




AKADÉMIAI KIADÓ

Testing the human superorganism approach to morality

ROBERT AUNGER*  and KATIE GREENLAND

Department of Infectious Disease, London School of Hygiene and Tropical Medicine, London, England

Received: March 14, 2021 • Accepted: August 30, 2022

DOI:

[10.1556/2055.2022.00007](https://doi.org/10.1556/2055.2022.00007)

© 2022 The Author(s)

Culture and Evolution

RESEARCH ARTICLE



ABSTRACT

Current theories in moral psychology do not agree about the kinds and range of offenses that people should moralize. In this study, a new approach to defining the moral domain, Human Superorganism Theory (HSoT), is presented and tested. HSoT proposes that the primary function of moral action is the suppression of cheaters in the unusually large societies recently established by our species (i.e., human ‘superorganisms’). It suggests that a broad range of moral concerns exist beyond traditional notions of harm and fairness, including actions that inhibit functions such as group-level social control, physical and social structuring, reproduction, communication, signaling and memory. Roughly 80,000 respondents completed a web-based experiment hosted by the British Broadcasting Corporation, which elicited a suite of responses to characteristics of a set of 33 short scenarios representing the areas identified by the HSoT perspective. Results indicate that all 13 superorganism functions are moralized, while violations of scenarios falling outside this area (social customs and individual decisions) are not. Several hypotheses derived specifically from HSoT were also supported. Given this evidence, we believe this new approach to defining a broader moral domain has implications for fields ranging from psychology to legal theory.

KEYWORDS

superorganism, morality, moral domain, web experiment

INTRODUCTION

Recently, we put forward a new approach to defining the moral domain, Human Superorganism Theory (Aunger, 2017). This theory proposes that the primary function of moral action is the suppression of anti-social behavior in the unusually large societies recently established by our species. This proposition is derived from a claim, based on ‘major transition theory’ (Maynard Smith & Szathmáry, 1995) that cities and nation-states have self-organised (through a ‘major transition’) into what can be called ‘superorganisms’. Major transition theory argues that such transitions require that the interests of those within the ostensible group be brought into alignment, or be suppressed in relation to their interests at group level such that the group can function and act as a new kind of unit. This usually means that some form of top-down control over conflicts of interest must arise. Major transitions include genes clustering into chromosomes, single cells coalescing into multi-cellular organisms, and multi-cellular organisms collecting together into social groups. The claim here is that this same argument can be extended one further step, to suggest that human social groups have not only become functioning units at the level of small-scale organisations, but also have formed nascent units at the level of cities and even nation-states (Christakis & Fowler, 2009; Kesebir, 2012; Szathmáry, 2015; Wilson, Vugt, & O’Gorman, 2007). A similar argument can be put forward using gene-culture co-evolutionary theory, in which traits supporting the maintenance of human superorganismal groups evolve via cultural group selection based on a population’s tendency to punish anti-social behavior (Richerson & Boyd, 1998).

*Corresponding author.

E-mail: Robert.Aunger@lshtm.ac.uk



If it is true that human society has undergone a major transition to superorganismal groups, then – as with any major transition – such superorganisms should exhibit adaptations at the ‘group’ level, reflecting significant selection pressures shaping organisations at that level. Our argument is that morality is such an adaptation which functions as a policing system, needed to keep these large groups of unrelated individuals functioning in the absence of kinship- or reciprocity-based overlaps of interests. Essentially, our suggestion is that moral sentiments and actions have evolved to regulate very large human populations via the informal mechanism of each member potentially punishing any other for anti-social forms of activity.

Is it possible that natural selection could have produced the psychological and behavioral adaptations underlying morality during relatively recent human evolution? Walled cities first appear in the archeological record only 11,000 years ago (in Jericho and Uruk) (Gangal, Sarson, & Shukurov, 2014). Nevertheless, the kinds of rapid rates of human genetic evolution necessary to support our claim are in evidence. For example, genes helping humans adapt to life in high altitudes have reached near ubiquity in some Tibetan populations in the past 3,000 years (Yi et al., 2010), and there are significant continental differences in the frequencies of genes underlying the ability to drink milk that have arisen in the past 7,000 years (Tishkoff et al., 2007). The genetic underpinnings for psychological traits can presumably face equally strong selection pressure and therefore become characteristic of human populations in the time required to support our hypothesis. However, research on human brain evolution has focused primarily on gross features such as brain size and shape. This research has shown that the human brain only reached its current shape relatively recently, between 100,000 and 35,000 years ago (Neubauer, Hublin, & Gunz, 2018). A few studies have shown genetic underpinnings for specific human features, such as language facility (e.g., the FOXP2 gene) (Graham & Fisher, 2015), and brain developmental patterns (e.g., NOTCH2NL) (Fiddes et al., 2018). Further, twin studies reveal a substantial heritability for prosocial behavior (related to morality), with genetic factors accounting for 30–50% of individual variance (Israel, Hasenfratz, & Knafo-Noam, 2015). Finally, in lab-based game-playing, variation in punishment as a response to unfair monetary offers (a crucial morality-related behavior) has been linked to particular gene differences and levels of activity in relevant brain areas (Enge, Mothes, Fleischhauer, Reif, & Strobel, 2017; Gärtner, Strobel, Reif, Lesch, & Enge, 2018). So genetic foundations for moral psychology are apparent; it is only the selection history for specific features of that mental facility which remain to be established.

Also, as with many evolved phenomena, morality has gone through gradual development. Precursors appear in other species: proto-morality, exemplified by two-party punishment, is present in our closest relatives, the apes, which have been shown to care about not being treated fairly themselves (e.g., being given equal portions of food) (Brosnan, 2013). However, full-blown morality, demonstrated

by third-party punishment, is present only in humans (Riedl, Jensen, Call, & Tomasello, 2012). This is a willingness to take umbrage when totally unrelated individuals do not deal responsibly with each other (the self is not included in the transaction, but gets involved to promote fairness and reduce harm) (Gintis, Henrich, Bowles, Boyd, & Fehr, 2008). Also, as we will show here, the sorts of behaviors that cause moral affront get generalized to any sort of anti-social behavior in humans. Essentially, according to this approach, morality is a way of making some acts justifiably punishable or laudable, depending on whether the intention of the actor was intended to support a superorganism function or not. It is a mechanism by which large groups of unrelated individuals can guarantee themselves means to regulate their social interactions and thereby increase social cohesion.

HSOT adds to major transition theory the argument that large-scale human societies such as cities or nation-states require people to perform thirteen different kinds of social roles to keep them functioning properly: the Boundary, Control, Structure, Production, Communication, Distribution, Reproduction, Excretion, Perception, Storage, Memory, Signaling and Enforcement functions (see Table 1 for short function definitions, with example institutions for each function, and the Supplementary Materials for an extended discussion). These functions were identified by consensus from a variety of fields such as ‘living systems’ theory (J. G. Miller, 1978), ‘minimal life’ theory (Bedau, 2011; Gánti, 2003), collective animal behavior (Sumpter, 2010), and eusocial insect ecology (Hölldobler & Wilson, 2008; Seeley, 1989; Wheeler, 1911) as well by comparison to the set of organs in multi-cellular organisms like Mammals, and the Indian caste system (taken as an example of a human society in which there are explicit functional divisions). For example, the Boundary function is that of holding the elements of the superorganism together, and regulating entry and exit, like the body envelope of a multicellular organism. The Control function concerns supervision of systemic organization, a role governmental agencies often perform. Reproduction in human societies occurs through the institution of families, who live together to nurture the next generation through a period of significant dependency (see Supplementary Material 1 for a detailed description of each superorganism function). The approach thus argues that a range of concerns can be theorized as moral – even to violations of group ‘memory’ (such as cultural traditions), inappropriate communication (e.g., lying) or not protecting valuable objects from damage or theft. Performing any of these functions is activity that helps to maintain the superorganism, so the failure to perform any of them can be seen as an abdication of responsibility as a member of the group.

Further, each function is hypothesized to translate into a type of moral concern, such that actions which inhibit that function can be seen as morally reprehensible (e.g., bestiality can be seen as wasting reproductive resources better spent on producing the next generation of people for society). Each function can also be associated with specific social institutions (e.g., the economy works to distribute goods and services through the social group, fulfilling the distribution



Table 1. Areas of superorganism functionality, with examples from several levels of organisation

Area	Function	Examples		
		Cell	Organism	Human super-organism
<i>Systemic functions</i>				
<i>Boundary</i>	Hold components together; regulate entry of elements from environment	Membrane	Skin	Military, Border patrol
<i>Enforcement</i>	Internal defense	Lysosomes	Immune system	Police/Courts
<i>Structure</i>	Maintain proper (spatial) relationships among units	Cyto-skeleton/ Cytoplasm	Skeleton	Physical infrastructure (e.g., road systems)
<i>Reproduction</i>	Create similar offspring	Miosis	Sexual reproduction	Households
<i>Control (Decision-making)</i>	Coordinate/regulate the system as a whole	Chromosomes	Brain	Government bureaus
<i>Material/energy functions</i>				
<i>Ingestion</i>	Acquire energy from the environment	Chloroplasts	Mouth	Fund-seeking agency
<i>Production</i>	Transform energy into materials or provide services for use within system	Ribosome, Mitochondria	Digestive system	Factory
<i>Maintenance</i>	Ensure proper functioning of system elements (repair/replace components)	Golgi complex	Kidneys/Liver/Lungs	Repair shop/Fitness centre
<i>Storage</i>	Retain material/energy within system for later use	Vacuole	Adipose tissue	Asylum/Prison
<i>Distribution</i>	Transport material/energy between system components	Endoplasmic reticulum	Circulatory system	Economy
<i>Excretion</i>	Remove wastes from system	Membrane vesicle	Excretory system	Sanitation system
<i>Information functions</i>				
<i>Perception (Info-production)</i>	Update information on external and internal conditions	Chemical exchange	Sensory organs	Science/Technology
<i>Memory (Info-storage)</i>	Retain information for later use	Chemical states	Endo-cannabinoid system	Education
<i>Communication (Info-distribution)</i>	Transmit information between internal components	Chemical signaling (internal)	Peripheral nervous system	Media organisations
<i>Signaling (Info-excretion)</i>	Indicate state/express identity; Send messages into external environment	Chemical signaling (external)	Phenotypic markers, Speech	Diplomatic corps, Public relations organisations

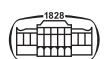
function with respect to physical and social resources – see Table 1 for examples).

The primary purpose of the current paper is to provide empirical support for this approach to understanding the moral domain. This will be achieved by showing that the areas of concern identified by HSoT are all moralized by a large sample of individuals from around the world, and that areas of concern not related to a superorganism function are not moralized. One kind of social violation which should fall outside the moral domain is violations of social folkways. These are social practices that are normative in the sense of being common but not sanctionable in the same way as moral rules. {Sumner, 2011 #12928} Such traditions, rituals,

fashions, manners, and etiquette, specify appropriate behaviors for specific situations and confer legitimacy to specific actions.

Even more generally, we expect behaviors which have no implications for group functioning (i.e., purely personal decisions) not to be moralized.

A second type of potential support for HSoT comes from tests of hypotheses derived specifically from HSoT. Two such tests will be investigated here, both derived from the central claim of Human Superorganism theory: that morality is a mechanism for promoting social cohesion in large groups of unrelated individuals. HSoT suggests that people living in larger communities will be more reliant on moral



action than alternative mechanisms for ensuring social cohesion, and therefore should be more willing to punish their fellows who act in anti-social ways. This is because genetic relatedness is lower in larger groups (meaning individuals can't rely on shared genetic interests), as is interpersonal familiarity (and hence the likelihood of a sense of reciprocal obligation). For these reasons, the need for moral action is even greater in these larger groups. In particular, what is required is a general willingness to engage in punishment of anyone who engages in a moral offense – so-called 'third party punishment' (because the individual is not directly offended themselves by the harm, but rather is upholding moral principles and ensuring group cohesion through this policing action) (Ernst Fehr & Fischbacher, 2004; Henrich et al., 2006). This can be called the 'Large Unrelated Group' hypothesis.

Types of punishment can also be distinguished in terms of their behavioral 'distance' from the original offense. *First-order* punishments are those meted out to the individual (or organization) which perpetrated the moral offense. A *second-order* punishment would be punishment of an individual who failed (in some context) to engage in their duty, as part of a superorganism, to punish a moral offender (Boyd, Gintis, Bowles, & Richerson, 2003; Fehr & Gächter, 2002). People living in larger, more complex societies should also be more willing to engage in these second-order punishments, due to their greater degree of reliance on moralistic sentiments and actions to maintain social cohesion. We will call this the 'Second-Order Punishment' hypothesis.

'Private' or 'victimless' harms (eating the 'wrong' food, 'deviant' sex, and self-abuse) are difficult to explain using the standard arguments of fairness and harm. HSoT suggests that such practices can be seen as causing harm to the body politic: these issues could be moralized because they represent violated obligations to the human superorganism. This kind of effect can be examined by looking at feelings toward an elderly, retired, terminally-ill man whose spouse and family have all died, who attempts to kill himself. Specifically, those who live in more inter-dependent groups (i.e., large societies) should feel suicide – even in such a case – is morally wrong. This will be called the 'Social Obligation Effect'.

Tests were also undertaken to establish that the dataset being used is consistent with known results concerning moral psychology and values. We performed four such tests:

Cultural Values Effect: Superorganisms can use different principles to regulate themselves (e.g., forms of government, degree of market orientation), which might be reflected in different cultural values being emphasized. One of the best-established cross-cultural findings – certainly at national level – is that collectivist countries tend toward social conformity, avoid conflict, and attempt to 'save face', while individualist countries are more domineering and emphasize self-expression (Hofstede, 2001; Nisbett & Cohen, 1996; Ting-Toomey, 2005). Those belonging to more collectivist nations should therefore score higher on reputational scenarios, while individualistic countries should be more concerned with issues of fairness and the division of spoils

(This effect is consistent with, but not definitive of, the HSoT approach.)

Action Principle: Actively causing harm is worse than causing harm through inaction, so moral outrage (wrongness, anger and disgust) should be higher on scenarios involving acts of commission rather than omission (Anderson, 2003; Cushman, Young, & Hauser, 2006; DeScioli, Christner, & Kurzban, 2011).

'Contact Principle': It feels worse to cause harm by one's own hand than to do so indirectly (via some proximal causal agency) (Cushman et al., 2006). This was tested by comparing average levels of wrongness for scenarios in which the protagonist engaged in physical contact with the victim compared to those in which no such contact takes place.

Victim Group Size Effect: A classic position in ethical thought is act-based consequentialism, which holds that acts should be judged by how much good that act produces (e.g., the 'greatest good for the greatest number' criterion associated with utilitarian philosophy) (Sinnott-Armstrong, 2012). The inverse of this would be a greater condemnation of acts which produce greater harm. Thus, moral judgments should be more severe as the number of people being victimized by an offense increases.

These hypotheses were tested by combining the scenarios or areas in different ways and performing appropriate statistical tests to compare mean scores between categories.

METHODS

This empirical testing was assayed via a web-based experiment. A large international sample of people voluntarily visited the study web-site (<http://www.bbc.co.uk/labuk/experiments/morality> – no longer active) as part of the British Broadcasting Corporation's effort to allow the general population to participate in scientific experiments. Anyone who completed the questionnaire was admitted into the study.

Participants were studied with respect to their responses about a wide variety of potentially offensive situations. We presented them with a set of 33 short verbal scenarios derived from the 13 categories of superorganism function identified by the HSoT perspective (Aunger, 2017). Several scenarios were included per area of concern (ranging from 1 to 4, plus 3 placebos) to illustrate various aspects of each superorganism function (see [Supplementary Material 2](#) for a listing of scenarios) (All within-area correlations except one have values for Cronbach's alpha >0.4; the exception being an alpha = 0.12, for Boundary, with five areas having values >0.6.) For example, the Storage function can require individuals to properly guard control over group-shared environmental resources or a group's financial reserves, so each kind of storage responsibility was incorporated into separate scenarios. Each scenario is purely behavioral, couched in terms of present action, by adults, without mention of psychological causes or consequences. For example, Scenario 2 is 'A woman burns her country's flag at a public demonstration.'



Immoral acts are typically found to be not only wrong, but disgusting or anger-inducing (J. Haidt, 2003; Prinz, 2007). People are also willing to punish immorality to different degrees – either through simple avoidance, or active aggression (Williams, Forgas, & Hippel, 2005). We therefore measured several aspects of the response to each scenario:

- Moral *judgment* about the act in question:
 - Wrongness
- Moral *feelings* toward the perpetrator:
 - Disgust
 - Anger
- Willingness to engage in moral *reactions*:
 - Avoidance
 - Punishment

In particular, each scenario was presented in written form, together with a pictorial representation of a salient moment in that situation, followed by the following set of questions:

- How **wrong** is what this person has done? [0 = ‘Not at all’ to 10 = ‘Very’]
- How **disgusted** do you feel towards this person? [0 = ‘Not at all’ to 10 = ‘Very’]
- How **angry** do you feel towards this person? [0 = ‘Not at all’ to 10 = ‘Very’]
- If you encountered this person, to what extent would you go to **avoid** interacting with them? [1 = ‘No extent at all’ to 10 = ‘A great extent’]
- Given the opportunity, how much would you **punish** this person? [1 = ‘Not at all’ to 10 = ‘Extremely’]

Scores for each kind of response to each scenario were averaged to give a mean score for each kind of response to each of the 13 functional categories. A similar design has been followed by others studying aspects of moral psychology (Gutierrez & Giner-Sorolla, 2007). Completion of the experiment required around half an hour’s time. Each respondent began by answering a number of background questions (see Questionnaire in [Supplementary Material 3](#)).

To the degree possible, the scenarios refer to everyday situations so that they do not overly tax sensibilities, can be compared with one another, and so that individuals of any kind can readily relate to them. They all require the respondent to reflect on a situation that does not involve themselves, nor members of any group to which they explicitly belong (except vaguely, by being based on situations in the context of a modern urban society, the predominant life-style of those who will be completing the survey). In this sense, the scenarios ask them about their feelings with respect to, and willingness to engage in, so-called ‘third-party punishment’, which appears to be uniquely human (Gintis, 2000) (J. A. Miller, 2017; Riedl et al., 2012). Third-party punishment occurs when someone is punished for a norm violation by a person not involved in the original infringement. Note that the two behavioral tendencies are expressed in terms of degree of effort expended (at least implicitly), that all the questions are

focused on the primary construct of interest (e.g., avoidance), and that all the response choices are phrased in similar fashion. These similarities should increase our ability to compare responses both within and between scenarios.

Pictures of each scenario were included to make them more vivid, so that respondents could better imagine themselves in that situation, and to help clarify whose action is being judged when the scenarios are a bit complicated socially or emotionally, while hopefully not introducing significant biases themselves (e.g., they were designed to be relatively ‘flat’ in valence, and generic in depiction, so that those of any culture or continent could respond similarly to them). This format should increase the ecological validity of responses to web-based stimuli. Pictures were in three colors (black, white and red), with the focal person in red (in cases where multiple people were depicted) (see [Supplementary Material 4](#) for an example, of Scenario 2).

Categorization of scenarios for each of the specific tests are as follows:

Second-order Punishment Effect: First order offense scenarios: all but 16,17, which are second order.

Action Principle: Scenarios characterized by an offense that involves the omission of an action: 7,8,10,12,14,15,17,20, 21,22,24,28,32; commission scenarios: 1,2,3,4,5,6,9,11,13,16, 18,19,23,25,26,27,29,30,31,33.

Contact Principle: Contact scenarios are 13,18,27.

Victim Group Size Effect: Scenarios, grouped by type of affected/offended party (i.e., level of organisation) are:

- Self: 7,10,24,29
- Other (known) individual(s) (e.g., friends, neighbours): 9,15,16,18,19,20,21,25,31,33
- Own family member(s): 6,13,22
- Organisation (e.g., company): 3,5,8,17
- Country/Society-at-large/Major social group (e.g., poor people): 2,4,12,14, 23,26,27,28,30,31,32
- World-at-large (e.g., ecology, immigrants): 1,11

Cultural Values Effect: To test for this effect, we used a country’s score on the individualism–collectivism dimension (IDV) developed by Geert Hofstede (available from <http://geert-hofstede.com/countries.html>), which is larger for individualist countries (Hofstede, 2001). (As IDV scores tended to become random as sample sizes from a country became smaller, we have restricted these analyses to countries represented by at least 100 respondents ($n = 76,758/79,389$; 96.7% of total survey population). These countries included: Great Britain, United States, Canada, Australia, India, Ireland, New Zealand, Germany, France, Romania, Netherlands, Sweden, Singapore, Norway, Spain, Mexico, Poland, Belgium, Brazil, Italy, Finland, Greece, South Africa, Turkey, Malaysia, Switzerland, Denmark, Philippines, Portugal, Pakistan, Republic of Serbia, Hong Kong, and Colombia.

Permission to conduct the study was obtained from the London School of Hygiene and Tropical Medicine Ethics Committee. The study also conformed to the British Broadcasting Corporation’s internal guidelines and review process. All participants consented to scientific use of their

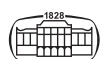


Table 2. Characteristics of survey respondents

		N	%
Gender (n = 78,538)	Male	41,777	46.2
	Female	36,761	46.8
Age (n = 79,968)	18-25	35,534	44.4
	26-30	11,220	14.0
	31-40	14,553	18.2
	41-50	10,077	12.6
	51-60	5,700	7.1
	over 60	2,884	3.6
Religion (n = 78,538)	Buddhist	1,849	2.4
	Christian	27,120	34.5
	Hindu	1,032	1.3
	Muslim	1,317	1.7
	Jewish	777	1.0
	none	40,577	51.7
	Sikh	199	0.3
other	3,456	4.4	
	Won't say	2,211	2.8
Education level (n = 78,538)	Incomplete schooling	1,223	1.6
	Schooling to GCSE-equivalent (age 16)	5,688	7.2
	Schooling to A'level equivalent (age 18)	12,030	15.3
	Vocational training	1,472	1.9
	Higher education	27,088	34.5
	Post-graduate degree	13,775	17.5
	Still in education	17,262	22.0
Social Class (n = 78,538)	Working	31,413	40.0
	Middle	38,189	48.6
	Upper Middle/Upper	8,936	11.4
Employment status (n = 78,538)	School	8,176	10.4
	University	17,449	22.2
	Full-time employment	33,907	43.2
	Part-time employment	5,741	7.3
	Self-employed	5,032	6.4
	Homemaker (stay-at-home parent)	1,539	2.0
	Unemployed	4,122	5.3
	retired	2,572	3.3
Type of work (n = 78,538)	prof/tech	28,258	36.0
	higher admin	4,638	5.9
	clerical	7,465	9.5

(continued)

Table 2. Continued

		N	%
	sales	4,577	5.8
	service	4,353	5.5
	skilled	3,493	4.5
	semi-skilled	1,700	2.2
	unskilled	3,059	3.9
	farm	299	0.4
	other	20,696	26.4

Denominators vary. Overall, 80,199 individuals began the survey and 78,357 completed all questions.

responses prior to completing the survey questionnaire by registering with the BBC Lab UK website (<http://www.bbc.co.uk/labuk/experiments/test-your-morality>). The study dataset has been archived with the UK Data Archive at the University of Essex (www.ukdataservice.ac.uk) for use by other scholars.

RESULTS

Sample characteristics

Between November 2011 and July 2012 a total of 80,199 individuals initiated the BBC survey (see Table 2), of whom 78,357 completed all scenario questions (a 97.8% completion rate). Sixty-seven percent of these respondents were British or Irish, 18% were American, 3% Canadian and 2% Australian; these 4 countries thus encompassed 90% of the total sample. However, at least one person from 202 different countries completed the survey. There was a preponderance of young people as well, with 58% of the sample being under 30.

Extent of the moral domain

Our first concern is to determine whether HSoT correctly identifies the range and extent of moral concerns. Figure 1 shows that the moral domain includes all of the issues expected by HSoT, but not the placebo scenarios, which exhibit significantly lower values (mean wrongness score for the customary 'hat' placebo = 1.27, for the 'golf' scenario = 1.38, and for the "suicide" scenario = 2.35; mean wrongness score for the non-placebo scenarios = 6.61; $P < 0.001$ for the comparison of each placebo with non-placebo scores).

HSoT-specific tests

Large Unrelated GroupEffect: We find that respondents living in larger communities express a stronger willingness to punish moral offenses than those living in smaller communities (one-way Anova: $F(4, 78,352) = 15.14, P = 0.001$). Post-hoc comparisons using the Bonferroni test found that the differences lie between all sizes of community and those living in a metropolis ($P < 0.001$ for all associations) and between



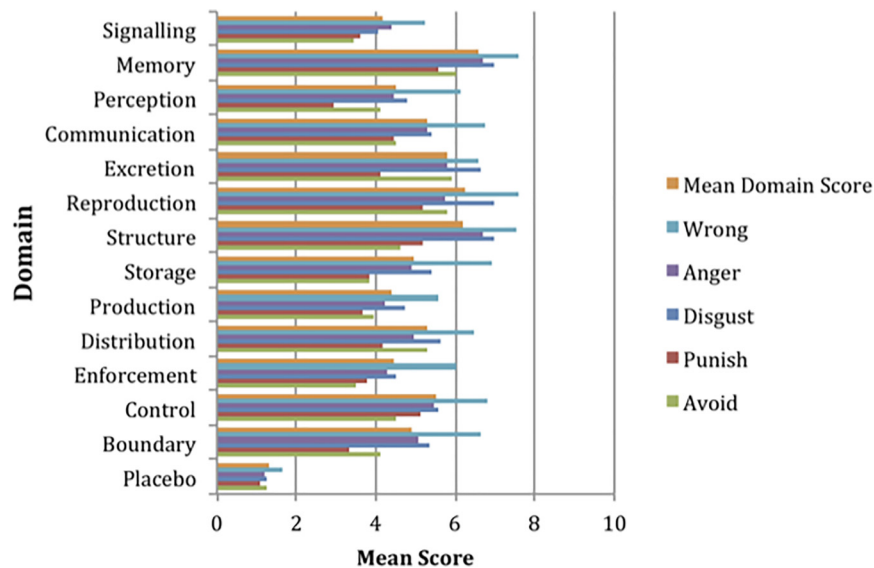


Fig. 1. Mean scores for each moral category for all measures. Scenarios comprising each category were scored from 1 – 10 on five factors: how wrong the offense was; whether the actor in the scenario angered or disgusted the respondent; and whether the respondent wished to avoid or punish the actor. Scores per category are the mean score given across all scenarios comprising the category

villages and large cities ($P = 0.029$) (see [Supplementary Material 5](#)). 10.8% of respondents reported living in a village, 28.3% in a town, 23.6% in a city, 21.7% in a large city and 15.7% in a metropolis.

Second-order Punishment Effect: Willingness to engage in second-order punishment (i.e., punishing someone who has failed to punish the original offender) increases significantly with self-reported community size (one-way Anova: $F(4, 78,362) = 12.94, P < 0.001$); the significant differences in particular lie between all smaller community sizes and metropolises (Bonferroni test $P < 0.001$ for all comparisons) and for the comparison between villages and large cities ($P = 0.025$).

Social Obligation Effect: The perceived wrongness of the act and willingness to punish a person thinking of committing suicide increases among those living in larger communities (wrongness: $F(4, 78,353) = 40.08, P < 0.001$; punishment; $F(4, 78,353) = 14.01, P < 0.001$).

Replication studies

Cultural Values Effect: We first compared a country's IDV score with its value for the difference between reputational and fairness scenarios (i.e., average Memory and Communication – average Distribution values). The correlation for mean response score with IDV was significantly negatively correlated ($r = -0.18; P < 0.001$), indicating that those living in more collectivist countries judged reputational violations more severely than those affecting fairness. We also find that, as expected, wrongness scores for the nepotism scenario correlate negatively with IDV scores by country ($r = -0.53, P < 0.001; N = 33$).

Action Principle: Mean scores for wrongness (omission: 4.85, 95%CI 4.84–4.86; commission: 6.99 (95%CI 6.98–7.00); $P < 0.001$), anger (omission: 3.98, 95%CI 3.97–3.99;

commission: 5.39 (95%CI 5.38–5.40); $P < 0.001$) and disgust (omission: 4.01, 95%CI 4.00–4.02; commission: 6.04 (95%CI 6.02–6.05); $P < 0.001$) were significantly higher for scenarios involving action than inaction.

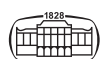
Contact Principle: The mean score for wrongness across scenarios causing physical harm was higher than scores for scenarios without (7.80 (95%CI 7.80–7.82) vs. 5.98 (5.97–5.99); $P < 0.001$).

Victim Group Size Effect: Offenses against larger groups were in fact judged more severely – that is, when victims were known individuals or family, punishment and wrongness were lower than when victims were whole societies or the world-at-large (although the trend is stronger for wrongness than punishment; see figure as [Supplementary Material 6](#)).

DISCUSSION

Our main result shows that moralizing responses to scenarios extended far beyond the concerns of fairness, harm and care traditionally covered by moral psychology to include all of the functions associated with human superorganisms. For example, HSoT argues that people can moralize other people's relationships with information or the environment, and thus that moral judgments are not restricted to social life ([Supplementary Material 7](#) shows in some detail how HSoT compares to other recent approaches vis a vis explaining the domain of moral situations.)

On the other hand, violation of a social custom (i.e., a woman not wearing a hat to a wedding) is less emotionally and behaviorally charged than superorganism-based violations. Social customs (arbitrary rules to increase social cohesion) and actions with repercussions that do not affect social functioning in any way do not elicit moralistic



responses, and so fall outside the moral domain, as expected if morality is about *regulating* social activity, not just *coordinating* it. It is not obvious that the moral line should occur at this juncture – why should failure to wear a hat to an important social event or lack of follow-through on a planned activity be judged less severely than a verbal act (lying) or harming a book? In effect, these results indicate that people moralize issues in just those areas covered by HSoT.

We also tested the proposition that people living in larger societies report engaging in more third-party punishment (i.e., punishment of those who have morally offended against someone else in your group). This replicates a previous study among a wider variety of populations, ranging from small-scale hunter-gather groups to modern city-dwellers (Marlowe & Berbesque, 2008), by using in this case a more comparable set of people all living in ‘modern’ countries but in communities of differing sizes. Further, people exhibit stronger moral sentiments against suicide if they live in superorganism-sized groups, implying that those who live in larger groups (and hence depend to a greater degree on the ‘kindness of strangers’ to accomplish everyday goals) feel more strongly that removing oneself from the group is wrong (Differences in responses to scenarios are either facultative – a function of the respondent’s life experience, reflecting the social environment in which they live) a consequence of their emotional demeanor at the time of response, or reflect more pro-social personalities, as some individuals preferentially migrate to live in larger agglomerations. In any case, it doesn’t require natural selection on generations of genetic lineages of city residents for moral concern, which would be unreasonable, given that few generations have passed since the origin of metropolises, and the ability of people to migrate in and out of societies of different sizes.)

Several oft-replicated effects associated with morality were also observed in the present dataset. These tests related to finding it more difficult to engage in direct punishment, or finding a violation involving more people to be more offensive, and that there is expected variation between countries along the individualist-collectivist dimension in the levels of moral concern of different kinds. All of these results lend greater credence to the novel results derived from the same dataset.

Study limitations. First, as the study is based on a web-survey, it may have low ecological validity; we fully support recent calls for moral psychology to move closer to the study of actual behaviors (Hofmann, Wisneski, Brandt, & Skitka, 2014; Teper, Zhong, & Inzlicht, 2015). Second, the web test did not allow for the possibility that there can be positive aspects to moral behavior – e.g., rewarding of good behavior or so-called ‘prescriptive’ morality (i.e., the more discretionary commendation of socially positive acts such as benevolence, charity, and generosity), which precluded testing the existence or extent of a positive moral domain (Janoff-Bulman, Sheikh, & Baldacci, 2008). In addition, some people could have viewed some scenarios as morally righteous (e.g., flag-burning as protest in country with a despicable government), although this was not allowed as a response. Third, the behavioral options are limited to avoidance and punishment; the test does not allow for respondents’ desires to engage in

rehabilitation or ‘talk therapy’ with the perpetrators (which might be more appropriate for some protagonists, such as the desperate gambler). Fourth, although the sample is large and multi-cultural in nature, it remains heavily biased toward educated individuals with internet access from developed English-speaking countries, so claims about universality must be tempered. Further empirical investigation would be needed to address questions about cross-cultural issues.

CONCLUSION

Current theories in moral psychology and related sciences do not agree about the extent of the moral domain (e.g. Curry, 2016; Haidt, 2007). In this study, a new approach to defining the moral domain, Human Superorganism Theory (HSoT) (Aunger, 2017), has been presented and tested. It is based on a single theoretical claim: that morality arose as a psychological mechanism to reduce social cheating in large groups of unrelated individuals (such as cities and nation-states), which can be considered human ‘superorganisms’. This claim is grounded in fundamental macroevolutionary theory, and compared to alternative approaches, is more parsimonious, while illuminating a larger range of moral behaviors than the other approaches. As a more basic, powerful approach, it should be preferred.

A number of empirical tests have supported this approach. The first kind of evidence shows that the moral domain extends to exactly the dimensions expected by HSoT. *All* of the HSoT areas are moralized, including those not previously considered by the broadest current theories, Moral Foundations Theory (J. Haidt & Joseph, 2007) and Morality-as-Cooperation Theory (Curry, 2016; Curry, Jones Chesters, & Van Lissa, 2019). These are functions such as Perception, Memory, Communication, Production, Storage, and Control. Further, situations involving information (not considered by other approaches) are judged just as severely as ‘concrete’ ones. That is, people report being willing to punish ‘talking’ violations (e.g., reputation disparagement) just as severely as violations against bodies (e.g., rape) or resources (e.g., theft). This implies that morality is not just about the interpersonal relations of harm, care and fair-dealing, but just as much about a wider range of individual responsibilities to the public sphere.

By contrast, violations of social customs and failures to achieve personal goals – neither of which are central to superorganism functioning – are not judged in the same way. This constitutes strong evidence that HSoT more accurately identifies the total range of moral feeling and behavior than other current approaches.

Further, hypotheses derived specifically from Human Superorganism theory’s central claim – that moral policing is required by everyone in the large social groups that characterize human populations – were substantiated empirically. In particular, willingness to engage in moral punishments is most often reported when people live in the largest (and hence most genetically diffuse and personally unfamiliar) groups. Second, people living in the largest groups are also most willing to punish those who haven’t punished moral



offenders (so-called ‘second-order’ punishment), further ensuring that social cohesion is achieved through moral policing, and reducing the likelihood of additional defections from these social responsibilities. Third, those living in larger social agglomerations are more likely to feel it is anti-social to attempt suicide.

Finally, that the dataset used to generate these conclusions is not idiosyncratic is suggested by our ability to replicate well-known findings in moral psychology, such as the ‘Action Principle’ and ‘Contact Principle’ (Cushman et al., 2006). For all of these reasons, we believe that the Human Superorganism approach to understanding the moral domain has been shown to be highly predictive of a broad range of moral sentiments and behaviors, and hope that those studying morality will find it useful in their future work.

Funding: No funding was required for this research.

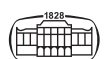
Declaration of interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

SUPPLEMENTARY MATERIAL

Supplementary data to this article can be found online at <https://doi.org/10.1556/2055.2022.00007>.

REFERENCES

- Anderson, C. J. (2003). The psychology of doing nothing: Forms of decision avoidance result from reason and emotion. *Psychological Bulletin*, 129, 139–167. Retrieved from <http://psycnet.apa.org/journals/bul/129/1/139/>.
- Aunger, R. (2017). Moral action as cheater suppression in human superorganisms. *Frontiers in Sociology*, 2, 2.
- Bedau, M. A. (2011). A functional account of degrees of minimal chemical life. *Synthese*. <https://doi.org/10.1007/s11229-011-9876-x>.
- Boyd, R., Gintis, H., Bowles, S., & Richerson, P. J. (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences of the United States of America*, 100, 3531–3535. Retrieved from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC152327/pdf/pq0603003531.pdf>.
- Brosnan, S. (2013). Justice-and fairness-related behaviors in nonhuman primates. *Proceedings of the National Academy of Sciences of the United States of America*, 110(Suppl 2), 10416–10423.
- Christakis, N. A., & Fowler, J. H. (2009). *Connected: The surprising power of our social networks and how they shape our lives*. New York: Little, Brown & Company.
- Curry, O. S. (2016). Morality as cooperation: A problem-centred approach. In T. K. Shackelford, & R. D. Hansen (Eds.), *The evolution of morality* (pp. 27–51). New York: Springer.
- Curry, O. S., Jones Chesters, M., & Van Lissa, C. J. (2019). Mapping morality with a compass: Testing the theory of ‘morality-as-cooperation’ with a new questionnaire. *Journal of Research in Personality*, 78, 106–124.
- Cushman, F., Young, L., & Hauser, M. D. (2006). The role of reasoning and intuition in moral judgments: Testing three principles of harm. *Psychological Science*, 17, 1082–1089. Retrieved from <http://pss.sagepub.com/content/17/12/1082.long>.
- DeScioli, P., Christner, J., & Kurzban, R. (2011). The omission strategy. *Psychological Science*, 22(4), 442–446.
- Enge, S., Mothes, H., Fleischhauer, M., Reif, A., & Strobel, A. (2017). Genetic variation of dopamine and serotonin function modulates the feedback-related negativity during altruistic punishment. *Scientific Reports*, 7(1), 2996.
- Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior*, 25(2), 63–87.
- Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415, 137–140.
- Fiddes, I. T., Lodewijk, G. A., Mooring, M., Bosworth, C. M., Ewing, A. D., Mantalas, G. L., ... Rosenkrantz, J. L. (2018). Human-specific NOTCH2NL genes affect Notch signaling and cortical neurogenesis. *Cell*, 173(6), 1356–1369 e1322.
- Gangal, K., Sarson, G. R., & Shukurov, A. (2014). The near-eastern roots of the neolithic in South Asia. *Plos One*, 9(5), e95714.
- Gánti, T. (2003). *The principles of life*. New York: Oxford University Press.
- Gärtner, A., Strobel, A., Reif, A., Lesch, K.-P., & Enge, S. (2018). Genetic variation in serotonin function impacts on altruistic punishment in the ultimatum game: A longitudinal approach. *Brain and Cognition*, 125, 37–44.
- Gintis, H. (2000). Strong reciprocity and human sociality. *Journal of Theoretical Biology*, 206, 169–179. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0022519300921118>.
- Gintis, H., Henrich, J., Bowles, S., Boyd, R., & Fehr, E. (2008). Strong reciprocity and the roots of human morality. *Social Justice Research*, 21(2), 241–253.
- Graham, S. A., & Fisher, S. E. (2015). Understanding language from a genomic perspective. *Annual Review of Genetics*, 49, 131–160.
- Gutierrez, R., & Giner-Sorolla, R. (2007). Anger, disgust, and presumption of harm as reactions to taboo-breaking behaviors. *Emotion*, 7, 853–868.
- Haidt, J. (2003). The moral emotions. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 852–870). Oxford: Oxford University Press.
- Haidt, J. (2007). The new synthesis in moral psychology. *Science*, 316, 998–1002.
- Haidt, J., & Joseph, C. (2007). The moral mind: How 5 sets of innate intuitions guide the development of many culture-specific virtues, and perhaps even modules. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The Innate Mind* (Vol. 3, pp. 367–391). New York: Oxford.
- Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., ... Henrich, N. (2006). Costly punishment across human societies. *Science*, 312(5781), 1767–1770.
- Hofmann, W., Wisneski, D. C., Brandt, M. J., & Skitka, L. J. (2014). Morality in everyday life. *Science*, 345(6202), 1340–1343.
- Hofstede, G. (2001). *Culture’s Consequences: Comparing values, behaviors, institutions, and organizations across nations*. Thousand Oaks, CA: Sage.



- Hölldobler, B., & Wilson, E. O. (2008). *The superorganism: The beauty, elegance, and strangeness of insect societies*. New York: W.W. Norton & Co.
- Israel, S., Hasenfratz, L., & Knafo-Noam, A. (2015). The genetics of morality and prosociality. *Current Opinion in Psychology*, 6, 55–59.
- Janoff-Bulman, R., Sheikh, S., & Baldacci, K. G. (2008). Mapping moral motives: Approach, avoidance, and political orientation. *Journal of Experimental Social Psychology*, 44(4), 1091–1099.
- Kesebir, S. (2012). The superorganism account of human sociality: How and when human groups are like beehives. *Personality and Social Psychology Review*, 16, 233–261.
- Marlowe, F. W., & Berbesque, J. C. (2008). More 'altruistic' punishment in larger societies. *Proceedings of the Royal Society B-Biological Science*, 275, 587–590.
- Maynard Smith, J., & Szathmáry, E. (1995). *The major transitions in evolution*. Oxford: Oxford University Press.
- Miller, J. G. (1978). *Living systems*. New York: McGraw Hill.
- Miller, J. A. (2017). *Chimpanzee third party behavior: Insights into the evolution of human conflict management*. (PhD). Washington D.C: The George Washington University.
- Neubauer, S., Hublin, J.-J., & Gunz, P. (2018). The evolution of modern human brain shape. *Science Advances*, 4(1), eaao5961.
- Nisbett, R. E., & Cohen, D. (1996). *Culture of honor: The psychology of violence in the South*. Boulder, CO: Westview Press.
- Prinz, J. J. (2007). *The emotional construction of morals*. Oxford: Oxford University Press.
- Richerson, P., & Boyd, R. (1998). The evolution of human ultrasociality. In I. Eibl-Eibesfeldt, & F. Salter (Eds.), *Indoctrinability, ideology, and warfare*. New York: Berghahn Books.
- Riedl, K., Jensen, K., Call, J., & Tomasello, M. (2012). No third-party punishment in chimpanzees. *Proc Natl Acad Sci U S A*, 109(37), 14824–14829. <https://doi.org/10.1073/pnas.1203179109>.
- Seeley, T. D. (1989). The honey bee colony as a super-organism. *American Scientist*, 77, 546–553.
- Sinnott-Armstrong, W. (2012). Consequentialism. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy*.
- Sumpter, D. J. T. (2010). *Collective animal behavior*. Princeton: Princeton University Press.
- Szathmáry, E. (2015). Toward major evolutionary transitions theory 2.0. *Proceedings of the National Academy of Sciences*, 112(33), 10104–10111.
- Teper, R., Zhong, C.-B., & Inzlicht, M. (2015). How emotions shape moral behavior: Some answers (and questions) for the field of moral psychology. *Social and Personality Psychology Compass*, 9(1), 1–14. <https://doi.org/10.1111/spc3.12154>.
- Ting-Toomey, S. (2005). The matrix of face: An updated face-negotiation theory. In W. B. Gudykunst (Ed.), *Theorizing about intercultural communication* (pp. 71–92). Thousand Oaks, CA: Sage.
- Tishkoff, S. A., Reed, F. A., Ranciaro, A., Voight, B. F., Babbitt, C. C., Silverman, J. S., ... Osman, M. (2007). Convergent adaptation of human lactase persistence in Africa and Europe. *Nature genetics*, 39(1), 31.
- Wheeler, W. M. (1911). The ant-colony as an organism. *Journal of Morphology*, 22, 307–325.
- Williams, K. D., Forgas, J. P., & Hoppel, W. v. (2005). *The social outcast: Ostracism, social exclusion, rejection, and bullying*. London: Psychology Press.
- Wilson, D. S., Vugt, M. V., & O'Gorman, R. (2007). Multilevel selection theory and major evolutionary transitions: Implications for psychological science. *Current Directions in Psychological Science*, 17, 6–9.
- Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z. X. P., Pool, J. E., ... Korneliussen, T. S. (2010). Sequencing of 50 human exomes reveals adaptation to high altitude. *Science*, 329(5987), 75–78.

