians and the general public to neglect the promise of addressing their prevalence through psychological and particularly socioeconomic means (Davies, 2021). Our results do not concur with this, though they do suggest the converse effect: stressing the potential effectiveness of political and social

approaches to psychological suffering would undermine the perception of effectiveness of drugs. A possible explanation is that, for the particular behaviours described in our vignettes, our participants thought that psychological and social interventions were much more effective than biological ones at baseline (see study 3 results). Thus, learning that an intervention-class they regarded as weaker (biological) was effective was not sufficient to diminish their belief in interventions they already regarded as stronger; whereas learning that interventions of their favoured class were indeed effective further inhibited the perceived effectiveness of the disfavoured class.

Learning that psychological and social interventions were *in*effective did not increase the perceived effectiveness of biological interventions. If anything, it tended towards decreasing it. A possible explanation is the ineffectiveness of psychological and social interventions is a cue that the behaviour is completely non-malleable. This is particularly true given that our participants thought the psychological and social interventions to be the strongest available at baseline. Thus, if even those do not work, it seems likely that a class of intervention perceived to be weaker in the first place (biological) would do so. Hence, learning about effectiveness (the behaviour is malleable, and responds to our best interventions) would not necessarily have the mirror image of the consequences of learning about ineffectiveness (the behaviour is not malleable even with our best interventions). We also found that learning about the effectiveness of a social intervention *increased* the perceived effectiveness of a psychological one, presumably because it provided evidence of malleability in general.

In summary, study 3b found some evidence that knowing that one class of intervention is effective can undermine perceived potential of other kinds of intervention. Specifically, this phenomenon only occurred across the biological/non-biological boundary: the effectiveness of psychological and social measures undermined the perceived effectiveness of biological measures like drugs, although the converse, did not hold.

General discussion

These studies have charted the mental maps of different explanations amongst our participants, and begun to probe the inferences people make from the choice of one explanatory framing over another. Although all three studies had pre-registered predictions and designs, we describe this work as primarily exploratory rather than testing a strong theory or causal model. Some

patterns emerged we believe to be worthy of discussion. We remind the reader that our participants were not professional academics and mostly not students. Although not a representative sample of the UK population, their responses are likely to be a reasonable guide to how explanatory framings are understood in the non-expert UK population at large. Moreover, given the misunderstandings that persist in the literature, we expect—though cannot here demonstrate—that much expert cognition about explanatory framings would be fairly similar. We begin by summarizing the observed patterns.

First, our participants perceived family resemblance between certain kinds of explanations. Specifically, they produced, unguided by us, 'biological', 'psychological' and 'sociocultural' clusters of explanations (study 1). They rated the constituent explanations of each of these groupings to be relatively similar to one another, and more different from the members of other groupings. This view arguably does not reflect our actual state of scientific knowledge (for example, hormones, genes and adaptive advantage, all 'biology' to our participants, are actually very different types of scientific explanation, with complex mappings to one another).

Second, our participants were not intuitive compatibilists. When considering the extent to which different explanations were compatible, almost all participants felt that at least some pairings of explanations were more incompatible than compatible (study 1). The more dissimilar explanations were rated as being, the more incompatible they were deemed to be. Again, this does not correspond well to the actual state of scientific knowledge, where explanations at very different levels (for example, the neural and the psychological or social) are widely considered complementary rather than competing; and much academic ink is spilt in the competition between explanations (for example in terms of culture versus social roles) that our participants would view as pretty similar. 'Biological' explanations, being rated as the most divergent from others in terms of similarity, were also seen as the most incompatible with other types.

Third, given an explanatory framing, our participants made spontaneous inferences about the locus of causation, and about malleability. Both 'biological' and 'psychological' framings led participants to infer that the causation is internal to the individual rather than external. However, 'biological' framings were taken to imply that the behaviour is not malleable, whereas 'psychological' framings implied a high degree of malleability (study 2). This finding echoes recurrent concerns expressed in the literature on psychological disorders. On the one hand, a 'biogenetic' conceptualization of psychological difficulties encourages the view that those

disorders are essential features of their sufferers, and not easy to change (Berent & Platt, 2021a; Haslam & Kvaale, 2015; Lebowitz, Ahn, & Nolen-Hoeksema, 2013). On the other hand, 'psychological' framings are sometimes 'cruelly optimistic': they seem to imply that change is easier than it actually is, especially when persistent socio-structural factors are integral to the actual network of causes (Brickman et al., 1982). Needless to say, the level of explanation researchers provide does not actually have clear implications for how likely the behaviour is to change: Gary Becker famously, though controversially, modelled addiction as choice (Becker, 1968), yet addiction is by definition hard to reverse. At the same time, many neural and hormonal states can be identified that are transient and reversible.

Fourth, our studies provide some evidence for an intuitive principle that 'the medicine must fit the cause'. That is, effective interventions should be congruent to the foregrounded explanation. Specifically, providing a 'biological' framing for a phenomenon both boosted the perceived effectiveness of intervening biologically, such as through drugs, and reduced the perceived effectiveness of psychological and societal change (study 3a). Moreover, providing evidence that psychological or societal interventions were effective reduced the perceived likelihood of a biological intervention working (study 3b). Since the cognitive function of providing causal explanations is to guide the search for ways of acting to make the world different (Quillien, 2020; Woodward, 2003), these effects make sense. They buttress the view that seizing hold of the explanatory framing is tantamount to seizing control of the kinds of remedies that are considered. For example, the popularity of individual-level nudges in behavioural public policy has distracted epistemic attention from structural solutions that might be more beneficial in the long run (Chater & Loewenstein, 2022); the food industry prefers to frame obesity in terms of psychological traits or lack of knowledge, whilst public health research stresses structural and systemic factors (Jenkin, Signal, & Thomson, 2011); and pharmaceutical interests strongly promote the framing of psychological problems as 'diseases of the brain' (Davies, 2021; Moncrieff, 2007; Read & Cain, 2013), because this naturally leads to the privileging of drug intervention. Inter-disciplinary turf wars about appropriate levels of explanatory framing often seem motivated by the feeling that allowing one explanatory framing to gain currency will guide attention toward some kinds of remedies, and away from others. Our results in study 3 make this concern appear a reasonable one.

One could interpret our findings through either of two lenses. Through one lens, the way our participants classified different explanatory framings, and reasoned from them, reflects reliably

developing features of human minds in general. Specifically, human core knowledge typically gives rise to distinct intuitive theories in the domains of biological, psychology, and sociology (Spelke & Kinzler, 2007). To the extent that these intuitive theories incorporate different priors and inferential principles, evoking one or the other of them is bound to produce different assumptions, for example about malleability. At the other extreme, these findings, which come after all from adults in just one population, could represent no more than the influence of a particular set of discursive practices, cultural and religious histories, or sets of educational institutions. The kinds of data we present here cannot adjudicate between these positions, or help develop the many intermediate positions that are possible (for example, that some form of biological/psychological distinction develops reliability in all populations, but can be culturally elaborated in different ways). Logical extensions to this work would involve studying different populations; studying the development of these distinctions in infancy, childhood, and adolescence; and studying the effects of scientific training. All of these strategies could be informative about the ubiquity of the patterns we have observed. We note the extensive developmental work on core cognition and intuitive theories. This has not yet been linked to the issues of explanatory framings we study here. We also note important cross-cultural research on concepts of mental life, which suggests that distinctions between psychological experience ('mind') and bodily sensation ('body') appear in very different populations, with subtle differences in which phenomena align to which category (Weisman, 2022). This work represents a potential paradigm for possible cross-cultural extensions of the current enquiry.

Regardless of what we can conclude about the causal origins of the patterns we observed, they do serve to illustrate potential hazards that can attend the scientific communicator. A researcher communicating about how phenomenon X works at the implementational, neurobiological level may be assumed by their audience to be claiming that phenomenon X can *only* (or most importantly) be viewed as having this kind of cause, and hence that their research is incompatible with social or psychological determinants also being important. They may also be taken as implying that phenomenon X is difficult to change, and if it can be changed, that only drug-type interventions are relevant candidates. The communicator may not intend any of these conclusions, or believe that the conclusions follow from their work. The communicator may not even realise the inferences are being made. Nonetheless, our research suggests they may well be made, leading to unnecessary conflicts, unproductive debates, and persistent misunderstandings. We thus recommend that communicators always take time to highlight the plurality of valid explanatory framings and the possible interplay between them. They may wish

to draw explicit attention to what their choice of explanatory framing *does not* imply. This might seem irrelevant or time consuming, but it might also lead to better understanding and forestall misplaced criticism.

This leads us to an important question, that of how easily the cognitive patterns we have described can be overcome. Overcoming them could have many benefits, including improving scientific literacy, producing better reasoning in policy-making, and avoiding academic turf wars. Generations of educators have structured their material around pluralistic frameworks such as Marr's three levels (Marr, 1982) and Tinbergen's four questions (Tinbergen, 1963), exactly because they provide a clear, reasoned antidote to the appearance that different framings are competitors to one another. To our knowledge, the extent to which they succeed has not been studied. Lebowitz et al. (2013) employed a brief audiovisual intervention explaining how neurobiological systems respond dynamically to environmental inputs. Seeing this intervention reduced prognostic pessimism amongst depression sufferers who endorsed biogenetic explanations, presumably because it broke the intuitive link between 'biology' and 'non-malleable'. In future work, we hope to examine the effects of such interventions on people's perceptions of compatibility, and implied malleability, when biological explanations are given. If successful, such work could suggest pathways to facilitating the acceptance of explanatory pluralism.

Data available

All data and code are available at: https://osf.io/wte2c/

Acknowledgements

The authors thank Emma Bridger and Eva Wittenberg for their feedback, and Coralie Chevallier and the team Evolution et cognition sociale at the Institut Jean Nicod for their support.

Funding

DN's research is supported by the EUR FrontCog grants ANR-17-EURE-0017 and ANR-10-IDEX-0001-02 to Université PSL; and ANR grant ANR-21-CE28-0009. WEF's contributions have been supported by the Dutch Research Council (V1.Vidi.195.130) and the James S. McDonnell Foundation (https://doi.org/10.37717/220020502).

References

Ahn, W., Proctor, C. C., & Flanagan, E. H. (2009). Mental Health Clinicians' Beliefs About the Biological, Psychological, and Environmental Bases of Mental Disorders. *Cognitive Science*, *33*(2), 147–182. https://doi.org/10.1111/j.1551-6709.2009.01008.x

Atran, S. (1998). Folk biology and the anthropology of science: Cognitive universals and cultural particulars. *Behavioral and Brain Sciences*, *21*, 547–609.

Becker, G. (1968). A theory of rational addiction. *Journal of Political Economy*, *96*, 675–700.

Berent, I., & Platt, M. (2021a). Essentialist Biases Toward Psychiatric Disorders: Brain Disorders Are Presumed Innate. *Cognitive Science*, *45*(4), e12970. https://doi.org/10.1111/cogs.12970

Berent, I., & Platt, M. (2021b). Public misconceptions about dyslexia: The role of intuitive psychology. *PLOS ONE*, *16*(12), e0259019. https://doi.org/10.1371/journal.pone.0259019

Black, T., & Dolgon, C. (2021). Zombie Sociology: Why Our Discipline Is so Susceptible to the Undead. *Critical Sociology*, *47*(3), 507–514. https://doi.org/10.1177/0896920520961808

Brickman, P., Rabinowitz, V. C., Karuza, J., Coates, D., Cohn, E., & Kidder, L. (1982). Models of helping and coping. *American Psychologist*, *37*, 368–384. https://doi.org/10.1037/0003-066X.37.4.368

Chater, N., & Loewenstein, G. (2022). The i-Frame and the s-Frame: How Focusing on Individual-Level Solutions Has Led Behavioral Public Policy Astray. *Behavioral and Brain Sciences*. https://doi.org/10.1017/S0140525X22002023

Cozzarelli, C., Wilkinson, A. V., & Tagler, M. J. (2001). Attitudes Toward the Poor and Attributions for Poverty. *Journal of Social Issues*, *57*(2), 207–227. https://doi.org/10.1111/0022-4537.00209

Davies, J. (2021). *Sedated: How Capitalism Created the Mental Health Crisis*. London: Atlantic Books.

Gerstenberg, T., Peterson, M. F., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2017). Eye-Tracking Causality. *Psychological Science*, *28*(12), 1731–1744. https://doi.org/10.1177/0956797617713053

Gottlieb, S., & Lombrozo, T. (2018). Can Science Explain the Human Mind? Intuitive Judgments About the Limits of Science. *Psychological Science*, *29*(1), 121–130. https://doi.org/10.1177/0956797617722609

Halpern, J. Y., & Pearl, J. (2005). Causes and Explanations: A Structural-Model Approach. Part II: Explanations. *The British Journal for the Philosophy of Science*, *56*(4), 889–911. https://doi.org/10.1093/bjps/axi148

Haslam, N., & Kvaale, E. P. (2015). Biogenetic Explanations of Mental Disorder: The Mixed-Blessings Model. *Current Directions in Psychological Science*, *24*(5), 399–404. https://doi.org/10.1177/0963721415588082

Iselin, M.-G., & Addis, M. E. (2003). Effects of Etiology on Perceived Helpfulness of Treatments for Depression. *Cognitive Therapy and Research*, *27*(2), 205–222. https://doi.org/10.1023/A:1023513310243

Jenkin, G. L., Signal, L., & Thomson, G. (2011). Framing obesity: The framing contest between industry and public health at the New Zealand inquiry into obesity. *Obesity Reviews*, *12*(12), 1022–1030. https://doi.org/10.1111/j.1467-789X.2011.00918.x

Jones, P., Mair, P., & McNally, R. (2018). Visualizing psychological networks: A tutorial in R. *Frontiers in Psychology*, *9*, 1742.

Kamps, F., Julian, J. B., Battaglia, P., Landau, B., Kanwisher, N., & Dilks, D. (2017). Dissociating intuitive physics from intuitive psychology: Evidence from Williams syndrome. *Cognition*, *168*, 146–153. https://doi.org/10.1016/j.cognition.2017.06.027

Kassambara, A., & Mundt, F. (2020). *factoextra: Extract and Visualize the Results of Multivariate Data Analyses*. Retrieved from https://CRAN.R-project.org/package=factoextra

Lebowitz, M. S., Ahn, W., & Nolen-Hoeksema, S. (2013). Fixable or fate? Perceptions of the biology of depression. *Journal of Consulting and Clinical Psychology*, *81*(3), 518–527. https://doi.org/10.1037/a0031730

Mair, P., Borg, I., & Rusch, T. (2016). Goodness-of-Fit Assessment in Multidimensional Scaling and Unfolding. *Multivariate Behavioral Research*, *51*(6), 772–789. https://doi.org/10.1080/00273171.2016.1235966

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Fransisco: W.H. Freeman.

Moncrieff, J. (2007). Co-opting psychiatry: The alliance between academic psychiatry and the pharmaceutical industry. *Epidemiology and Psychiatric Sciences*, *16*(3), 192–196. https://doi.org/10.1017/S1121189X00002268

Piff, P. K., Wiwad, D., Robinson, A. R., Aknin, L. B., Mercier, B., & Shariff, A. (2020). Shifting attributions for poverty motivates opposition to inequality and enhances egalitarianism. *Nature Human Behaviour*, *4*(5), 496–505. https://doi.org/10.1038/s41562-020-0835-8

Quillien, T. (2020). When do we think that X caused Y? *Cognition*, *205*, 104410. https://doi.org/10.1016/j.cognition.2020.104410

Read, J., & Cain, A. (2013). A literature review and meta-analysis of drug company–funded mental health websites. *Acta Psychiatrica Scandinavica*, *128*(6), 422–433. https://doi.org/10.1111/acps.12146

Sagarin, B. J., Ambler, J. K., & Lee, E. M. (2014). An Ethical Approach to Peeking at Data. *Perspectives on Psychological Science*, *9*(3), 293–304. https://doi.org/10.1177/1745691614528214

Shutts, K., & Kalish, C. W. (2021). Intuitive Sociology. *Advances in Child Development and Behavior*, *61*, 335–374.

Spelke, E. S. (1990). Principles of object perception. *Cognitive Science*, *14*, 29–56.

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, *10*, 89–96.

Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift Fur Tierpsychologie*, *20*, 410–433.

Weisman, K. (2022). Similarities and differences in concepts of mental life among adults and children in five cultures. *Nature Human Behavior*.

Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.