



**HAL**  
open science

# Moral conviction, political polarisation, and susceptibility to misinformation

Antoine Marie

► **To cite this version:**

Antoine Marie. Moral conviction, political polarisation, and susceptibility to misinformation. Philosophy. Université Paris Cité, 2021. English. NNT : 2021UNIP5079 . tel-04300037

**HAL Id: tel-04300037**

**<https://theses.hal.science/tel-04300037>**

Submitted on 22 Nov 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**Université de Paris**

*Ecole Doctorale 474 Frontières de l'Innovation en Recherche et Education*

*Université de Paris*

*Sciences cognitives, psychologie, linguistique, philosophie de la pensée*

*Laboratoire Cognitions Humaines et Artificielle (CHArt), EA 4004, Université Paris 8,*

*Université Paris-Est Créteil, EPHE*

**“MORAL CONVICTION, POLITICAL POLARISATION, AND  
SUSCEPTIBILITY TO MISINFORMATION”**

Par Antoine Marie

Thèse de doctorat de sciences cognitives

Dirigée par Jean Baratgin

Présentée et soutenue publiquement le 25 mars 2021

Devant un jury composé de :

Jean BARATGIN, Maître de conférences en psychologie cognitive, Université Paris 8  
Vincennes Saint-Denis, Directeur

Gérald BRONNER, Professeur, Laboratoire Interdisciplinaire des Energies de Demain,  
Université de Paris, membre invité

Sabine GUERAUD, Maîtresse de Conférences en psychologie cognitive, Université Paris 8  
Vincennes Saint-Denis

Hugo MERCIER, Directeur de recherche CNRS, Rapporteur

Cathal O'MADAGAIN, Assistant Professor of Cognitive Science, Université Mohammed VI  
Polytechnique, Rapporteur

Gloria ORIGGI, Directrice de recherche CNRS, Ecole des Hautes Etudes en Sciences  
Sociales, Examinatrice

Brent STRICKLAND, Chargé de recherche CNRS, Examinateur



# Table of Contents

ACKNOWLEDGMENTS.....	5
ABSTRACT.....	8
RÉSUMÉ.....	10
RÉSUMÉ SUBSTANTIEL.....	13
CHAPTER 1: GENERAL INTRODUCTION .....	19
OUR INTUITIONS DO NOT PREDISPOSE US TO FOLLOW THE FACTS .....	23
THE EVOLUTIONARY COGNITIVE SCIENCES (IN A NUTSHELL) .....	24
“THE COGNITIVE FOUNDATIONS OF MISINFORMATION ON SCIENCE” .....	27
MORAL CONVICTIONS AND TRADE-OFF INSENSITIVITY .....	29
“MORAL CONVICTION PREDICTS SHARING PREFERENCE FOR POLITICALLY CONGRUENT HEADLINES” .....	32
“INTENTIONS MATTER A LOT, AND EFFICIENCY LITTLE, IN FOLK JUDGMENTS OF POLICY DECISIONS” .....	35
CHAPTER 2: THE COGNITIVE FOUNDATIONS OF MISINFORMATION ON SCIENCE .....	38
INTRODUCTION .....	38
HOW COMMUNICATION WORKS .....	39
HUMAN MINDS ARE MADE OF INTUITIVE MECHANISMS .....	40
WE TRUST FAMILIAR SOURCES WE SEE AS BENEVOLENT AND COMPETENT .....	41
WE TEND TO BE PARTIAL IN FAVOR OF THE BELIEFS WE AND OUR GROUPS HOLD .....	42
WE HAVE NAÏVE THEORIES ABOUT MANY THINGS .....	43
RUNNING ON EMPTY: THE ILLUSION OF UNDERSTANDING .....	44
INFORMATIONAL MARKET WORKERS KNOW MANY BIASES AND EXPLOIT THEM .....	45
RELEVANCE = REVENUE: THE LURE OF EXAGGERATED CLAIMS.....	45
TURNING SCIENTIFIC RESEARCH INTO TALES .....	46
INFORMATION FITTING PRECONCEIVED NOTIONS CAN SPREAD LIKE WILDFIRE .....	47
CONSPIRACY THEORIES.....	48
ADDRESSING MISUNDERSTANDINGS: RECOMMENDATIONS TO SCIENTISTS .....	49
CONCLUSION .....	52
REFERENCES .....	54
CHAPTER 3: MORAL CONVICTION INCREASES SHARING PREFERENCE FOR POLITICALLY CONGRUENT HEADLINES .....	56
ABSTRACT:.....	57
INTRODUCTION .....	58
EXPERIMENTS 1-3: WILLINGNESS TO SHARE TO FRIENDS .....	61
Method .....	61
Results.....	65
EXPERIMENT 4: MOTIVATIONS TO SHARE NEWS HEADLINES .....	68
Method .....	69
Results.....	70
Discussion of Experiments 1-4.....	71
EXPERIMENT 5: POLITICALLY CONGENIAL VS. UNCONGENIAL AUDIENCE .....	73
Method .....	74
Results.....	74
EXPERIMENT 6: ANONYMOUS VS. PERSONAL ACCOUNT .....	76
Method .....	76
Results.....	79
EXPERIMENTS 7 AND 8: INTERVENTION MESSAGE .....	80
Method .....	81
Results.....	82
GENERAL DISCUSSION .....	85
ELECTRONIC SUPPLEMENTARY MATERIALS (ESM) .....	89
Coding of predictors in regression tables.....	90
Experiment 1: Willingness to share true headlines, 4-point DV .....	91
Experiment 2: Willingness to share true headlines, dichotomous DV.....	93
Experiment 3: Willingness to share false headlines, dichotomous DV .....	95
Experiment 4: Motivations for sharing, 4-point DV.....	97

<i>Experiment 5: Sharing to politically congenial vs. uncongenial audience, dichotomous DV</i> .....	100
<i>Experiment 6: Willingness to pay for sharing from anonymous vs. personal account, 0-100 DV</i> .....	102
<i>Experiment 7: Intervention message on real news items</i> .....	104
<i>Experiment 8: Intervention message on false news items</i> .....	106
<i>Choice of the 4 controversial issues</i> .....	108
<i>Real political news items</i> .....	109
Raters' perceptions of real political news items' slant .....	114
<i>Real non-political news items</i> .....	115
<i>False political news items</i> .....	117
Raters' perceptions of false political news items' slant.....	121
<i>Political attitude on issue (continuous)</i> .....	124
<i>Moral importance of issue (dichotomous)</i> .....	125
<i>Vignettes not provided in main text</i> .....	126
<i>Attention check 'videogame'</i> .....	129
<b>CHAPTER 4: INTENTIONS MATTER A LOT, AND EFFICIENCY LITTLE, IN FOLK JUDGMENTS OF POLICY DECISIONS</b> .....	<b>131</b>
<b>ABSTRACT:</b> .....	<b>131</b>
<b>INTRODUCTION</b> .....	<b>133</b>
<b>EXPERIMENTS 1-4</b> .....	<b>136</b>
<i>Method</i> .....	136
<i>Results</i> .....	140
<i>Discussion</i> .....	144
<b>EXPERIMENT 5</b> .....	<b>145</b>
<i>Method</i> .....	145
<i>Results</i> .....	147
<b>GENERAL DISCUSSION</b> .....	<b>152</b>
<b>ELECTRONIC SUPPLEMENTARY MATERIALS (ESM)</b> .....	<b>156</b>
<i>Experiments 1-4</i> .....	157
Vignettes.....	157
Main multiple regression models on aggregated data .....	161
Main multiple regression models on aggregated data, data from 1st condition only.....	163
Additional simple regressions models of commendability on moral commitment:.....	164
<i>Experiment 5</i> .....	165
Main multiple regression models of support for policy judgments .....	165
Additional simple regression models of support on moral commitment.....	167
<b>CHAPTER 5: CONCLUSION AND FUTURE AVENUES</b> .....	<b>169</b>
<b>CONCLUSION</b> .....	<b>169</b>
<i>Misleading intuitions</i> .....	169
<i>Political polarization of factual beliefs</i> .....	173
<i>Social motivations to communicate and believe</i> .....	175
<i>From diverging priors to political hostility</i> .....	176
<b>MORAL CONVICTIONS AND SELECTIVE COMMUNICATION: FUTURE AVENUES</b> .....	<b>178</b>
<b>INTENTIONS AND EFFICIENCY IN JUDGMENTS OF POLICIES: FUTURE AVENUES</b> .....	<b>182</b>
<b>EXPLORING THE FUNCTIONS OF DEMONIZING POLITICAL OPPONENTS</b> .....	<b>186</b>
<b>APPENDIX</b> .....	<b>191</b>
<b>MORAL RIGIDITY AS A PROXIMATE FACILITATOR OF GROUP COHESION AND COMBATIVENESS</b> .....	<b>191</b>
<b>GENERAL REFERENCES</b> .....	<b>197</b>

# Acknowledgments

Difficile d'avoir des paroles originales lorsqu'il s'agit de s'acquitter du rituel des remerciements. Cette thèse est certes passée un peu plus vite que prévu, mais je ne dirais pas qu'elle a filé à la vitesse que j'avais d'abord imaginé.

Je commencerai par remercier mon directeur, Jean Baratgin, de m'avoir accueilli en psychologie en 2017 lorsque j'étais alors un étudiant en sociologie aux convictions trop naturalistes, et à l'intérêt pour la cognition déjà trop développés, pour que les « sciences sociales standards » puissent beaucoup plus longtemps constituer ma demeure intellectuelle.

Merci ensuite à Brent Strickland, mon co-encadrant, et compagnon de route principal au cours de cette thèse, de m'avoir transmis un certain nombre de ficelles du métier, d'avoir dirigé mes recherches expérimentales, et de m'avoir aidé à financer la fin de ma thèse au travers d'un séjour d'enseignement à l'université Mohammed VI Polytechnique, à Ben Guérir (Maroc).

Je suis infiniment redevable, ensuite, à la petite bande des Naturalistes de l'ENS et de La Sorbonne : Aurélien Allard, Sacha Yesilaltay aka Sacha Altay, Charlotte Barot, Matthias Michel, Josselin Bonnamour, Joffrey Fuhrer, Loïa Lamarque, Léo Fitouchi, Laurent Cordonier. Merci aussi aux comparses Can Konuk, Charles Arnal, Slawa Loev, Marco Inchingolo, Céline BL, Martin Fortier, Mélusine Boon Falleur, Hualin Xiao, Mauricio Martins, Hugo Trad. Tous ont contribué à ce que ces cinq dernières années (à la louche), et en particulier mon passage au 29 rue d'Ulm, aient été à la fois une période d'émulation intellectuelle intense, et d'amitiés magnifiques qui je l'espère, dureront longtemps.

Merci ensuite à tous les chercheurs qui ont exercé une influence décisive sur ma pensée. Parmi eux, l'école sperbérienne en sciences sociales cognitives—Dan Sperber, Hugo Mercier et Nicolas Baumard en particulier—mais aussi Jon Haidt, Scott Atran, Philip Tetlock,

Steven Pinker, Philip Fernbach et Steven Sloman, Pascal Boyer, et Michael Bang Petersen, mon futur patron.

Merci aussi aux encouragements à croire en mes intuitions théoriques et évolutionnaires de Hammad Sheikh, qui m'a aidé à me sentir un peu moins comme un étudiant, et un peu plus comme un chercheur.

Spéciale dédicace à Manvir Singh, génial anthropologue hipster bariolé et enturbané.

Enfin, ma gratitude va également à mes parents, Anne-Sophie et Jean-Paul, et à mes frères et sœurs, Lucas et Julia, à qui je dois l'environnement familial sain, aimant et soutenant dont j'ai bénéficié tout au long de ma vie.

Cette thèse a été financée par une bourse IIFR (Institut Innovant de Formation par la Recherche), Université de Paris.





# Abstract

In the political arena, rational decision making should be based on accurate assessments of the costs and benefits associated with technologies and policies—from GMO foods to nuclear power to global warming and gun control—lest immense opportunities may be missed. However, human minds are home to many cognitive biases, many of which likely evolved by natural selection. Those biases distort information processing and transmission, standing in the way of people indexing their beliefs on the scientific consensus when it exists, and maintaining nuanced representations of complex issues when the evidence is equivocal. This thesis explores various cognitive biases contributing to this outcome through one theoretical and two empirical papers.

“The cognitive foundations of misinformation on science” (Chapter 2) draws on extant research in cultural evolution, science communication, and evolutionary psychology to review some of the main psychological mechanisms involved in causing widespread misbeliefs on policy-relevant scientific and technological issues. Among other culprits, the article highlights the reconstructive character of human communication and the role played by expectations of relevance in exaggerating factual claims; the myside bias and our tendency to moralize and politicize factual topics; our preference for homophily; our susceptibility to threatening information and conspiratorial thinking; and the illusion of understanding.

“Moral conviction predicts sharing preference for politically congruent headlines” (Chapter 3) explored, through 8 experiments, what type of political information US participants decide to share on simulated social media on controversial topics such as gun control or abortion. In analogy with our myside bias when processing information, we found that participants have a sharing preference for politically congruent news stories which increased with the moral importance of the issue—whether the news was true or false. We also found that this sharing preference was little swayed by manipulations of the audience composition,

anonymously of the account from which sharing is done, and exposition to an intervention message warning against political bias. Perceived accuracy and usefulness for one's political goals were among the main predictors of sharing. We suspect that a tendency to do selective communication may accelerate discrepancies in prior beliefs across political subcultures, undermining the ability of liberals and conservatives' partisans to find common ground on important policy issues.

Across 5 experiments run on French subjects, "Intentions matter a lot, and efficiency little, in folk judgments of policy decisions" (Chapter 4) found that laypeople spontaneously prefer altruistically motivated policies that achieve little at a high cost over selfishly motivated but highly efficient and profitable policies. This preference appeared to be driven by a combination of low responsiveness to differences in efficiency expressed in numeric format—a tendency magnified by a participant's degree of moral conviction on the issue—and of high sensitivity to the intentions driving the person implementing the policy.

Taken together, this work contributes to portray humans as a moralistic and paranoid species, wired for the righteous defense of political causes and moral principles more than for the pursuit of what may effectively promote collective welfare.

## Résumé

Dans la vie publique, la prise de décision rationnelle doit être basée sur des évaluations précises des coûts et des bénéfices associés aux technologies et aux politiques—qu’il s’agisse des OGM, de l’énergie nucléaire civile, de l’implication de l’homme dans le réchauffement climatique ou du droit au port d’arme—faute de quoi nous risquons de passer à côté d’immenses opportunités. Malheureusement, l’esprit humain est le siège de biais cognitifs, dont beaucoup ont probablement évolué. Ces biais, qui interviennent tant au niveau du traitement de l’information que de sa transmission, rendent difficile pour les gens d’avoir des croyances relatives aux coûts et aux bénéfices des politiques qui soient correctes. Cette thèse explore quelques-uns des biais cognitifs contribuant à ce résultat, à travers 1 article théorique et 2 articles empiriques.

“The cognitive foundations of misinformation on science” (Chapitre 2) s’appuie sur les recherches en évolution culturelle, en communication scientifique et en psychologie évolutionnaire pour examiner certains des principaux mécanismes à l’origine de fausses croyances répandues sur les questions scientifiques et technologiques. Entre autres biais, l’article met en avant le caractère reconstituteur de la communication humaine, et le rôle joué par les attentes de pertinence qui poussent les locuteurs à l’exagération ; le biais de confirmation et notre tendance à politiser les sujets factuels ; notre préférence pour l’homophilie ; notre susceptibilité aux informations menaçantes et à la pensée conspirationniste ; ou encore l’illusion de compréhension.

“Moral conviction predicts sharing preference for politically congruent headlines” (Chapitre 3) explore, à travers 8 expériences, quel type d’information politique des participants américains décident de partager en ligne sur des sujets controversés tels que le contrôle des armes à feu ou l’avortement. Nous constatons que les participants préfèrent partager des articles dont le titre est politiquement congruent avec leurs attitudes politiques sur chaque sujet, et que

cette préférence augmente avec l'importance morale du sujet en question—que les articles soient vrais ou faux (« fake news »). Nous observons également que cette préférence est peu influencée par des manipulations expérimentales de la composition politique de l'audience, du caractère anonyme ou non du compte utilisé, et de l'exposition ou non à un message mettant en garde contre les biais politiques. La vérité perçue d'un article, et son utilité pour les buts politiques du participant, figuraient parmi les principales raisons de le partager. Cette tendance à la communication sélective pourrait contribuer à accélérer les divergences de croyances entre sous-cultures politiques, rendant plus difficile pour la gauche et la droite de s'entendre sur des faits communs.

Au travers de 5 expériences menées sur des participants français, “Intentions matter a lot, and efficiency little, in folk judgments of policy decisions” (Chapitre 4) montre que des citoyens ordinaires préfèrent des politiques altruistes mais peu efficaces et ayant un coût élevé à des politiques animées par des motivations égoïstes mais très efficaces et rentables. Ce résultat fut observé sur quatre thèmes distincts : protection de l'environnement, promotion de l'égalité de sexe, de l'autorité de la France dans le monde, et régulation de l'immigration. La préférence pour des politiques peu efficaces mais bien intentionnées semblait motivée par la combinaison d'une faible sensibilité aux différences d'efficacité exprimées sous forme numérique—une tendance amplifiée par le degré de conviction morale du participant—et d'une forte sensibilité aux intentions animant la personne mettant en œuvre la politique.

Ces travaux contribuent à dépeindre l'homme comme un animal paranoïaque et moralisateur, dont les instincts cognitifs ne le prédisposent guère au pragmatisme dans le domaine de l'action publique.



## Résumé substantiel

Le monde est tel qu'il est indépendamment de ce que nous en pensons. Dans la vie publique, la prise de décision rationnelle se doit donc d'être basée sur des évaluations précises des coûts et des bénéfices associés aux technologies et aux politiques publiques—que ce soit dans le domaine des biotechnologies, de la médecine, de l'énergie, ou de la politique économique—faute de quoi d'immenses opportunités pourraient être manquées. La « vérité » sur ces questions n'étant jamais directement observable, elle ne peut être obtenue que par un examen systématique des données disponibles, suivant les méthodes minutieuses et contre-intuitives de la science : recours à l'expérimentation, attention aux taux de base, réplication et méta-analyses, etc.. Une morale rationaliste exige donc tant des décideurs que des citoyens de suivre l'avis des experts sur les questions sur lesquelles la science est claire—par exemple, que l'homme contribue au changement climatique et que le nucléaire civil constitue un atout important dans la lutte contre ses effets, ou que les OGM et les vaccins actuellement commercialisés sont sans danger—et de se retenir d'avoir des opinions polarisées sur les questions sur lesquelles les preuves sont pour l'heure équivoques.

Le problème, c'est que l'esprit humain est le siège de biais cognitifs, intervenant tant au niveau de la compréhension, de la mémorisation que de la communication, qui rendent parfois difficile d'acquérir des représentations des risques et des bénéfices des politiques et des technologies qui soient alignées avec ce que la meilleure recherche suggère. Cette thèse, d'obédience naturaliste ou évolutionnaire, explore quelques-unes des prédispositions cognitives contribuant à ce résultat à travers un article théorique (Chapitre 2) et deux articles empiriques (Chapitres 3 et 4). Les travaux qui y sont présentés contribuent à dépeindre l'homme comme une espèce paranoïaque et moralisatrice, dont les instincts cognitifs évolués ne le prédisposent guère au pragmatisme dans le domaine de l'action publique.

“The cognitive foundations of misinformation on science” (Chapitre 2) est un article de revue s'appuyant sur les recherches en évolution culturelle, en communication scientifique et en psychologie évolutionniste pour examiner certains des principaux mécanismes contribuant à alimenter un décalage entre les croyances populaires et le consensus scientifique, comme les croyances erronées selon lesquelles les OGM ou les vaccins sont dangereux pour la santé. L'article prend pour point de départ l'idée que les gens ont généralement des difficultés à comprendre ce qui rend l'expertise scientifique beaucoup plus fiable que leurs propres intuitions et celles de leurs pairs. Nous acquérons l'essentiel de nos croyances politiques par la parole des autres—famille, amis, collègues, journalistes ou hommes politiques partageant notre vision du monde—auxquels nous sommes réceptifs lorsque des indices de bienveillance (« pourquoi essaieraient-ils de nous tromper ? ») ou d'affiliation (« ils sont comme nous ») sont présents, ou lorsque les opinions qu'ils défendent semblent répandues. Toutefois, nous réalisons rarement que ces personnes en qui nous avons confiance manquent de compétence technique sur les questions dont elles parlent, et que leurs déclarations ne respectent pas les exigences méthodologiques qui donnent en principe sa fiabilité à une croyance empirique—comme le fait qu'elle soit basée sur des études expérimentales répliquées, une inférence statistique contrôlée, etc.. Ce que la plupart des gens considèrent comme des « preuves » ne sont souvent que des anecdotes, à la représentativité douteuse. Victimes d'une « illusion de compréhension », ils ont aussi tendance à surestimer leur compréhension du fonctionnement causal de la plupart des phénomènes, qu'ils soient économiques, sociaux ou biologiques.

La plupart des affirmations auxquelles la communication sociale nous expose sont sémantiquement vagues et fragmentaires, ne contenant aucune spécification quantitative des phénomènes qu'elles décrivent (« les OGM sont dangereux » ; « les armes à feu tuent des gens » ; « les vaccins peuvent causer le cancer »). Mais même lorsque l'information est exacte à un certain point de la chaîne de communication (par exemple, dans un article de journal que nous lisons ou la présentation d'un expert que nous entendons), celle-ci est sélectivement écoutée,

mémorisée, reconstruite et retransmise aux autres d'une manière qui suit les pentes naturelles de nos intuitions cognitives. Bien que certaines de ces intuitions existent probablement en raison des services qu'elles rendirent à nos ancêtres, elles contribuent parfois à nous faire dévier d'une appréciation des risques et des bénéfices qui soient justifiée par les données. Par exemple, la communication humaine est régulée par des attentes de pertinence, qui conduit facilement à ce que les descriptions des résultats scientifiques que nous faisons circuler dans les conversations soient graduellement exagérées. Nous avons aussi des « théories naïves » de nombreux processus, qui peuvent biaiser la manière dont nous nous représentons certaines technologies—comme nos intuitions biologiques essentialistes, par exemple, qui semblent jouer un rôle dans la peur des OGM.

L'esprit humain est également prédisposé à sur-réagir aux informations relatives à la menace. Si cette sensibilité est adaptative d'un point de vue individuel, elle peut nous conduire à focaliser notre attention sur les seuls risques des politiques et des technologies, au détriment de leurs bénéfices, moins psychologiquement saillants. Cela peut être le cas, par exemple, d'informations suscitant la crainte de la contamination ou le dégoût. Notre système d'évitement des pathogènes semble jouer un rôle dans la peur de l'énergie nucléaire civile, les radiations étant intuitivement représentées comme un agent pathogène potentiellement contaminant. Dans un registre similaire, les théories du complot surfent sur des systèmes cognitifs ayant évolué pour détecter des actions coordonnées hostiles, qui furent vraisemblablement un défi adaptatif récurrent. Ces systèmes cognitifs deviennent problématiques lorsque, renforcés par la connaissance de conspirations et de mensonges publics historiquement avérés, ceux-ci conduisent la population à nourrir des soupçons disproportionnés vis-à-vis des activités des industriels et des scientifiques, sapant la confiance en eux.

Le fait que nous ne réalisons pas à quel point nos intuitions sont souvent inaptes à se représenter les conséquences réelles des politiques publiques et des technologies n'est guère surprenant. Nos esprits, après tout, ont évolué dans de petites communautés traditionnelles dans



lesquelles personne ne possédait d'expertise théorique avancée, et à une époque où nous n'avions qu'un contrôle technique limité sur la nature, de sorte que les politiques économiques ou énergétiques complexes, étant absentes, n'exercèrent aucune pression sélective.

Parfois les gens se fourvoient de manière à peu près universelle. Mais bien souvent, la direction dans laquelle ils s'écartent de la vérité peut être prédite par leurs convictions politiques. Les opinions des gens sur des questions sans rapport les unes avec les autres ont tendance à corrélérer entre elles et à se polariser, délimitant des sous-cultures politiques. Par exemple, la gauche a tendance à croire que les activités industrielles contribuent à réchauffer la planète, que l'énergie nucléaire n'est pas une source d'énergie sûre et efficace, que nos sociétés sont systématiquement racistes et sexistes, tandis que la droite a tendance à croire le contraire. Cette polarisation politique des croyances est problématique pour la démocratie, car elle réduit notre capacité à tomber d'accord sur la nature des problèmes auxquels la société est confrontée, et sur les solutions concrètes envisageables pour les régler.

Mais comment la polarisation politique des croyances émerge-t-elle ? La recherche suggère plusieurs mécanismes. Pour commencer, des différences de résidence géographique ou de classe sociale exposent les gens à des représentations différentes du monde et à des médias différents dès le début de leur vie. Plus tard, les gens tendent à s'entourer de personnes qui partagent leurs croyances, leurs traits de personnalité et leurs préférences politiques (homophilie), ce qui réduit les chances que leurs représentations soient mises à l'épreuve par de bons arguments en provenance d'autres sous-cultures politiques que les leurs (par exemple, venant des Républicains lorsque l'on est un Démocrate aux Etats-Unis). Et lorsqu'ils sont effectivement confrontés à des informations qui vont contre leur idéologie et leurs croyances, les individus sont prompts à leur appliquer davantage de scepticisme qu'à des informations qui confirment leurs attentes et leurs intérêts (biais de confirmation).

L'article "Moral conviction predicts sharing preference for politically congruent headlines" (Chapitre 3) explore une autre source potentielle de cette polarisation politique des

croyances : les gens préfèrent-ils *communiquer* sur les sujets politiques d'une manière qui privilégie les convictions qu'ils ont déjà, de la même manière que ceux-ci tendent à donner plus de poids aux informations confirmant leurs convictions et leurs croyances ? L'article entreprend de répondre à cette question à travers 8 études expérimentales examinant quels types d'articles politiques des participants Américains, libéraux et conservateurs, décident de partager sur internet sur 4 thèmes controversés : le contrôle des armes à feu, le droit à l'avortement, l'égalité de sexe et l'égalité raciale. Nous observons que les participants préfèrent partager des articles politiquement congruents avec leur position (plutôt qu'incongruents) sur chaque sujet, et que cette préférence augmente avec l'importance qu'a le sujet en question pour eux—que les articles soient vrais ou faux (« fake news »).

Nous avons également constaté que cette préférence était peu influencée par diverses manipulations expérimentales : de la composition politique de l'audience (peuplée des gens qui sont d'accord *vs.* en désaccord avec soi politiquement), du caractère anonyme ou non du compte à partir duquel le partage est effectué, ou encore de l'exposition ou non à un message mettant en garde contre les biais politiques dans le traitement de l'information. La vérité perçue d'un article, et son utilité pour les objectifs politiques du participant (e.g. discréditer le parti politique rival) figuraient parmi les principales raisons de le partager. Ceci suggère que la mésinformation et des « fake news » peuvent se propager largement à l'intérieur des réseaux s'ils surfent sur les idées préconçues et les intérêts politiques partagés par leurs membres.

Au travers de 5 expériences, “Intentions matter a lot, and efficiency little, in folk judgments of policy decisions” (Chapitre 4) étudie les facteurs influençant les jugements moraux de participants français sur des politiques publiques touchant à 4 domaines distincts : la protection de l'environnement, la promotion de l'égalité de sexe, de l'autorité de la France dans le monde, et la régulation de l'immigration. Nous trouvons que les informations relatives aux intentions des acteurs politiques, qui devraient être peu importantes d'un point de vue pragmatique, pèsent lourdement sur le jugement moral. En revanche, les informations relatives

à l'efficacité des politiques et à leur coût financier exprimés sous forme numérique, qui devraient influencer fortement les jugements, n'ont qu'un impact marginal. Nous avons également constaté que l'insensibilité des participants au degré d'efficacité des politiques était d'autant plus grande que ceux-ci avaient de fortes convictions morales vis-à-vis du domaine abordé par la politique (e.g. promouvoir l'égalité de sexe), ce qui corrobore l'idée que de fortes convictions morales nous rendent moralement plus déontologiques. S'ils se transposent dans le monde réel, la combinaison de ces effets pourrait amener les gens à soutenir des politiques inefficaces tant qu'elles sont mues par de nobles intentions, plutôt que des politiques nettement plus efficaces mais semblant guidées par l'égoïsme, et à communiquer sur les sujets politiques d'une manière qui privilégie les valeurs et les symboles mais qui néglige l'impact concret des politiques sur la société.

Plus généralement, les résultats du Chapitre 4 sont cohérents avec une vision du jugement moral comme étant câblé pour se focaliser sur les vertus morales des autres, plus que pour le pragmatisme et l'évaluation des conséquences. De nombreuses recherches suggèrent en effet que notre esprit moral a évolué principalement pour juger des partenaires de coopération potentiels. Nous serions comparativement peu équipés pour représenter et réagir aux informations sur l'efficacité des politiques, en particulier lorsqu'elles sont exprimées sous forme abstraite et numérique. Après tout, les politiques publiques et les organisations modernes, par définition, n'existaient pas pendant la majeure partie de notre évolution, et ne pouvaient donc pas constituer des pressions de sélection, et les systèmes numériques sont une innovation culturelle récente.

# Chapter 1: General Introduction

The world is the way it is independent of what people believe it to be. It is therefore a basic axiom of rational action theory that decisions should be made based on careful examination of their concrete consequences if they are to reach their goals. What *is* and what one *ought to do* are logically distinct spheres (Hume 1739/2000), but few normative theories would deny that anticipation of risks and benefits should play a central role in guiding our practical decisions (Von Neumann & Morgenstern, 1944).

Keeping this premise in mind may be particularly important in the political domain, where policy decisions do not just have consequences for the agent and her immediate surroundings, but potentially for the collectivity as a whole. Policies can result in considerable progress on societal and economic fronts, but also cause vast missed opportunities and wasted resources if they are based on wrong beliefs about the way the world works, and which technology can achieve what. Think for instance of yellow rice. Numerous specialists insist that growing this genetically modified rice enriched in vitamin A on a large scale could make a big difference in the fight against the effects of malnutrition in Africa—from blindness to death due to vitamin-A deprivation—at a very low cost (Wessler & Zilberman, 2014). Yet, many laypeople and non-governmental organizations the world over have misbeliefs about GMOs being dangerous for health and the environment despite wide scientific consensus to the contrary (Blancke et al., 2015; National Academy of Sciences, 2016; Fernbach, Scott, Rozin, 2019; McPhetres et al., 2019). This has led almost all countries to *ban* cultures of

yellow rice, contributing<sup>1</sup> to massive years of life and human happiness wasted (Wessler & Zilberman, 2014).

Likewise, people hear and pass along catastrophic stories about accidents at Fukushima (the only death of which was caused by evacuation, not radiation) and Chernobyl (which killed less than coal does every day). Many of them end up opposing civil use of nuclear energy, which may be one of our best chance of curbing global warming in the short run thanks to its vanishingly low carbon footprint and tremendous efficiency at producing electricity. Instead, many people wholeheartedly support solar or wind energies as panaceas without spending much time inquiring on how much mineral resources and square kilometers would have to be eaten up by an energy system entirely dependent on them (Ausubel, 2007; Brand, 2009; Pinker, 2019). Experts estimated that had Germany and Japan *not* shut down their nuclear power plants, 28 000 deaths caused by increase in air pollution, and 2400 megatons of released CO<sub>2</sub> could have been avoided between 2011 and 2017 (Kharecha & Sato, 2019).

In yet another domain, many people believe the world is increasingly violent and war-torn and that most countries are poorer than before, despite clear historical evidence that humanity is, in terms of global averages, more peaceful and vastly more affluent than in 1945 (Levy & Thompson, 2011; Pinker 2011, 2018; Roser, 2016a, 2016b).

The need to possess accurate factual knowledge of the concrete consequences of human decisions may primarily be incumbent on elected officials, CEOs and managers, since those sit at the top of powerful organizational chains of command that give them huge amounts of decision power on society. But democratic societies, through their polls, elections, and media, also put significant amounts of power in the hands of ordinary citizens, journalists,

---

<sup>1</sup> Even if they were authorized and their development encouraged, GMO foods may still remain difficult to access by the poor because of their price. But legally allowing them is surely a necessary condition for their contributing to fight hunger and malnutrition.

intellectuals, and pundits, whose beliefs, at the aggregate level, also have an impact on politics. From a moral perspective, acquiring and disseminating true beliefs on policy-relevant issues, whatever the position one occupies in society, might be considered as an ethical obligation in many circumstances.<sup>2</sup>

What is the landmark by which beliefs on policies' risks and benefits should be assessed? A surprising number of ordinary citizens, policy makers, and influencers of all stripes do not seem to be aware of this, but the response is unequivocally *science*, the collective and dialectic examination of the best empirical evidence and arguments available at any given time, the most reliable way humanity has ever found to arrive at accurate representations of how natural and social phenomena work, and of the consequences of our actions (Levy, 2017; Mercier & Heintz, 2013). “Factfulness”—the intellectual habit of relying on data and analytic thinking, of tolerating uncertainty, of approaching issues in terms of costs/benefits ratios, orders of magnitudes, probabilities and percentages, etc.—rather than intuitions, images, and anecdotes, is an intellectual virtue essential to evidence-based decision making, which should be part of the intellectual toolkit of any educated citizen and leader of the modern world (Pinker, 2019; Rosling et al., 2018).

An evidence-based ethics thus commands to follow the science on the issues on which it is clear—for instance that human activities contribute to make the planet warmer, that vaccines and commercialized food GMOs are safe, or that nuclear power contributes little to global warming. And on all the issues that are too complex for research to yield unequivocal

---

<sup>2</sup> Please note, however, that advocating the idea that consequences should play a central role in our moral and practical choices does not mean that *only* consequences should matter in moral judgment. I am just saying that assessments of consequences—concrete costs and benefits for human welfare—should attract *more* attention in how we form political attitudes and make decisions. But some deontological intuitions, according to which certain categories of acts are intrinsically forbidden, may be correct in certain cases. For instance, I agree that reaping the organs of an individual at the doctor's waiting without her consent to save five in need of a transplant—the right thing to do according to hardcore consequentialism—is immoral. Or that the end of protecting the nation against terrorists does not justify resorting to torture.

conclusions—e.g. on the impact of immigration on economic growth—the reasonable thing to do is to accept uncertainty, and refrain from holding polarized and over-confident opinions.

Of course, people have different fundamental moral intuitions about which goals societies should pursue in priority between which it is difficult to adjudicate. What is the optimal balance between the goal of reducing poverty and of promoting economic opportunity, or between personal freedom and collective safety? How much destruction of the natural environment should be incurred to protect our materialistic lifestyle? What balance must be found between the precautionary principle, and the potentially immense promises of technological innovation? There is evidence that political conservatives display stronger physiological responses to threat, and dedicate more resources to protect themselves against it (Hibbing et al., 2014). Likewise, people who grew up in more hostile and poorer environments tend to display more temporal discounting, more materialistic preferences, and lower interest in health and education than those who developed in safe and affluent environments. Part of those differences may be rooted in genetic differences between individuals (Hatemi et al., 2014), or in epigenetic differences corresponding to different life history strategies (Brumbach et al., 2009; Mell et al., 2018; Pepper & Nettle, 2017).

Individuals can disagree passionately about the priorities of existence, and what makes a just society, and many of those moral disagreements may be undecidable (there may be no “right” answer to them). But all humans are by definition part of the same species, and there is something akin to a universal human nature (Pinker, 2002). As a consequence, human beings, despite their ethnic, political and cultural differences, do share basic universal needs, and can in principle agree on the legitimacy of at least *some* policy goals, the advancement of which objectively benefits the majority: from reducing poverty and inequality to promoting economic affluence to protecting nature to fighting cancer, malnutrition, crime, or air pollution.

## **Our intuitions do not predispose us to follow the facts**

Policies' and technologies' risks and benefits are not directly observable states of affairs. They are facts that take place beyond the range of personal experience and which can only be assessed at an aggregate level of analysis, and in the mid- to long-run, through systematic examination of the data. The problem is that many cognitive distortions converge to make it difficult for most people to access reliable factual information about them, and to find the motivation to act upon this information when it is made available. This is what makes the above claim on the importance of basing our actions on an accurate understanding of the facts not trivial or redundant. As Pinker (2019) puts it, just as fostering a culture of factfulness may be desperately necessary to tackle humanity's problems, "an awareness of the infirmity of unaided human intuition should be part of the conventional wisdom of every educated person."

For instance, people tend to trust their own intuitions too much, as well as those of trusted sources from which they derive most of their beliefs (family, friends, politically congenial media) to the detriment of true experts (Fiske & Dupree, 2014; Mercier, 2017; Morin et al., 2020). They represent, remember and pass long information about risks and potential losses more readily than information about the benefits of technologies and policies (Kahneman & Tversky, 1988; Moussaïd, Brighton & Gaissmaier, 2015). They are miserable at representing risks as abstract probabilities, and differences between them (Gigerenzer, 2018). They heavily discount future well-being, whether of the self, strangers, and future generations (Ainslie, 2001; Singer, 2009). They are more powerfully moved by the fate of one individual child than by that of entire communities ("A single death is a tragedy, a million deaths are a statistic", as Stalin is purported to have said, and he knew what he was talking about; Bloom, 2017). Humans minds are wired for over-ascribing hostile intent under uncertainty, and for making systematically uncharitable interpretations of people with whom



they are competing (Boghossian & Lindsay, 2019; Stone, Patton & Heen, 2010; van Prooijen & van Vugt, 2018; Walmsley & O'Madagain, 2020).

The present dissertation aims to document a number of cognitive tendencies that, arguably, *may contribute to making it difficult for us humans—including yours truly*<sup>3</sup>, of course—to *access reliable information on important policy issues, and to have that information guide our attitudes and choices*. The dissertation is built around 1 review article and 2 empirical articles of which I am lead author:

- “The cognitive foundations of misinformation on science” (Chapter 2)
- “Moral conviction predicts sharing preference for politically congruent headlines” (Chapter 3)
- “Intentions matter a lot, and efficiency little, in folk judgments of policy decisions” (Chapter 4)

The dissertation’s interdisciplinary approach is inspired by the cognitive and evolutionary social sciences, or “naturalistic” social sciences, a research field at the crossroads between the social sciences, cognitive science, and evolutionary biology. This means that along with providing descriptions of our cognitive dispositions, and how they make us potentially deviate from evidence-based pragmatism, this approach also involves *proposing testable evolutionary hypotheses for why we might have the cognitive dispositions we do*. The following pages succinctly delineate the principles at the core of the evolutionary social sciences, before introducing each paper.

## **The evolutionary cognitive sciences (in a nutshell)**

---

<sup>3</sup> I recently discovered that this expression meant ‘I’ in old-fashion English, and I really like it.

Societies ultimately rest on networks of individuals exchanging information and constantly transforming it as it travels from brain to brain. Yet, people behave in many similar ways across the world, and some misbeliefs are extraordinarily common. To explain why many of our spontaneous ways of thinking and acting deviate from evidence-based pragmatism, social scientists must focus on psychology, on the natural slopes of human understanding. They must identify the biases that make us understand, interpret, memorize, transmit and profess certain representations rather than others to our peers, and in certain situations rather than others (Sperber & Hirschfeld, 2004).

Providing full-blown *explanations* of human behavior in turn requires doing more than just inventories of the underlying cognitive biases or tendencies, as is often the case in psychology and the social sciences (Sperber & Hirschfeld, 2004; Boyer, 2018). To explain *why* the human mind has the cognitive biases it does in the first place, human scientists have much to gain from reaching out to evolutionary biology: they should recall that human brains are biological organs that evolved by natural selection. Those brains are home to a wide diversity of specialized, domain-specific cognitive systems or “mental functions” that operate in parallel. Those systems output intuitive emotions, representations, inferences, and behaviors, the computational steps of which take place below the radar of conscious inspection. Evolution being a parsimonious engineer, there are good reasons to believe that many of those intuitive cognitive systems were carved by natural selection to trigger the reactions that, on average, would have promoted reproductive success, given the particular challenges that dominated the history of our species (and, in particular, our hunter-gatherer lifestyle; Barkow, Cosmides, Tooby, 1992; Tooby & Cosmides 1990; Boyer, 2018).

The reason why evolved cognitive systems make us behave in ways that can appear as “irrational” from the standpoint of evidence-based pragmatism is because those systems sometimes pursue goals that are orthogonal to factual accuracy and moral prudence. For instance, our minds’ predisposition to over-react to threat-relevant or disgusting information

may be regrettable to the extent that it seems to contribute to make some citizens prone to conspiratorial thinking, distrust in legitimate democratic institutions, or oppose vastly beneficial technologies such as GMOs or nuclear power to which the label “danger” has been ascribed (Blancke et al., 2015; Hacquin et al., 2020). But adoption of a biological perspective on this bias helps understand why having evolved it makes sense: because, ancestrally, there would have been much fewer costs in being paranoid about plots or pathogens (beyond a few unnecessary drops of sweat) than in *not* being on hair trigger alert about the infinitely many things that could potentially kill us (Haselton & Nettle, 2006; Haselton, Nettle, Andrews, 2015).

Likewise, consider the popular resistance to the theory of evolution, which drives scientists and rationalists particularly crazy when it is entrenched in regions of the advanced United States, despite evolution being taught in schools. As a matter of fact, part of this puzzlement may fade away if one considers that humans are coalitional or “tribal” creatures that were under strong selection pressures to affiliate to groups in contexts of group conflict (Tooby & Cosmides, 2010). A predicament which likely selected for cognitive systems the function of which is to hold and profess beliefs—in particular controversial ones—as signals of cultural and political membership (e.g. that one is part of the “true Christian patriots”, not “libtards” from the “snobbish coastal elites”); Kahan et al. 2011a; Levy, 2017; Mercier, 2017, 2020; Pinker, 2018).

Importantly, the claim that many human biases are the outputs of domain-specific systems that evolved does *not* mean those systems are rigid, insensitive to cultural variation, or history. While they were shaped by selection to meet specific challenges and as such can be attributed biological functions, most of the systems that populate our minds also evolved to be flexible and context-sensitive to adapt to varying ecologies. They are neurocognitive programs that provide us with no more than a preparedness to learn, represent, memorize, feel, and act, on the basis of input information they receive from perception and

communication—and this input information can vary widely across historical times and places (Al-Shawaf et al., 2018; Boyer, 2018; Sperber & Hirschfeld, 2004). For instance, while part of the public distrust in technologies and institutions we are witnessing today may be plausibly attributed to a promiscuous tendency to detect coordinated hostile action directed towards oneself, even where there is none (van Prooijen & van Vugt, 2018), it goes without saying that the conspiracy theories of ancient Greece are in many respects different from those of 1930s Germany or today’s Trump supporters. Taking into account people’s prior beliefs, which vary across time and space, is no less important than pinning down universal cognitive tendencies if one is to propose full-blown explanations of human behavior (Boyer & Petersen, 2018; Boyer, 2018; Bronner, 2011; Cordonier, 2018; Sperber & Hirschfeld, 2004).

How do we know that human behaviors rest on a myriad of relatively simple (but not rigid) computational systems? This of course is a complex debate, which I will not be able to cover here. One clue is the link between certain brain regions and certain psychological functions, revealed by neuropsychological disorders (Damasio, 1994). Another comes from comparative biology: biologists consistently find that the behaviors of plants, animals, and bacteria are caused by simple structures and relatively automatic instincts, which suggests that evolution constructs biological design with computational parsimony (Callebaut & Rasskin-Gutman, 2005; Clune, Mouret & Lipson, 2013; Schlosser & Wagner, 2004). So, given the unity of life (we all have a common ancestor), there is no apparent reason why our own complex cultural behaviors should not also be realized by simple specialized systems. Yet another argument is that trying to have artificial intelligence reproduce our behaviors and cognitive operations inevitably requires decomposing them into simpler operations (Marr, 1982).

### **“The cognitive foundations of misinformation on science”**

Let me now introduce the papers contained in this dissertation. The paper presented in Chapter 2, “The cognitive foundations of misinformation on science” adopts the approach favored by the evolutionary cognitive sciences of striving to explain puzzling recurrent social phenomena in terms of the evolved psychological systems that underpin them. It takes the issue with which we began this introduction as its explanatory target: the fact the public often holds beliefs on the risks and benefits associated with key technological and policy issues that are in stark contrast with the scientific consensus—such as the belief that currently commercialized genetically modified foods’ and vaccines are not safe, that nuclear power releases significant CO<sub>2</sub>, or that research on the genetic determinants of behavior advance a dangerous, “deterministic” view of human existence.

Along with co-authors Sacha Altay and Brent Strickland, I draw on the literature in cultural evolution, science communication, and evolutionary psychology to review what we deemed to be the most important psychological mechanisms responsible for this outcome. Among other processes, we highlight the role played by expectations of relevance and vagueness in communication in generating exaggerated and distorted accounts of scientific findings (Sperber & Wilson, 1987). We pin down our propensity to overestimate our own causal understanding of most physical and social processes (Fernbach et al., 2013, 2019), and put forward the perverting influence of evolved, “naïve theories” of those processes, such as the role of biological essentialist intuitions, in entertaining fear of GMOs, for instance (Blancke et al., 2015; National Academies of Sciences, 2016; Fernbach et al., 2019; McPhetres et al., 2019). We also underscore our inclination to discount information that runs against our priors and political preferences (Mercier & Sperber 2011; Nickerson, 1998; Stanovich & West, 2007), and to surround ourselves with people, and privilege news sources, that on the whole tend to already agree with us politically (Alford et al. 2011; McPherson, Smith-Lovin & Cook, 2001).

The paper came out in *EMBO Reports*, a molecular biology journal whose readership is mostly made of natural scientists, hence our efforts to avoid psychological jargon. We meant our review to mention the cognitive biases responsible for what seemed to us to be socially impactful popular misbeliefs, but we obviously make no pretense to exhaustivity.

The article ends with practical recommendations to scientists on how they might want to frame their findings in order to mitigate chances that they be misunderstood and twisted. Our recommendations are generally structured around the idea that scientists should make more efforts to anticipate predictable misconceptions. Think for instance of the recurrent but illogical inference that findings in genetics “imply” certain things in how we should treat people ethically—contradicted by Hume’s *is/ought* gap maxim according to which facts have no moral implications (Hume 1739/2000), and the fact that compensating inequalities requires understanding their causes. Or the notion that existence of cognitive biases in individual scientists can suffice to make entire areas of research unreliable—largely debunked by the fact that scientific research is a collective enterprise in which peers have professional incentives to pin down their peers’ blind spots (Levy, 2017; Mercier & Heintz, 2013).

### **Moral convictions and trade-off insensitivity**

The “Cognitive foundations” paper presented in Chapter 2 reflects my interest in explaining why our evolved cognitive biases can grant spectacular “cultural success” to certain misbeliefs, thereby potentially making us lose track of scientifically-assessed risks and benefits. But my core intellectual interests are in studying how moral and political processes specifically distract us from important facts and pragmatic thinking, the adoption of which would make us wiser.

Among the moral and political biases that make us irrational, I am captivated by the consequences of *strong moral convictions* on reasoning and behavior (Skitka et al. 2005; Skitka & Mullen, 2002; Skitka, Bauman & Sargis, 2005). What social psychologists mean by moral convictions are commitments to certain goals or values—such as protecting the environment, one’s religion, or fighting for sex or racial equality—that individuals claim to pursue as “sacred” or “absolute” moral imperatives (Atran, 2010, 2016; Baron & Spranca, 1997; Graham & Haidt, 2012; Tetlock, 2003; Tetlock et al., 2000). At an operational level, those convictions can be measured based on agreement to statements such as “*Protecting [e.g. nature] is an absolute moral imperative*”; “*The conviction that one must fight to protect [the environment] is central to my identity*”; or “*Protecting [the environment] should be the government’s priority*” (Skitka et al., 2005; Ståhl, Zaal, Skitka, 2016)<sup>4</sup>.

On the subjective side, a strong moral conviction is experienced as a rule or principle an individual feels she must honor consistently, regardless of circumstances (Singh & Hoffman, 2020). Contradicting the axiom that rational action should be based on cost/benefits considerations, people who reach high levels of moral conviction will claim to refuse deviating from their cherished moral goal, or trading it off against other goals, even when doing this implies incurring important losses and renouncing significant rewards. Strong moral convictions also often contain an element of moral objectivism: the value is regarded as being a “fact”, imposed onto the mind from the outside (Goodwin & Darley, 2010; Stanford, 2018), which must be respected for intrinsic reasons, i.e. reasons irreducible to instrumental considerations about benefits<sup>5</sup>. Strong moral commitments are also typically experienced as being tied to moral identity: they are regarded as definitional to the kind of people individuals

---

<sup>4</sup> Other common measures of moral convictions ask how much a participant would have to be paid to accept to violate her core moral conviction, for instance “*Signing a secret-but-binding pledge to only hire people of your race in your company*” with respect to racial equality, and using responses such as “*Never for any amount of money*” as indicating a high level of conviction or sacralization (Graham & Haidt 2012).

<sup>5</sup> Note that this does not prevent people from producing rationalizations to justify their core intuitions that do appeal to consequences.

want to be, to “what it means” to be the member of a certain community (“*We, the true Feminists/Environmentalists/Patriots*”).

Individuals rarely, if ever, have moral commitments alone. Our moral convictions presuppose a community of ingroup members sharing the same convictions, and they always prescribe some form of cooperative behavior: protecting, caring, remaining loyal, enforcing equality, etc.. But the community emerges all the more easily as it posits outgroups (“*Them, Patriarchy/the Infidels/the Capitalists*”), perceived at best as rivals, at worse as an existential threat. It is in those conditions that the insensitivity to trade-offs implied by moral convictions can lead individuals to feel “mandated” to sacrifice other moral norms, from free speech to consideration of scientific evidence to democratic due process to respect for outgroup members as legitimate human beings (Atran, 2010, 2016; Skitka, Bauman & Sargis 2005; Skitka & Mullen, 2002). Think for instance of Trump supporters’ attacks on Mexican immigrants, Democrats and political correctness in the name of their political battles (fighting illegal immigration, the establishment, free speech); or, in the more extreme cases, calls to behead teachers who showed caricatures of Muhammad in the name of Islam.<sup>6</sup>

My interest in moral convictions and their ability to push people to refuse trade-offs fed two parallel research streams during my PhD. On the one hand, I put forward possible evolutionary functions for trade-offs insensitivity in commentary articles, one of which is presented in the Appendix: “Moral rigidity as a proximate facilitator of group cohesion and

---

<sup>6</sup> Now, it is important to highlight that these are features of the way individuals *experience* their strongest moral convictions. In practice, at a behavioral level, honoring them consistently is impossible (Tetlock 2003; Baron & Leshner, 2000). The world is a place of rival resources in which all things must ultimately take on a finite value, and be mutually fungible. As pointed out above, there are objective trade-offs between protecting nature, and human welfare; between fighting poverty and reducing inequalities. In certain circumstances, individuals may acknowledge this basic fact, that their “sacred” moral commitments aren’t that unconditional after all; that they, too, can ultimately be traded off against other goals. This may seem contradictory, but it shouldn’t come as a surprise to the cognitive social scientist. Our moral minds, just like the rest of cognition, are made of multiple systems operating in somewhat parallel ways (Cushman, Young & Greene, 2010; Kurzban, 2010; Tooby & Cosmides 1990). So it may be that some systems in the mind evolved to perform pragmatic, cost/benefit calculations, while others are here to give us deontological intuitions that a given value “is an unnegotiable imperative”.



combativeness”, published in *Behavioral and Brain Sciences* as a commentary of De Dreu & Gross (2019). The ideas I am defending there will be expanded in a more comprehensive theoretical paper currently in preparation. In a nutshell, I propose that cognitive systems making us relate to moral goals as unconditional imperatives or rules may have been selected because they are a parsimonious way of simultaneously attracting the trust of one’s cooperation partners, and of protecting oneself against exploitation attempts from within and outside one’s group, following an error management logic (Haselton et al., 2015; see Appendix).

But the main angle from which I approached moral convictions during my PhD was that of their consequences on judgment and behavior, a question central to the two empirical papers presented in this dissertation: “Moral conviction predicts sharing preference for politically congruent headlines” (Chapter 3), and “Intentions matter a lot, and efficiency little, in folk judgments of policy decisions” (Chapter 4).

### **“Moral conviction predicts sharing preference for politically congruent headlines”**

“Moral conviction predicts sharing preference for politically congruent headlines” (Chapter 3), examined the influence of people’s degree of moral conviction on the type of information they decide to communicate to others on political topics. It also drew inspiration from research on one of the cognitive biases most often accused of making people drift away from the truth on political matters, the myside bias (or confirmation bias; Mercier & Sperber 2011; Nickerson, 1998; Stanovich & West, 2007). In the political domain, the myside bias is the tendency to apply higher standards of skepticism to information that runs against one’s

political attitudes, identity, or interests than to information that is congenial to one's political views—at the cost of potentially missing occasions to correct one's beliefs on issues where one is misinformed. For instance, a scientific article highlighting the role of human activities in global warming (a scientific fact) is more likely to have the credibility of its methods or the probity of its authors questioned by a conservative industrialist than by a liberal professor (Kahan et al., 2011a). Likewise, an article claiming that nuclear energy does not emit much CO<sub>2</sub>, and that its efficiency in producing electricity can be an essential tool in efforts to reduce carbon emissions (two scientific facts) is more likely to have its methods questioned by left-wing environmental activists than by right-wing voters (Nuclear Energy Agency, 2010). To complicate things further, the more people are committed to an issue, the more they apply double standards of skepticism (Kahan, 2016; Van Bavel & Pereira, 2018).

The extent to which the myside bias should be considered as “irrational” or not is a matter of current debate in psychology today, and the response may largely depend on the mechanisms at play (which is hard to assess experimentally). According to a popular view (e.g. Kahan 2013, 2016; Kunda, 1990; Van Bavel & Pereira, 2018), the reason why individuals tend to give more or less credence to new information depending on its consistency with their political values is because they do “motivated reasoning”. In this psychological scenario, a scientific article highlighting the role of human activities in global warming would be perceived as threatening to an American industrialist's political identity and preferences, which would trigger an unpleasant feeling. This feeling would in turn motivate the reasoner to (unconsciously) search selectively for evidence and arguments in memory that can only justify skepticism towards the study, for instance by questioning the competence of the researchers who made it, or the rigor of the method they employed.

According to an alternative, and more parsimonious view, Bayesian updating, the link between people's political attitudes and the ease with which they accept or reject new information can be accounted for by the weight of their prior beliefs only, and need not

involve any search for evidence biased by emotion (Baratgin & Politzer, 2006, 2010; Pennycook, 2020; Tappin, Pennycook, Rand, 2020). People's political attitudes, the Bayesians argue, do not come from nowhere: they emerge from and covary with people's factual beliefs about what is true in the world, and what should be done or avoided to make society fairer. Individuals acquire their prior beliefs through repeated exposure to personally trusted sources. The American industrialist, for instance, *sincerely believes* that anthropogenic climate change is a hoax, because that's what Fox News, along with his friends and colleagues in his hometown, have told him for two decades. And since the industrialist trusts his beliefs, just like everyone else, he simply uses them as benchmarks, as default hypotheses to judge that this article highlighting the role of human activities in the warming of the earth can safely be discarded as statistical "error". As Bago, Rand & Pennycook (2020:2) highlight, doing this may sometimes mislead, but it is fairly rational from a subjective standpoint:

When considering evidence that is inconsistent with your prior factual beliefs, it can be rational to conclude that it is more likely that (i) the information source is unreliable than that (ii) the accumulation of all your prior knowledge is wrong. For example, if a stranger tells you that he was abducted by aliens, is it not irrational to conclude that that information is probably unreliable. Thus, it is essential to account for prior factual beliefs when attempting to test for politically motivated reasoning.

Regardless of what is the exact mechanism underlying the correlation between political attitudes and how much scrutiny people put when processing politically-relevant information, no one disputes the existence of the correlation. The project "Moral conviction predicts sharing preference for politically congruent headlines" was inspired from an analogy: Do people display a preference for politically congenial information when deciding to *communicate* it to others on controversial topics such as abortion or gun control, just like they

have a congeniality bias when processing political information? Moreover, my interests in what happens when people become strongly committed to political topics made me wonder: Does a person's degree of moral conviction on a political issue *reinforce* their propensity to do selective communication on that issue, just like stronger moral convictions make us more partial when we evaluate novel information?

Spoiler alert, we found that participants do indeed prefer to communicate political information that is congruent to their prior beliefs and political attitudes over information that runs against them, and that the more morally committed to an issue they are, the more they display this tendency. Why should this matter? Since people tend to spend most of their time surrounded by like-minded others (Alford et al., 2011; McPherson, Smith-Lovin & Cook, 2001), we suspect that a tendency to communicate on politics in ways that favor the beliefs or political interests one already has, and to remain relatively silent about counter-attitudinal or unexpected information, will not help individuals acquire richer and more nuanced representations of complex, hot button policy issues (such as gun control or abortion).

### **“Intentions matter a lot, and efficiency little, in folk judgments of policy decisions”**

Knowing that people pass along information to their networks at a higher rate when it is congenial to their values and priors than when it isn't allows us to make fairly robust predictions about how individuals, given their attitudes, relate to new information. But it does not tell us which type of politically relevant information is more attractive to the human mind *in general*, regardless of a person's attitudes—what may be called a content bias (Mesoudi & Whiten, 2008; Aarøe & Petersen, 2018). Classic instances of content biases relevant to the

processing of political information are the biases for threatening or disgusting information (Moussaïd, Brighton & Gaissmaier, 2015). There is a clear adaptive advantage in giving priority importance to information about potential fitness costs—plots, sources of pathogens—so our minds evolved predispositions for over-attending and over-reacting to threat-relevant information.

Similarly, people like and remember more easily arguments in favor or against certain policy issues when it is presented through anecdotes based on the lives of particular individuals than when it is provided in abstract statistical format (Aarøe & Petersen, 2018). In the same vein, considerations about the trials and tribulations of particular politicians—e.g. who they married or have had affairs with—often with moralistic tones, garner much more attention than the details of their policy positions and past decisions, which people easily find boring and are quicker to forget. A whole industry of “yellow” and celebrity journalism is based on leveraging this bias to its financial benefit (*Paris Match*, *The Sun*, etc.). The attractiveness of stories about particular individuals likely comes from the fact that humans relied so heavily on reputational information about others’ vices and virtues and who-did-what-to-whom in the community that it would have been adaptive to develop an intrinsic taste for telling and hearing them (Baumard, André & Sperber, 2013; Sperber & Baumard, 2012).

The project that led to the paper “Intentions matter a lot, and efficiency little, in folk judgments of policy decisions” (Chapter 4) was led by an interest in what seemed to be one particularly prevalent content bias in how people spontaneously think and talk about politics: the importance they give to political actor’s *intentions* when judging their political accomplishments. In casual conversations on politics, laypeople’s attention is often powerfully captured by speculations on what are the “true” motives behind opaque business and political choices. Recently, for instance, many of us surely took part in discussions about what are the exact goals pursued by our governments in implementing confinement policies, or what are the primary motives behind the pharmaceutical industries trying so hard to

develop vaccines against Covid-19 (is it because “they really care about humanity”? Or because they “simply want to make profits”? Or something in between?).

The problem is that, from a pragmatic point of view, what intentions are driving policy makers and businesses should be of little import, if any. As alluded above, in large-scale modern societies, well-designed policies can massively increase the welfare of the community as a whole. So whether the persons in charge of implementing them are virtuous enough to be the kind of people you would like to have a beer or play karaoké with should only be given secondary importance. What should dominate the conversation of pragmatic individuals is whether and to what extent the policies are *efficient* in reaching their stated goals, at an acceptable cost—for instance whether the pharmaceutical industries are able to produce vaccines that are safe, abundant, and cheap enough, to allow for a quick reboot of our social lives and the economy.

The experiments presented in the paper “Intentions matter a lot, and efficiency little, in folk judgments of policy decisions” (Chapter 4) were inspired from observations of this sort. On several policy issues, it systematically examined whether laypeople spontaneously prefer altruistically motivated policies that achieve little at a high cost, over selfishly motivated but highly efficient and profitable policies. In line with my interests in the behavioral consequences of having strong moral convictions, we also looked at the influence of individual differences in moral conviction on each issue, with the expectation that greater commitment would decrease the weight of cost/benefit considerations in participants’ attitudes towards the policies.

## Chapter 2: The cognitive foundations of misinformation on science

This section is based on the article:

“The cognitive foundations of misinformation on science”

Published in EMBO Reports, 21(4). <https://doi.org/10.15252/embr.202050205>

2020

Antoine Marie, Sacha Altay & Brent Strickland

References cited in this paper are after the text, rather than in the general references section at the end of the dissertation, because the editors requested the references to be in *Nature* format, not APA format.

Please note that the editors required a maximum of 10 scientific references. We would have put more of them if we could.

### **Introduction**

In spite of fast and continuous progress in science, general education, and easy access to high-quality information, many non-scientists are surprisingly uninformed, or misinformed, about the state of expert consensus on scientific debates and technologies. A sizeable number of laypeople in the USA are still skeptical of global warming or the theory of evolution, and a significant proportion of people in the West believe that genetically modified foods are

dangerous, that vaccines are not safe, or that nuclear power plants release significant amounts of CO<sub>2</sub> into the atmosphere.

Not only does misinformation about scientific facts contribute to feeding unnecessary anxieties, it also has hazardous political and economic consequences. Public policies are partly oriented by moral choices about which values society should pursue or uphold, but their implementation critically relies on factual beliefs about how the world works and which technology can achieve what. The accuracy of these beliefs conditions the policies' success. Whether they are in the position of voters, journalists, or policy makers, it is therefore crucial that citizens be informed about the state of the art if they are to make decisions that effectively achieve stated goals.

This commentary draws on research in the cognitive social sciences to outline a simplified model of how scientific findings become distorted, impoverished, exaggerated or ignored when being transmitted among non-scientific audiences. It focuses on the psychological mechanisms that make certain misbeliefs more likely to emerge, to be retained, and widely transmitted, thereby constituting obstacles to accurate science communication. The last part of the paper puts forward a few practical recommendations on how scientists could present and discuss findings on controversial topics if they want to mitigate risks that they be misunderstood and twisted.

### **How communication works**

Society is ultimately made of networks of individuals interacting and exchanging information. In the case of scientific information, communication can be described as a two-step process: information first emerges in mass media sources based on reading of the scientific literature



and press releases by journalists, and it is then reinterpreted by the lay public through informal discussions [1].

Crucially, and in contrast with genetic or electronic information, communicated information gets significantly modified when travelling from one individual to another. The beliefs we express are transcribed in language in terms that are often vague or underspecified, and language comprehension is often piecemeal and incomplete. When someone says ‘there have been concerns about GMOs being toxic’, do they mean that a *majority* of GMOs may be dangerous, or only just one? What negative outcomes are they supposed to cause, and in which dosage? Assuming speakers even know the answers to those questions, their addresses often do not, which leaves room for reinterpretation and misinterpretation. Information is thus inevitably reconstructed at the level of understanding, and often selectively retransmitted at the level of emission. To understand how everyday communication works, and how it contributes to science misinformation, it is therefore indispensable to explore the mechanics of the human mind.

### **Human minds are made of intuitive mechanisms**

The mind relies on specialized and intuitive inference systems, shaped by natural selection to quickly react to the practical challenges and opportunities faced by our hunter-gatherer ancestors in their everyday lives. The function of these intuitive systems is to attend to a certain type of information in the environment in order to quickly and often automatically generate representations, attitudes and behaviors that, ancestrally, would on the whole have promoted survival and reproduction. For instance, to the extent a certain pattern of movement in the grassland could have indicated a lurking predator or enemy, it would have been adaptive to

develop automatic systems dedicated to detecting such motion patterns and triggering fight-or-flight responses.

While these intuitive processes owe their existence to the favors they did to our ancestors, our environments have significantly changed since the Pleistocene, so these processes now sometimes lead us astray by producing attitudes and behaviors that are ill-adaptive or irrational from the point of view of modern standards. Think for instance of our addiction to sugar and fat, whose original function was to help us store as much calories as possible for bad times, but which is now one of the primary causes of cardio-vascular diseases all over the world (Fig. 1).

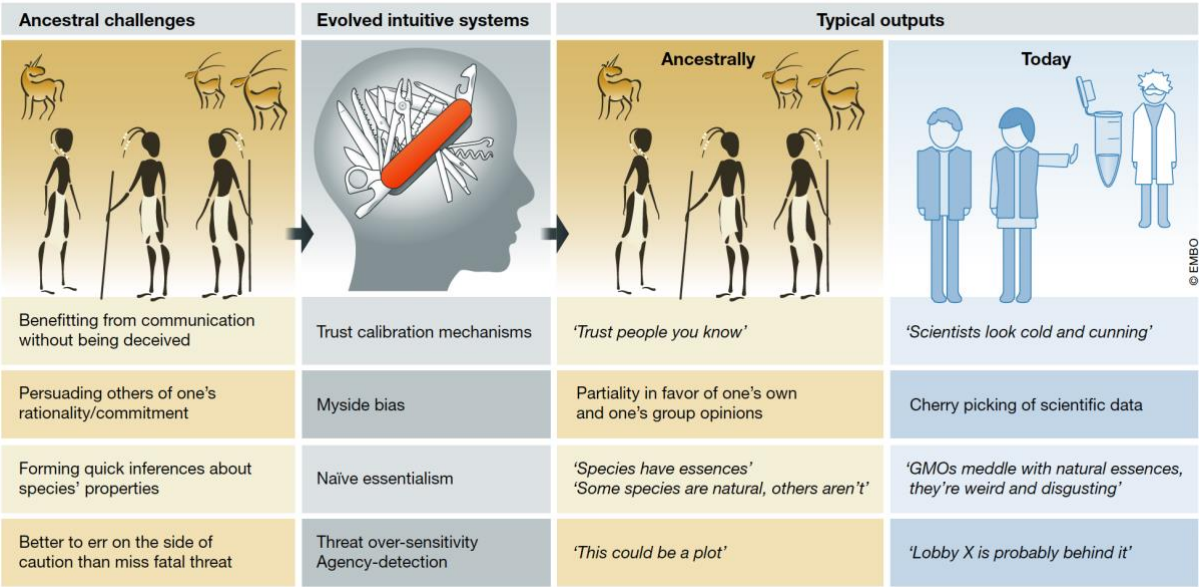


Figure 1. From ancestral challenges to evolved intuitive systems to misbeliefs about scientific facts.

**We trust familiar sources we see as benevolent and competent**

Intuitive processes can shape how we derive our personal positions on complex topics such as nuclear power, GMOs or economic policy. We do not typically form judgments on such issues through arduous study, but rather through the word of others—whether mass media or informal discussions with friends, family or colleagues. Given that it is impossible to personally assess the accuracy of all claims we encounter, and that there are immense benefits in trusting others—thereby taking advantage of the “division of cognitive labor”—our minds evolved intuitive mechanisms of trust calibration that dispose us to trust sources we see as competent and benevolent, and opinions we see as being widespread.

Our trust-calibration mechanisms perform pretty well in most everyday situations. However, they evolved to work in tight-knit communities in which personally knowing others was a good proxy for assessing their trustworthiness, and in which no one possessed advanced scientific expertise. As a consequence, without formal teaching about the experimental method, peer-reviewing, meta-analyses, or other features distinguishing personal experience from a scientific study, our minds are ill-prepared for understanding what makes scientific beliefs better grounded in reality than other beliefs. People end up giving too much authority to family and friends or people who hold ideologically similar beliefs (other Democrats, Republicans, Muslims, environmental activists, etc.), and they have difficulties distinguishing self-appointed pundits from scholars recognized as true experts in their fields.

### **We tend to be partial in favor of the beliefs we and our groups hold**

In addition to their inclination to not trust experts enough, people are also endowed with a robust and unconscious tendency to expose themselves to, interpret, and selectively retain information in ways that are partial to the beliefs and hypotheses they already favor. This “myside bias”, or

“confirmation bias”, is especially strong with regards to topics in which they have invested emotionally, such as nuclear power, GMOs or economic policy, often because those are part of the moral matrices and identities of the groups they belong to. It is on those polarizing topics that information running against favored narratives is most likely to be discounted and forgotten, or to be selectively retained and exaggerated when it fits preconceived notions. Some studies even suggest that this form of “motivated thinking” can bring partisan individuals to forgo small monetary gains to avoid exposure to cross-cutting political information [2]. In interaction with our tribal dispositions, our inclination to see and defend only one side of otherwise complex issues makes it difficult to recognize that alternative standpoints often have a kernel of truth.

Moreover, the informational isolation effects of the myside bias are all the stronger as people preferentially interact with like-minded others in ideologically homogeneous networks, thereby depriving themselves of well-argued, dialectic attacks on their own positions. It is worth noting however that while social media have been accused of constituting such ideological “echo chambers”, careful examination of the information exchanged on those virtual networks suggests that they are actually more ideologically diverse than the information people are exposed to in real life, face-to-face interactions.

### **We have naïve theories about many things**

Of critical importance for understanding scientific misinformation are also naïve theories, such as our naïve understanding of physics (e.g. we expect solid objects to be impenetrable and to fall to the ground), of human psychology (we tend to see others as driven by unambiguous beliefs, desires, and intentions), or our “intuitive essentialism”. As regards the latter, our

ancestors were surrounded by plant and animal species about which they had to draw quick inferences: Can it move? Is it edible? Is it dangerous? Evidence suggests that those selection pressures endowed us with a set of intuitive, innate expectations, emerging early in child development, about what constitutes a biological species: that all living things belong to mutually-exclusive taxonomic categories, which are fixed in time, and whose members, e.g. roses, contain a set of intrinsic traits that distinguish them from tigers, bathtubs, and sunflowers.

Considered from the perspective of modern biology, our intuitive essentialism is partially false: there is genetic diversity within each species, which are all subject to changes caused by mutation, drift, gene flow, and selection. But essentialist intuitions provide us with a conceptual template about what biological species are that, at the conscious level, makes it counter-intuitive and even dubious to a lot of people that species could have evolved by natural selection, and *a fortiori* that it should be possible to engender new ones by simply modifying their genome in the lab [3]. This likely triggers distrust and even disgust in a lot of people, who tend to conceive of GMOs as degenerate hybrids, which in turn causes cognitively attractive beliefs, such as ‘GMOs are weird because they involve meddling with natural species’ or ‘GMOs are dangerous for health’, to enjoy spectacular cultural success.

### **Running on empty: the illusion of understanding**

An important consequence of our minds being full of naïve expectations about how the world works, tuned to the practical demands of prehistorical life, is that they are very much unprepared for forming scientifically accurate representations of *most* facts on the basis of the input they receive. This applies to processes that, either because they are too big, too small, or because they could not possibly have occurred in our ancestral past—such as genetic engineering or

nuclear fission—take place beyond the range of evolutionary meaningful experience. Our default intellectual predicament is therefore one of ignorance on most matters covered by the sciences, coupled with a general ignorance of our own ignorance [4], at least until special circumstances oblige us to engage with the details of a specific area of knowledge—for example, when attending a course at a university, or when being personally affected by a disease that will force one to acquire basic notions of biology.

### **Informational market workers know many biases and exploit them**

Needless to say, many journalists, university press staffs, bloggers, and politicians exploit these cognitive biases to further their own careers, reputations, and incomes. This is entirely rational: workers in the information market are economically incentivized to leverage cognitive bias in their audiences. Journalists, for instance, get rewarded by their peers and employers for their ability to capture audiences' attention and generate traffic on the websites. By summarizing a scientific article's claims so as to make it easier to understand and attractive for a specific readership, they are in turn more likely to attract clicks and eyeballs, which translate into revenue through sales, advertising, donations, and so on.

### **Relevance = Revenue: The lure of exaggerated claims**

In order to sound more interesting and competent, and to distinguish ourselves from others, we naturally exaggerate things—at least when we think the odds that our audiences will find out are low. This tendency to artificially inflate relevance is so ubiquitous that we often do it unconsciously. Yet, we might believe that contrary to ordinary individuals, institutions like

universities' communication departments refrain from it, especially since, as the main link between researchers and the media, they are responsible for correctly informing the public.

Unfortunately, a recent study by Sumner et al. [5] found that press releases from universities are often an early locus of exaggeration of research findings' significance in the information chain. The authors showed that 40% of press releases that gave health-related recommendations to the public "contained more direct or explicit advice than did the journal article"; that "33% of primary claims in press releases were more strongly deterministic than those present in the associated journal article"; and that "for studies on animals, cells, or simulations, 36% of press releases exhibited inflated inference to humans compared with the journal article". In the category of inflated causal claims, confusion between relative and absolute risk is also a frequent culprit. Reports of health research may state, for instance, that risks of cancer have tripled (relative risk), which can sound quite frightening without any contextual information about their prevalence in the population (absolute risk), which may be quite low—learning that risks of cancer have increased from merely .001% to .003% is indeed less impressive news.

Why such exaggerations? Humans' evolved psychology of communication is constructed on mutual expectations of *relevance*, which are like an implicit contract between communicators that the information they exchange should be worth it. When we speak, specialized systems work to package what we are saying to others in a way that maximizes its informativity or impact, while keeping the mental cost required for understanding it low.

### **Turning scientific research into tales**

A common technique for repackaging information so as to enhance its relevance is through storytelling. Narratives involving protagonists with sharply contrasting interests and worldviews, competing with one another by resorting to tactics are found in all cultures, and do not spare conversations and publications about scientific discovery. Stories' cultural success comes from them simultaneously exploiting many folk psychological biases, such as our inclination to ascribe simple purposes and traits to people, our disposition to imagine all sorts of social plots, and our taste for gossiping about the trials and tribulations of others. By exploiting those inclinations to attract readers, scientific journalists like to frame the logic of scientific investigation in terms that attribute more influence and determination to individual figures than what happens in reality, where teamwork, uncertainty, hesitation, and repeated attempts are more prevalent.

### **Information fitting preconceived notions can spread like wildfire**

While the claim that humans are strongly gullible is likely overstated, our abilities to spot misinformation are weakest when this information is simultaneously cognitively attractive and comes from sources we have no reasons to distrust. One paradigmatic form of relevant information is that eliciting fear or disgust. Through its ability to trigger negative emotions, threat-related information tends to attract more attention, to be more easily remembered, and to make speakers appear more competent than emotionally neutral or positive information. Lab experiments simulating real-world communication on transmission chains suggest that, as information gets passed on from one individual to the next, messages become shorter and their content increasingly unfaithful to the original, except risk-related information that gets gradually exaggerated [6].



Incidentally, our high sensitivity to threatening information also facilitates the fallacious attribution of moral “implications” to merely empirical facts. For instance, people are often quick to infer that discoveries about the heritability of intelligence will inevitably lead to the “justification” of inequalities, despite the logical fact that, since descriptions imply no prescriptions, such discoveries can just as well provide ground for educational policies targeting underprivileged individuals.

Note that the general susceptibility to threatening information makes a lot of sense from an adaptationist standpoint, as it is better for organisms to err on the side of caution, even at the cost of slight paranoia, than to risk missing potentially lethal threats.

But non-threatening information can be quite successful as well when it emanates from trusted sources, and taps into naïve theories or ingrained stereotypes (which, regarding the latter, depend on people’s priors, social and professional position, and so on). On social media such as Facebook or Twitter, sharing is often based on skim-reading of articles’ titles or chapeau, and skim-viewing of YouTube videos. Content can therefore travel fast without undergoing much quality check, giving rise to the spectacular spread of attractive but false beliefs about scientific findings.

### **Conspiracy theories**

In extreme cases, the interplay of our proclivities for narratives and threatening information, coupled with our promiscuous agency-detector and naïve sociological intuitions, derail into conspiracy theories that do not spare scientists. Most people see scientists as competent, but not necessarily trustworthy, individuals who are in a position to take advantage of the majority’s

*naiveté* for their own benefit [7]. This leads some individuals to ascribe hostile intentions and relationships to scientists that never existed [8]. Or, more frequently, people will simply jump to the conclusion that a given scientific finding cannot possibly be trusted merely because private agencies partly funded the research. Of course, conflicts of interest exist and can be very consequential. However, this heuristic-based reasoning becomes faulty when it is left unchecked and ignores the organizational functioning of science. Scientists indeed have very strong career incentives to disprove accepted findings—demonstrating that a widespread GMO corn is in fact dangerous would be worthy of an article in a top journal—and free scientific inquiry, provided material means are available, tends towards objectivity in the long run.

### **Addressing misunderstandings: recommendations to scientists**

Knowing all these facts, what can scientists do to counter misunderstandings and misinformation?

***Knowing psychology better.*** As a start, and regardless of the field they are working in, scientists should update their knowledge of human cognition, which will put them in a better position to take preemptive action. The English-speaking scholarship is rich with numerous scientific popularization essays, written to be accessible for wide audiences, that researchers can draw on (see Further Reading).

***Framing information to make it intuitive, in the right way.*** A second set of strategies to reduce risks that scientific research be misunderstood or twisted is to make it more intuitive by framing it in controlled ways. Research questions and findings will always have to be framed one way or another, and the ability of various frames to persuade and be memorized by target populations

should be tested experimentally. On the issue of increasing vaccines acceptance, for instance, it has been recently shown that highlighting that “90% of medical scientists are in favour of vaccination” (a positive frame) is more efficient than highlighting that “only 10% are against it” (a negative frame); contrary to the frequent finding that people are more sensitive to negative information, the former frame ended up being more convincing and more widely shared by participants. This result is however contingent and the opposite effect could be observed on other topics [9].

As a word of caution, scientists have to be careful when appealing to naïve theories in order to illustrate scientific notions. Metaphors are a priceless tool for quickly conveying representations, but, as simplifying analogies, they are susceptible to mislead the public into overlooking key aspects of the processes they were meant to explain. A telling example is the use of artificial selection as a metaphor for natural selection, used by Darwin himself, which had the negative side-effect of contributing to sustain the creationist notion that there is an intelligent designer behind biological evolution.

The inevitability of framing also means that scientists should pay special attention to headlines, whether of their own publications or of the universities’ press releases that cover their research. Avoiding misleading and exaggerating headlines is a way to mitigate misunderstandings from the start of the communication process, before the research gets interpreted by mass media and citizens.

Data presentation is another important issue since a graph is worth a thousand words. Graphs must give readers an accurate sense of effect sizes at stake by providing the entire y axis scale, rather than focus on differences between experimental conditions. Readers should be able to *see* that a “doubling” of the efficacy of available treatments is a modest achievement when it

amounts to moving from .1 to .2 on a scale from 0 to 5. As regards quantitative data explained to lay audiences, it is preferable to do without  $p$ -values, which require statistical education to be understood. It is also helpful to remind readers that a percent increase is not the same thing as a percentage *point* increase: moving up from 40 to 44% amounts to a 4 percentage points increase, but a 10% increase in the proportion measured.

***Actively responding to objections.*** Most importantly though, scientists need to be more active by spontaneously producing caveats, reminders, and counterarguments meant to preemptively debunk the most likely misunderstandings and objections. Some misunderstandings are quite general, for instance about what distinguishes science from mere intuition, in which case scientists should remind their audiences of why controlled experimentation, peer review, replication and so on can be trusted to generate reliable knowledge. When, by contrast, rejection of scientific findings is motivated by moral fears, caveats should target common moral concerns likely to surround the topic. In the case of a study on the genetics of, say, IQ or personality, scientists should highlight that a trait being heritable does not mean it is “100% determined by genes”, that causal explanations are not moral justifications, and that the best way of mitigating the societal impact of a genetic predisposition is to study it, not deny it.

As regards the locations where to express caveats and counterarguments, social media such as Twitter constitute important tools for reaching the public, not just for professional exchanges. Authors can also resort to dedicated sections in the scientific reports themselves to formulate their reminders, inspired for instance from *PNAS*' “Significance” boxes, or create websites or pages linked to the papers containing pedagogical images and videos. Face-to-face encounters with the public, for instance during science festivals or certain radio and TV shows, are also excellent occasions for scientists to change people's minds—including on such heated topics as GMOs, nuclear energy, or vaccination [10]. Contrary to unidirectional messaging typically

used in written science communication, informal interactive exchanges make it possible to address laypeople's objections precisely and in real time, and to undermine distrust by showing the public that scientists are just ordinary people with good intentions.

***Other recommendations.*** Lastly, the media and informal conversations mainly transmit attention-grabbing results, so failed replications and refutations are likely to fall under their radar. Scientists should therefore systematically mention them, both in order that public perceptions do not lag behind the actual state of the art, and to illustrate the Popperian logic of validation or refutation on which science constructs itself.

## **Conclusion**

Populated by many misleading naïve theories, a difficulty understanding the foundations of scientific expertise and conversely, a tendency to trust one's own intuitions too much and ignore one's own ignorance, a built-in myside bias, paranoid tendencies, a propensity to simplify when remembering and to exaggerate when communicating, and so forth, the human mind's evolved complexion is, to say the least, little predisposed to form accurate scientific beliefs. But this needs not be the end of the story, as many of those cognitive proclivities can, under certain conditions, be leveraged to favor accurate belief formation. Just as journalists and politicians, science communicators can for instance resort to framing to maximize intuitiveness and memorability of scientific information in relatively controlled ways. And, in analogy with professional scientists who get their methods and conclusions critically examined by others during the peer-reviewing process, laypeople's myside bias and misinformation can be overcome by science communicators motivated to engage with them actively and directly.

## **CONFLICT OF INTEREST:**

The authors declare having no conflict of interest.

## **ACKNOWLEDGEMENTS:**

The Département d'Etudes Cognitives (DEC) is supported financially by the EUR Frontiers in Cognition and the Idex PSL: ANR-17-EURE-0017.

## **FURTHER READING**

For an introduction to evolutionary psychology:

- Boyer, P. (2018) *Minds Make Societies*, Yale University Press.
- Haidt, J. (2012) *The Righteous Mind: Why Good People Are Divided by Politics and Religion*: Vintage Books.
- Mercier, H., & Sperber, D. (2017) *The Enigma of Reason*. Cambridge, MA: Harvard University Press.
- Pinker, S., (2002) *The Blank Slate: The Modern Denial of Human Nature*. New York, NY: Penguin
- Stuart Williams, S. (2018) *The Ape that Understood the Universe*. Cambridge University Press.
- Tooby, J. & Cosmides, L. (1990) The past explains the present. *Ethol. Sociobiol.* 11, 375–424.

On information transmission and evaluation:

- Mercier, H. (2020) *Not Born Yesterday. The Science of Who We Trust and What We Believe*. Princeton University Press.
- Morin, O. (2016) *How traditions live and die*. Oxford: Oxford University Press.

On the importance of evolved intuitions in explaining misbeliefs on scientific topics:

- Blancke, S., Tanghe, K. B., Braeckman, J. (2018) Intuitions in Science Education and the Public Understanding of Science, in Rutten, K., Blancke, S., Soetaert, R. (Ed.) *Perspectives on science and culture*. Purdue University Press.
- Gelman, S. (2004) Psychological essentialism in children. *Trends Cogn. Sci.* 8, 404–409.

On the illusion of understanding:

- Sloman, S. & Fernbach, P. (2017) *The Knowledge Illusion. Why We Never Think Alone*. MacMillan.

## References

1. Aarøe, L. & Petersen, M. B. (2018) Cognitive Biases and Communication Strength in Social Networks: The Case of Episodic Frames. *Br. J. Polit. Sci.* 1–21
2. Frimer, J. A., Skitka, L. J. & Motyl, M. (2017) Liberals and conservatives are similarly motivated to avoid exposure to one another's opinions. *J. Exp. Soc. Psychol.* 72, 1–12.
3. Blancke, S., Van Breusegem, F., De Jaeger, G., Braeckman, J. & Van Montagu, M. (2015) Fatal attraction: the intuitive appeal of GMO opposition. *Trends Plant Sci.* 20, 414–418.
4. Rozenblit, L. & Keil, F. (2002) The misunderstood limits of folk science: an illusion of explanatory depth. *Cogn. Sci.* 42.

5. Sumner, P. *et al.* (2014) The association between exaggeration in health-related science news and academic press releases: retrospective observational study. *BMJ* 349, g7015–g7015.
6. Moussaïd, M., Brighton, H. & Gaissmaier, W. (2015) The amplification of risk in experimental diffusion chains. *Proc. Natl. Acad. Sci.* 112, 5631–5636.
7. Fiske, S. T. & Dupree, C. (2014) Gaining trust as well as respect in communicating to motivated audiences about science topics. *Proc. Natl. Acad. Sci.* 111, 13593.
8. Scheufele, D. A. & Krause, N. M. (2019) Science audiences, misinformation, and fake news. *Proc. Natl. Acad. Sci.* 116, 7662–7669.
9. Altay, S. & Mercier, H. (2018) Framing messages for vaccination supporters. doi:10.31234/osf.io/ubp5s.
10. Altay, S. & Lakhlifi, C. (In press) Are Science Festivals a Good Place to Discuss Heated Topics? *Journal of Science Communication*. 19: A07 <https://doi.org/10.22323/2.19010207>



## Chapter 3: Moral conviction increases sharing preference for politically congruent headlines

This section is based on the following preprint:

“Moral conviction predicts sharing preference for politically congruent headlines”

OSF Preprint: [10.31219/osf.io/twq3y](https://osf.io/twq3y)

2020 (posted on OSF archive but not yet submitted)

Antoine Marie, Sacha Altay & Brent Strickland

Data, scripts, and materials to be published on the Open Science Framework

## **ABSTRACT:**

Political polarization in the United States may be due, in part, to partisans having sharply different perceptions of facts, as a consequence of being exposed to information streams that are very different in content and which rarely challenge their background beliefs. We approached the issue of biased access to political information by asking what kind of political headlines people preferentially share online. Across 8 studies, we focused on decisions to share news headlines touching on four controversial issues—gun control, abortion, sex equality and racial equality—on simulated social media. We found robust evidence that participants have a sharing preference for politically congruent news items over incongruent ones, and that this sharing bias increases with the moral importance of the issue, whether the news were true or false. Perceived accuracy and coalitional motivations to share headlines to advance political goals were among the main motivations to share. The transmission preference for congruent content and its interaction with issue importance held whether sharing was done from an anonymous or a personal social media account, and whether the audience was described as political congenial to the participant or not. Intervention messages reminding participants of their susceptibility to political bias had little moderating influence.

**Keywords:** Coalition; Fake News; Misinformation; Political bias; Polarization; Selective communication; Sharing; Social Media

## Introduction

In mass modern societies, political debate is partly shaped by information that is shared and discussed online. Thus any biases in how we consume and share information on the internet, even slight, have the potential to create large-scale impact.

One particularly salient type of information we readily encounter online pertains to divisive political debates. Such debates partly rest on irreducible disagreements about moral values, such as preferences in trade-offs between collective safety and individual liberty, or between equality and economic opportunity. But they also hinge on *factual claims*, the truth of which citizens and policy makers should be able to converge around: on the concrete consequences of public policies and technologies, on who-did-what-to-whom in the political arena, or on what is actually going on economically in the world. In an ideal world, citizens of democratic societies would have access not just to true information, but to all the important facts relevant to a given policy domain in order to form accurate opinions and pragmatically choose the programs and politicians best able to tackle them.

In practice, however, access to an impartial flow of accurate information is difficult. Over the course of their political socialization, individuals acquire contrasted views about many factual aspects of society (Finkel et al., 2020; Haidt, 2012; Kahan, 2013, 2016; Van Bavel & Pereira, 2018). In the US, liberals by and large believe that citizens' right to carry weapons can only increase crime, whereas many conservatives typically think this right will drive homicide rate down by allowing people to protect themselves. Liberals typically see free universal healthcare as a necessary condition for an economically efficient society, while conservatives by and large regard it as a counter-productive meddling in the market's laws (Rubin, 2003, 2014).

A bias for congenial information probably contributes to explain salient partisan differences in perceptions of social facts. Individuals tend to selectively expose themselves to

sources and content that are consistent with their prior political beliefs and identities over ideology-inconsistent information (Iyengar & Hahn, 2009; Stroud, 2008). They tend to preferentially befriend, work with, and marry people who share their moral and political worldviews (Alford et al., 2011; McPherson, Smith-Lovin & Cook, 2001). When exposure to information cannot be avoided, their myside bias (Mercier & Sperber, 2011; Nickerson, 1998; Stanovich & West, 2007) makes them more likely to discount ideologically incongruent information (Lord, Ross, & Lepper, 1979)—whether one views this tendency as rational Bayesian reasoning (Tappin, Pennycook, Rand, 2020), as mediated by an “argue-to-win” mindset (Fisher, Knobe, Strickland, & Keil, 2016), or as stemming from irrational emotion as in “motivated thinking” (Kahan, 2013, 2016; Kunda, 1990). Individuals also tend to remember pro-attitudinal political information better than counter-attitudinal information (Frenda, Knowles, Saletan, and Loftus 2013). Moreover, people’s propensity to selective exposure and the myside bias may be most pronounced on the polarizing issues (e.g. gun control, racial equality) on which they have strong moral convictions—for instance because they see them as sources of considerable societal costs or benefits, or as definitional to the identities of the groups they want to be part of (Haidt, 2012; Lord et al., 1979; Skitka, Bauman & Sargis 2005; Skitka & Mullen 2002; Van Bavel & Pereira, 2018).

Building on evidence of exposure to, and processing of, political information in ways partial to one’s priors, political interests and identity, recent work has shown that communication to others similarly favors individuals’ preconceived notions and goals. This has been found to be true both in real life group discussions, where individuals tend to convey information that confirms their preferences more than information that challenges them (Faulmüller et al., 2012), and in online environments, where individuals favor politically congenial sources (Bakshy et al., 2015; Barbera et al., 2015; Liang, 2018) and contents (An, Quercia & Crowcroft, 2014; Arendt et al., 2016; Ekstrom & Lai, 2020; Pogorelskiy & Shum, 2019; Shin & Thorson, 2017) over incongruent ones when deciding what to share.

To our knowledge, however, little research has examined how people's degree of moral conviction on an issue (Skitka, Bauman & Sargis, 2005; Skitka & Mullen, 2002) modulates their predisposition for selective communication of political information—a question this paper aimed to address. Operationally, we restricted the scope of our studies to sharing decisions on simulated social media of political news items touching on 4 controversial issues—gun control, abortion rights, sex equality, and racial equality—on which US liberals and conservatives can be expected to be polarized (Finkel et al., 2020; Pew Research Center, 2014; Van Bavel & Pereira 2018). Social media allows one to post political news articles to one's network at no cost, and thus has become an important source of citizens' political information consumption. We selected news stories that make factual claims only—they are not opinion pieces—that are typically congenial either to liberals or conservatives on each one of the 4 issues. All our surveys were designed in Qualtrics and all participants were American residents recruited on Amazon Mechanical Turk.

Experiments 1-3 explored how the political congruence (*vs.* incongruence) of a news headline to a participant's ideology, as well as the moral importance of the issue it touches on, predicted sharing decisions. Our first research interest was to see whether our design, based on news sharing on simulated social media, would reproduce earlier findings that participants favor politically congruent content when deciding what to share. Our second goal was to examine how sharing decisions would interact with participants' degree of moral conviction on the issue covered by the news, or, in other words, with the issue's moral importance to the participant.

Next, Experiment 4 investigated participants' explicit motivations for sharing news. Finally, Experiments 5 to 8 tested various experimental manipulations meant to test the robustness of the sharing patterns observed in Experiments 1-3. Experiment 5 explored whether the composition of participants' audience (politically congenial *vs.* uncongenial) affected their sharing decisions. Experiment 6 examined whether sharing from an anonymous *vs.* personal social media account would influence sharing decisions. Finally, Experiments 7 and 8 tested

the efficacy of a simple intervention message, informing participants of their propensity to political bias, in mitigating their sharing preference for congruent headlines.

### **Experiments 1-3: Willingness to Share to Friends**

Experiments 1 to 3 explored willingness to share real political news headlines as a function of a news item's congruence with each participant's position on the issue, and the moral importance of the issue to the participant. Willingness to share was measured using a 4-point scale in Experiment 1, and a dichotomous scale in Experiment 2. Experiment 3 used the same design and measures as Experiment 2 to explore willingness to share false political news items.

## **Method**

### ***Participants and design***

331 U.S. participants (Experiment 1), 318 participants (Experiment 2) participants and 433 participants (Experiment 3) were recruited on Amazon Mechanical Turk (henceforth MTurk), in exchange for a small compensation (\$.60). In Experiments 1 and 2, participants were randomly exposed to 8 real political news items touching on 4 controversial issues and to 4 real neutral news items, in within-subjects design, and asked to report their willingness to share each news. Experiment 3 adopted the same design as Experiments 1 and 2 but the stimuli was 8 false political headlines (no neutral news), which touched on the same 4 controversial issues. Experiments 1-3 also collected information on participants' political attitudes to determine whether each news item was congruent or incongruent, and whether it touched on a low or high moral importance issue.

### ***News items database***

The real or true news items used in Experiments 1 and 2 were actual headlines taken from press articles found on mainstream news media websites. The false news items of Experiment 3 were taken from prominent fact checking websites (e.g. Snopes). Each headline's titles was followed by a short introductory snippet and a picture chosen to illustrate its content (see Figure 1). We removed information on the items' sources and gave them the same press-looking display. News items can be found in the Electronic Supplementary Materials, henceforth ESM, after the References section.

All political news items, whether real (Experiment 1-2) or false (Experiment 3), were organized in pairs touching on one controversial issue: gun control, racial equality, gender equality and abortion. All political headlines had a factual tone: they reported a societal event or fact, as opposed to expressing opinions. One news story on each issue was typically congruent for liberals on the issue (e.g. gun control supporters), and one news story was typically congruent for conservatives on the issue (e.g. gun rights supporters). More specifically, there was two ways in which a news story was defined as being "congruent" for a given political side or coalition: either because the news made a factual claim that *fits the typical prior beliefs* of that coalition (e.g. for gun rights supporters, that gun control does not reduce crime), or because the news portrays an event that *is good news* for that coalition's political goals (e.g. for pro-choice supporters, that abortion has become free in Ireland) (see ESM).

To be retained in the stimuli database, the political headlines, whether real (Experiment 1-2) or false (Experiment 3), had to clearly be identified by independent MTurk raters as favoring one political side on each issue. We used the following question to probe participants on their perception of each item's slant: "*Assuming that this news story is true, is it more favorable to Democrats or Republicans?*", with responses collected on a 0-100 slider scale ranging from 0, "*Democrats*" to 100, "*Republicans*", with 50 as default slider position. False headlines had to respect the additional criterion of falling within an intermediate level of perceived plausibility, namely between "*Somewhat inaccurate*" and "*Somewhat accurate*", on

a 7-point scale, in response to the question “*How accurate do you think this news story is?*” (see ESM).

The 4 non-political, neutral news items used in Experiments 1 and 2 were used as control items. They relate historical or biological facts that do not advance any particular political narrative or interest.



**Abortion will soon become cost-free in Ireland**

Until recently, abortion was illegal in Ireland. However, according to Health Minister Simon Harris, a new law is about to be submitted to the legislature in October proposing to make abortion cost-free...



**New evidence suggests a majority of abortion cases are on-demand and do not involve rape victims**

One of the most common reasons given to justify abortion is that it may allow mothers to postpone childbearing to a more suitable time and provide already born children with more care and resources...



**Allowing abortion for rape victims will only incentivize women to report fictitious rapes, Pence said off the record**

“The only thing allowing abortion for rapes can do is to motivate women to falsely report having been raped” said Vice President Pence in...



**Bernie Sanders wants to set cut off date for abortions up to 7 months**

Says Bernie Sanders wants “to make the cut off date for abortions up to 28 weeks”, arguing “a baby at that point feels no pain.”

**Figure 1:** Examples of real political headlines, typically congruent for liberals (top left) and conservatives (top right) on the issue of abortion, and of false political headlines, typically congruent for liberals (bottom left) and conservatives (bottom right), on the same issue.

### ***Materials and procedure***



Experiments 1-3 started by requesting informed consent and displayed the following message: *“You will be asked to answer a few questions about news stories that were found on the internet. The reliability of the news items’ sources is uncertain. The sources will not be made visible to you.”* Participants then saw the 8 real political and 4 real neutral headlines in Experiments 1 and 2, and the 8 false political headlines in Experiment 3, in a random order. Each headline, its picture, and the willingness to share question were displayed on the same page. The willingness to share question read: *“How likely would you be to pass along this news item to friends or people you like?”*. It offered 4 choices in Experiment 1 (1, *“Very unlikely”*, 2, *“Unlikely”*, 3, *“Likely”*, 4, *“Very Likely”*) and a dichotomous 0, *“Not Share”* vs. 1, *“Share”* alternative in Experiments 2 and 3, so as to better match real social media use.

When participants were done going through the headlines, they responded to two types of political attitudes questions. The first type of attitude information we collected was a participant’s position on each one of the 4 political issues. All responses were collected using 0-100 slider scales ranging from 0, the most conservative position on each issue, to 100, the most liberal position, with 50 as default slider position. In Experiments 1 and 2, the attitude questions on racial and gender equality were worded *“To what extent do you think [gender equality] should be the government’s priority?”*, with responses anchored in 0, *“I don’t care at all”* and 100, *“Absolute priority”*. In Experiments 1 and 2, the attitude questions on the gun control and abortion issues were worded *“On the issue of abortion, would you consider yourself as being:”*, and responses ranged from 0, *“Absolutely Pro Life”* and *“Absolutely Pro Gun rights”*, to 100, *“Absolutely Pro Choice”* and *“Absolutely Pro Gun control”*, respectively. We had initially adopted two formulations because we thought that issues such as abortion and gun control allow for clearly polarized “for” vs. “against” positions, while on issues like gender and race equality, the conservative position is more often expressed in terms of indifference than opposition. However, in Experiment 3 and all the other studies in this paper, the 4 attitude questions on each issue were simplified into a unique formulation:

“*What is your position on the issue of [e.g. gender equality]?*”, and response choices were kept the same as before. Information on participants’ position on each one of the 4 issue was later used to determine whether each political news item was congruent or incongruent to the participant (see Results section for how this was done).

The second type of attitude information we collected was whether the issue a headline covered was of high or low moral importance to the participant, or, in other words, whether the participant had a high degree of moral conviction on the issue or not. In Experiments 1-3, issue importance was probed with the question: “*Is one of the following debates or issues of absolute importance to you? (several choices possible)*”. Responses were collected as a set of binary outcomes by participants ticking boxes next to each issue label (“*Gun Control*”, “*Racial Equality*”, “*Gender Equality*”, “*Abortion*”) or specifying other issues in text format (“*Other: \_\_*”). We interpreted ticking of a box as indicating high moral importance, and no ticking as meaning the issue was of low importance.

The questionnaires also measured political ideology in general on a one item slider scale anchored in 0, “*Strongly Democrat*” and 100, “*Strongly Republican*” with 50 as default position. The studies ended with demographic questions (sex, age, level of education, vote at the last general and mid-term election) and an attention check used in all the studies reported in this paper (see “videogame” in ESM).

## Results

33 participants were excluded from Experiment 1 because they either passed the questionnaire twice (3), stopped during completion (1), or failed at the attention check (29 individuals), leaving 298 participants ( $M_{age} = 34.7$ ,  $SD_{age} = 9.6$ , 47.9% female). 12 participants were excluded from Experiment 2 because they either passed the questionnaire twice (4), or failed the attention check (12 people), so 302 participants remained in the dataset ( $M_{age} = 39$ ,

$SD_{age} = 12.3$ , 49% female). Experiment 3 was the control condition of an intervention experiment, Experiment 8, presented below. Participants who passed the questionnaire twice, failed the attention check, or stopped during completion were deleted, leaving 330 participants in Experiment 3 ( $M_{age} = 38.5$ ,  $SD_{age} = 12.6$ , 42% women).

In order to facilitate interpretation of the data, participants' scores on the 4 attitude questions were reverse coded prior to analysis, such that 0 now represented the most liberal position, 100 the most conservative position on each issue, and 50 the middle of the scale. This information was used to determine whether each political news item was congruent or incongruent for each participant. Our method was the following: a political news story advancing a conservative narrative on an issue was considered as *congruent* to a participant if that participant's attitude on the issue was comprised between [50, 100], and as *incongruent* if the participant's attitude on the issue was comprised between [0, 50]. The reverse was true for headlines advancing a typically liberal narrative.

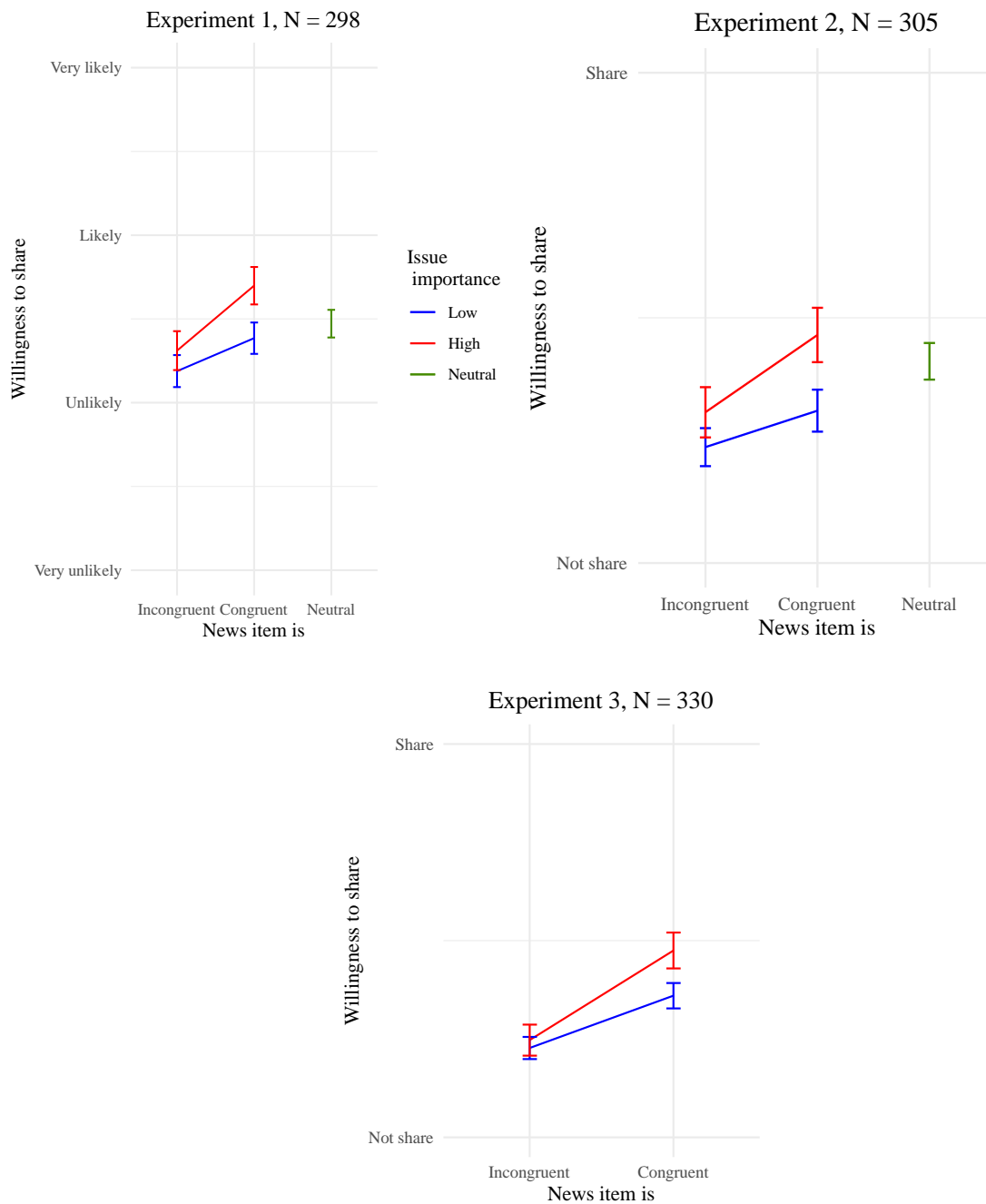
All the data analyses in this paper were performed in R (version 3.6.2) using R Studio (Version 1.2.5033). In Experiments 1-3, our default analysis strategy was to run mixed-effects models (using the "lmer4" package) because each participant rated several news headlines (12 in Experiments 1-2 and 8 in Experiment 3). Willingness to share news item was the dependent variable, news item congruence (dichotomous) and issue importance (dichotomous) were fixed effects, and participant was the random effect. Main effects of news item congruence and issue importance are reported from models containing only the main effects, and the interaction of news item congruence and issue importance is reported based on models containing the main effects and their interaction (see ESM for regression tables). We report standardized regression coefficients, coefficients' 95% confidence intervals between brackets, and *ps*. When this was relevant, we also compared means of specific groups using paired *t*-tests after having averaged repeated measures at the relevant level.

In all three studies, participants reported higher willingness to share politically congruent than incongruent news headlines (i.e. main effect of a news item being congruent *vs.* incongruent: Experiment 1:  $\beta = 0.26$ , [0.19,0.32],  $p < 0.001$ ; Experiment 2:  $\beta = 0.21$ , [0.14,0.28],  $p < 0.001$ ; Experiment 3:  $\beta = 0.36$ , [0.29,0.42],  $p < 0.001$ ). Moreover, willingness to share incongruent headlines was consistently greater on high than on low importance issues (main effect of high *vs.* low issue importance: Experiment 1:  $\beta = 0.20$ , [0.07,0.33],  $p < 0.001$ ; Experiment 2:  $\beta = 0.18$ , [0.04,0.32],  $p < 0.01$ ; Experiment 3:  $\beta = 0.20$ , [0.07,0.33],  $p < 0.001$ ). Most importantly, the difference between willingness to share congruent and incongruent news items was always greater on high than on low importance issues (interaction between news congruence and issue importance: Experiment 1:  $\beta = 0.17$ , [0.10,0.24],  $p < 0.001$ ; Experiment 2:  $\beta = 0.18$ , [0.10,0.25],  $p < 0.001$ ; Experiment 3:  $\beta = 0.14$ , [0.07,0.22],  $p < 0.001$ ). See Figure 2, and ESM for regression tables.

The increased sharing preference on high (rather than low) importance issues was mostly driven by *congenial* items (Experiment 1:  $t(776) = 4.81$ ,  $p < 0.001$ ,  $d = 0.3$ ; Experiment 2:  $t(859) = 4.49$ ,  $p < 0.001$ ,  $d = 0.27$ ; Experiment 3:  $t(909) = 4.03$ ,  $p < 0.001$ ,  $d = 0.23$ ). By contrast, willingness to share *incongruent* items only increased very marginally, if at all, when issue importance increased (Experiment 1:  $t(775) = 1.60$ ,  $p = 0.10$ ,  $d = 0.09$ ; Experiment 2:  $t(870) = 1.74$ ,  $p = 0.08$ ,  $d = 0.10$ ; Experiment 3:  $t(917) = 0.81$ ,  $p = 0.4$ ,  $d = 0.05$ ).

Overall, in Experiments 1 and 2, willingness to share was *lower* for political headlines than for real neutral ones, with the exception of politically congruent news items on high importance issues.

Interestingly, in Experiments 1 and 2 on real news headlines, participants who had a strongly liberal stance were equally likely to share politically congruent *and* incongruent headlines on low importance issues (see ESM for graphs showing this effect).



**Figure 2:** Willingness to share news item as a function of headline congruence and issue importance in Experiments 1, 2 (real news) and 3 (false news). Means are surrounded by 95% CIs.

### Experiment 4: Motivations to Share News Headlines

Experiment 4 examined the conscious motivations driving headlines sharing in our simulated social media. Our hypotheses were mainly about political items. We expected three main motivations: perceived accuracy, perceived relevance or informativeness, and usefulness of a headline for participants' political commitments.

## **Method**

### ***Participants and design***

421 US participants were recruited on MTurk to participate in Experiment 4. The study built on Experiment 1's design: 8 real political and 4 real neutral news items exposed in a random order (within-subjects), with willingness to share measured on a 4-point scale. This design was complemented by 3 additional questions on participants' motivations to share. In order to compensate fatigue caused by the additional questions, participants were randomly allocated to one of two groups which exposed them to only 6 of the 12 news items.

### ***Materials and procedure***

The procedure was identical to Experiment 1's except for the fact the willingness to share question was followed by 3 questions on motivations to share, displayed on the same page. Responses to these additional questions were all collected on 0-100 slider scales with 50 as default choice. The first question asked participants to rate the perceived accuracy of the headline ("*How accurate do you think this news story is?*", from 0, "*Not at all accurate*", to 100 "*Completely accurate*"). The second question asked about its informativeness ("*How informative do you find that news item?*", from 0, "*I knew that already*" to 100, "*Extremely informative*"). The third question asked about the headline's usefulness for the participant's

personal political goals (“*How useful is this piece of news for your personal political commitments?*”, from 0, “*Not at all*” to 100, “*Extremely useful*”).

## Results

94 participants were removed because they had already taken one of our questionnaires (82), or failed the attention check (12), leaving 315 participants in the database. Following the same analysis strategy as in Experiments 1-3, we found that participants showed a sharing preference for politically congruent over incongruent news items (main effect of news congruence:  $\beta = 0.19$ , [0.11,0.27],  $p < 0.001$ ), as well as for news touching on high rather low importance issues (main effect of issue importance:  $\beta = 0.18$ , [0.08,0.27],  $p < 0.001$ ). The sharing preference for congruent items increased when moving from low to high importance issues, but the interaction between news congruence and issue importance was not significant ( $\beta = 0.13$ , [-0.03,0.29],  $p = 0.12$ ).

As regards reported motivations to share, the strongest predictor of willingness to share was headlines’ usefulness for one’s political commitments, whether the headlines were politically incongruent ( $\beta = 0.42$ , [0.34,0.49],  $p < 0.001$ ), congruent ( $\beta = 0.40$ , [0.32,0.48],  $p < 0.001$ ), or, surprisingly, neutral ( $\beta = 0.35$ , [0.27,0.43],  $p < 0.001$ ). Perceived informativeness or relevance was the second strongest predictor of sharing of incongruent ( $\beta = 0.24$ , [0.17,0.32],  $p < 0.001$ ) and congruent headlines ( $\beta = 0.22$ , [0.14,0.29],  $p < 0.001$ ), as well as neutral items ( $\beta = 0.24$ , [0.16,0.31],  $p < 0.001$ ). Perceived accuracy also significantly predicted sharing of incongruent ( $\beta = 0.14$ , [0.07,0.21],  $p < 0.001$ ), congruent ( $\beta = 0.17$ , [0.10,0.24],  $p < 0.001$ ), and neutral headlines ( $\beta = 0.19$ , [0.12,0.26],  $p < 0.001$ ). Time spent viewing the headline did not predict willingness to share any type of news (incongruent news:  $\beta=0.01$ ,  $p=0.77$ ; congruent news:  $\beta=0$ ,  $p = 0.96$ , neutral news:  $\beta=0.04$ ,  $p=0.14$ ). See ESM.

## Discussion of Experiments 1-4

Experiments 1-4 repeatedly found that US MTurkers were more likely to share politically congruent than incongruent news items, and that this preference was magnified by increased moral importance of the issue. Our studies conceptually replicate prior evidence of selective communication on political topics with a novel design, inspired from social media environments (Ekstrom & Lai, 2020; Faulmüller et al., 2012; Bakshy et al., 2015; Barberà et al., 2015; Liang, 2018; Arendt et al., 2016; Pogorelskiy & Shum, 2019; Shin & Thorson, 2017), and uniquely show that moral conviction on an issue operates as a magnifier of this tendency.

What explains the imbalance in sharing between politically congruent and incongruent headlines? First, it is important to note that, in line with most research on social media, and contrary to what is actually possible to do on Facebook and Twitter, our design did not allow for shares to be accompanied by a comment, e.g. a critical one. To the extent that the act of sharing an article online is interpreted as endorsement of its claim (so long, at least, as the claim is in the vicinity of what the person may plausibly believe), the impossibility to leave a critical comment on headlines may have de-incentivized sharing of politically incongruent content.

Regarding the politically congruent headlines, they were selected to be plausible within a given group given their typical prior beliefs—e.g. pro-choice supporters, gun rights supporters. The real news were by definition true, so they were plausible, and the false news were selected to fall within an intermediate level of credibility. It is no surprise therefore that in line with earlier work (Altay, de Araujo, Mercier, 2020; Tappin, Pennycook, Rand, 2020), perceived accuracy predicted sharing of congruent political news headlines ( $\beta = 0.17, p < 0.001$ ).

However, truth is not a sufficient reason for information to be communicated, and there is a lot of evidence that communication on politics is also driven by instrumental, coalitional goals. Humans, when they have moral and political convictions, intuitively behave as team



members. They spontaneously frame polarizing issues in terms of group conflict and are concerned to secure social support and to promote their political coalition's interests (Altay, de Araujo, Mercier, 2020; Osmundsen et al., 2020; Pennycook, Epstein, et al., 2020; Pennycook, McPhetres, et al., 2020; Petersen, Osmundsen, Tooby, 2020; Tooby & Cosmides, 2010; Tooby, 2017). Our experiments were specifically meant to activate those instincts, both with the choice of the polarizing societal issues—interpreted by the mind as instances of group conflict—and because the audience of news sharing in Experiments 1-4 was assumed to be politically congenial to participants—and consequently constitutes allies with whom bonds should be maintained.

In line with this coalitional perspective, participants reported a congruent headline's usefulness for their political commitments to be the strongest motivation for sharing it in Experiment 4 ( $\beta = 0.40, p < 0.001$ ). More specifically, many politically congruent headlines point at threat (such as the real headline denouncing racism in sports commentary for liberals on racial equality; see ESM). In a coalitional perspective, threat-relevant information may be used to mobilize one's group against an enemy, or to signal to allies that one cares about their welfare (Petersen et al., 2020). Other headlines are congruent to a coalition because they highlight successes for that coalition (such as the one announcing free abortion procedures in Ireland for pro-choice supporters). Sharing them may be done to persuade new allies to join the cause and motivate those already involved, as well as to signal devotion to the coalition to enhance one's moral standing. In turn, the fact that participants showed a greater sharing preference for congruent content on high than low importance issues may suggest that greater moralization of a given issue makes people *more* proselyte.

Interpreting why participants did sometimes share politically incongruent headlines is more difficult than explaining why they shared congruent ones. A disinterested motivation to foster political debate may have played a role. Denouncing the events depicted—e.g., free abortion in Ireland for pro-life conservatives—likely was a reason too, despite the fact that, as

already noted, doing so without a disambiguating comment might have opened the risk of being perceived as endorsement.

\*\*

The remainder of the paper presents studies meant to test the robustness of the transmission preferences documented in Experiments 1-4. Experiment 5 tested whether manipulation of the political congeniality *vs.* uncongeniality of the fictitious social media's audience could affect sharing decisions. Experiment 6 examined whether manipulating the anonymous *vs.* personal character of the social media account from which news items are shared would influence transmission. Finally, Experiments 7 and 8 tested whether informing participants of their susceptibility to political bias when processing political information could reduce their sharing preference for politically congenial content.

### **Experiment 5: Politically Congenial *vs.* Uncongenial Audience**

All the studies so far put participants in the mindset of sharing news stories to “*friends or people [they] like*”. There is a wealth of evidence that most people tend to interact with people whose opinions by and large align with theirs (e.g. Van Bavel & Pereira, 2018), so most of those “friends” could reasonably be assumed to be close to the participant’s ideology. There. With Experiment 5, however, we chose to explicitly vary the political congeniality of the audience. Compared to a politically congenial one, we expected a politically *uncongenial* audience to stimulate sharing of incongruent content and to inhibit sharing of congruent headlines.

## Method

### *Participants and design*

290 US MTurkers were recruited for Experiment 5. The study built on the same design as Experiment 2: 8 real political and 4 real neutral headlines exposed in a random order (within-subjects), and willingness to share measured on a dichotomous scale. This time, however, participants were randomly allocated to one of two conditions: sharing to people they agree vs. disagree with politically.

### *Materials and procedure*

Experiment 5 used the same true headlines as Experiments 1 and 2. Depending on condition, the willingness to share question following each news item was worded as follows: “*Would you share this news story to people you agree with politically?*” vs. “*Would you share this news story to people you disagree with politically?*”. Responses were given on a dichotomous scale “*Not Share*” vs. “*Share*”.

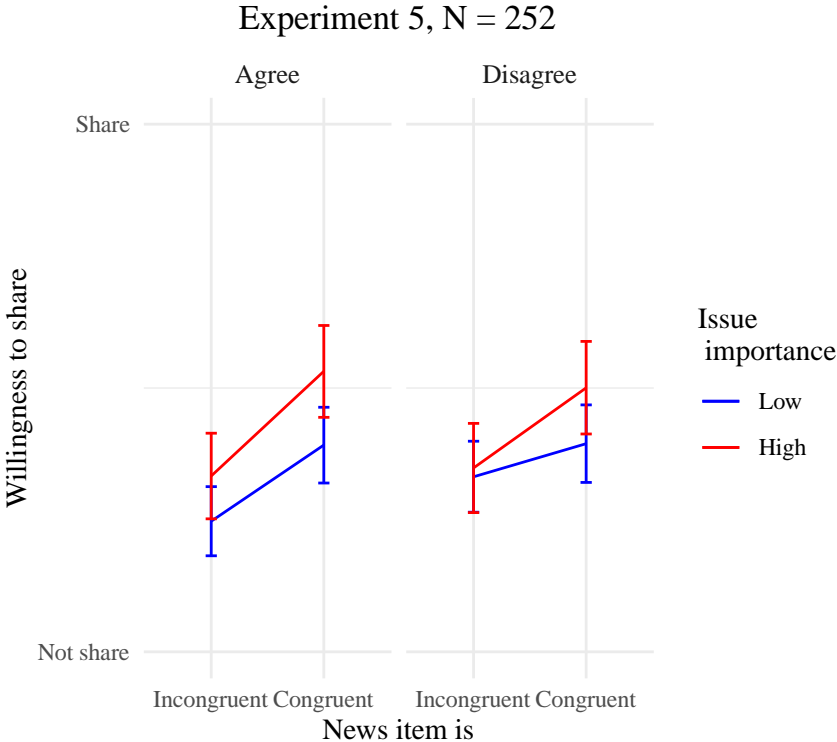
## Results

39 participants were removed from the data because they passed the questionnaire twice (7), were the result of preview trials (1), failed the attention filter (11), or had already passed one of our questionnaires (20), leaving 252 valid participants ( $M = 36.3$ ,  $SD = 11.2$ , 46% female).

Neutral headlines were not included in the analyses as the experimental manipulation was targeting sharing of political headlines. As in previous studies, politically congruent news stories were more shared than incongruent ones (main effect of news congruence:  $\beta = 0.26$ ,  $[0.19, 0.34]$ ,  $p < 0.001$ ), and moving from a low to a high importance issue increased willingness to share (main effect of issue importance:  $\beta = 0.19$ ,  $[0.10, 0.27]$ ,  $p < 0.001$ ). Sharing was highest for congruent content touching on high importance issues, but the interaction between news

congruence and issue importance was not significant ( $\beta = 0.12$ ,  $[-0.02, 0.27]$ ,  $p = 0.10$ ). Willingness to share in the politically congenial and in the politically incongruent audience conditions did not differ (main effect of experimental condition:  $\beta = 0.06$ ,  $[-0.09, 0.22]$ ,  $p = 0.44$ ). See ESM.

In order to facilitate interpretation of the results, the effect of experimental condition was assessed on politically incongruent and congruent news separately, using two different mixed-effects models (see ESM). As regards incongruent headlines, the only effect that moving from a politically congenial to a politically incongruent audience seemed to have was to slightly increase willingness to share, but the effect was not significant (main effect of condition:  $\beta = 0.15$ ,  $[-0.03, 0.32]$ ,  $p = 0.10$ ). Moving from a politically congenial to a politically incongruent audience did not decrease sharing of congruent headlines (main effect of condition:  $\beta = -0.02$ ,  $[-0.20, 0.16]$ ,  $p = 0.86$ ). See Figure 3.



**Figure 3:** Manipulations of the political congeniality of the audience (Experiment 5). Means are surrounded by 95% CIs.

## **Experiment 6: Anonymous vs. Personal Account**

Our studies so far implicitly put participants in the mindset of sharing news from their own social media account. However, we know that humans have an intuitive ability to inhibit certain actions to preserve their reputation, in particular in contexts where there are witnesses (Mercier & Sperber, 2011; Sperber & Baumard, 2012). Experiment 6 tested whether the sharing preference for congruent headlines and its interaction with issue importance would be affected by whether the headlines are shared from the participant's personal social media account vs. from an *anonymous* account on which their reputation is not at stake. We expected that decreasing reputational concerns would increase sharing of politically congruent news stories, as it might make participants less watchful of the potential reputational costs of passing along controversial content.

## **Method**

### ***Participants and design***

325 US participants were recruited on MTurk to pass Experiment 6. We used the same 8 real political and 4 neutral headlines as in Experiments 1-2, and willingness to share was measured on a dichotomous scale. However, participants were randomly assigned either to an anonymous account or a personal account condition, following a mixed design.

### ***Materials and procedure***

The procedure was identical to Experiment 2 except for the following changes. Experiment 6 started with an English proficiency check in which participants were asked to

describe what they had done the previous day in 150 characters. The goal was to eliminate people unable to understand the headlines. This check was systematically included in the following studies.

Our previous designs had implicitly assumed that news sharing was done from a *personal* social media account. Asking participants to imagine they would now be sharing news items anonymously would imply that their account would potentially have no followers, which was unecological (there is *prima facie* no reason to follow an anonymous account online, unless its content is especially interesting). In both conditions, the questionnaire therefore began with a vignette asking participants to imagine they were working for an association that gives them money to promote any social cause they care about through news sharing (see ESM for vignettes). In both the personal account and the anonymous account conditions, sharing was made symbolically costly by telling participants they were given a fictional \$100 endowment per news item. They were informed that they could spend whichever proportion of that endowment to promote content online, and keep the rest for themselves. The dependent variable was now a willingness to pay for sharing each headline.

To emphasize anonymity in the anonymous account condition, participants were told they were working for an association called “Public Support”, and shown its (fictitious) Twitter home page at the beginning of the questionnaire (see ESM). In contrast, the personal account condition did not mention any association name nor any Twitter account, and the vignettes

made it clear that news sharing would be done from participants' own social media account.

Below are the vignettes employed in both conditions:

***Anonymous account***

Imagine you work for an association called 'Public Support', that gives you a budget for promoting any social cause you care about through sharing news online. The budget the association gives you is limited, but allows you to have a social impact on potentially thousands of users through news sharing.

To promote content you care about, you are *using the association's social media account, not your own*. This means that *your promotion is done anonymously: no one will be able to match the content you publish to your name*.

["Public Support" Twitter account home page displayed, see ESM]

Currently preparing your *anonymous* promotion, you are about to be presented with a series of news items to potentially share online. For each of them, you are given an endowment of \$100, from which you must decide how much money you are willing to spend in order to share the news (thereby promoting its content), or how much to keep for yourself. For instance, if you decide to invest \$60 in order to promote a news item on a topic you care about, that means you can keep the remaining \$40 for yourself.

Note that the reliability of the news' sources is uncertain. The sources will not be made visible to you.

["Public Support" Twitter account home page displayed, see ESM]

***Personal account***

Imagine you work for an association that gives you a budget for promoting any social cause you care about through sharing news online. The budget the association gives you is limited, but allows you to have a social impact on potentially thousands of users through news sharing.

To promote content you care about, you are using *your own social media account* (but the money you are investing comes from the association, not your own pocket). This means that *all your friends will be able to match the content you publish to your personal identity*.

[No picture displayed]

Currently preparing your *nominative* promotion, you are about to be presented with a series of news items to potentially share online. For each of them, you are given an endowment of \$100, from which you must decide how much money you are willing to spend in order to share the news (thereby promoting its content), or how much to keep for yourself. For instance, if you decide to invest \$60 in order to promote a news item on a topic you care about, that means you can keep the remaining \$40 for yourself.

Note that the reliability of the news' sources is uncertain. The sources will not be made visible to you.

[No picture displayed]

In the anonymous account condition, the willingness to pay for sharing question displayed under each item read: "*From a \$100 provided by the association, how much would you spend on sharing this news? (from an anonymous social media account, no one will know it's you)*". In contrast, in the personal account condition, the question read: "*From a \$100 provided by the association, how much would you spend on sharing this news item? (from your*

*personal social media account, all your friends will know*)”. Responses were collected using 0-100 slider scales anchored in 0, “*Keep all the money*” and 100, “*Spend all the money*”, with 50 as default slider position.

## Results

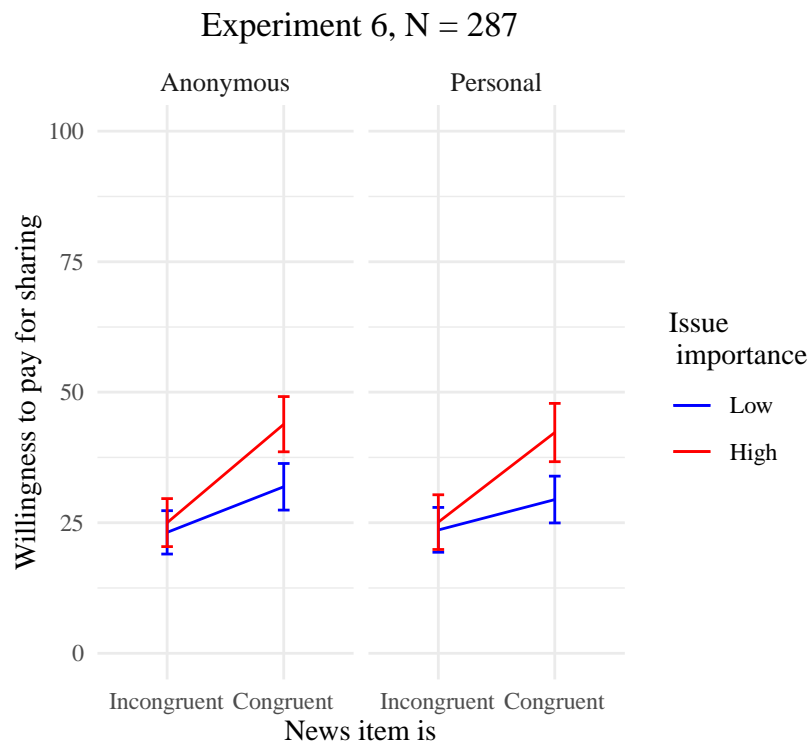
We excluded 38 participants who either failed the English proficiency test or did not fill the entire questionnaire (24), or failed the attention check (14). 287 participants remained in the dataset ( $M_{age} = 37$ ,  $SD_{age} = 10.8$ , 46% women).

The data was analyzed without including the neutral items, as the hypothesis driving the experimental manipulation—greater sharing of congruent items in the anonymous account than in the personal account condition—only concerned the political headlines. To simplify reading, “willingness to pay for sharing” is abbreviated as “willingness to share” or “sharing”. Using the same analysis strategy as in previous studies, we found that congruent news items were more shared than incongruent ones (main effect of news congruence:  $\beta = 0.34$ , [0.28,0.40],  $p < 0.001$ ), and that news on high importance issues were more shared than those on low importance issues (main effect of issue importance:  $\beta = 0.20$ , [0.13,0.27],  $p < 0.001$ ). The sharing preference for congruent news was again greater on high than low importance issues (interaction between news congruence and issue importance:  $\beta = 0.34$ , [0.22,0.46],  $p < 0.001$ ). Overall willingness to share in the personal account condition did not differ from sharing in the anonymous account condition (main effect of condition:  $\beta = -0.03$ , [-0.19,0.13],  $p = 0.68$ ).

The effect of our experimental manipulation was assessed using distinct mixed-effects models for incongruent and congruent headlines (see ESM). Having one’s reputation on the spot by sharing headlines from one’s personal social media account rather than from an anonymous account had almost no effect. It did not affect sharing of incongruent items (main



effect of condition:  $\beta = 0.02$ ,  $[-0.16, 0.21]$ ,  $p = 0.80$ ), nor did it significantly decrease sharing of congruent items (main effect of condition:  $\beta = -0.08$ ,  $[-0.25, 0.09]$ ,  $p = 0.37$ ). See Figure 4.



**Figure 4:** Manipulation of account anonymity (Experiment 6). Means are surrounded by 95% CIs.

### Experiments 7 and 8: Intervention Message

Psychological research has recently explored ways of influencing behavior deemed problematic through nudge interventions. For instance, Pennycook et al. (2020) recently found evidence that people’s willingness to share false news in online studies can be reduced by priming them with a concern for accuracy. Experiments 7 and 8 examined whether informing participants of their inclination to process political information in ways partial to their prior beliefs and interests, and spotlighting the issues on which they were most likely to fall prey to it, might influence sharing decisions. We made two hypotheses: that the intervention message

would decrease sharing of politically congruent headlines, and that it would increase sharing of incongruent ones.

## Method

### *Participants and design*

1166 participants were recruited on MTurk to pass Experiment 7, the intervention experiment on real news items. Participants in the control condition merely reported their willingness to transmit each one of the 8 real political news stories (no neutral news), presented in a random order, whereas in the intervention condition, the headlines were preceded by a message about our inclination to process political information in ways partial to our political values. Experiment 8 recruited 937 participants and followed the same design as Experiment 7 but on false news items.

### *Materials and procedure*

The procedure for Experiments 7 and 8 was identical to previous studies except that both experiments contained a control condition with no message and an experimental condition with the intervention message. The message was inspired from scientific research on politically motivated thinking and provided a scientific reference (Kahan, 2016). It warned participants of a tendency to “*favor information that fits one's goals and values, and to disregard information that doesn't*”, in particular on the issues one regards as being “of absolute importance” (i.e. high moral importance issues). In order for the message to embed the labels of those latter issues, participants were asked to report their political attitudes on each issue, and the “absolute importance” multiple choice question, *before* reaching the randomizer allocating them to one of the two conditions (control, or intervention message). Below is the message we used (emphasis appeared as bold text in questionnaire):

Previous research has shown that the issue(s) you judge as having absolute importance are the ones on which you are *most likely to be politically biased*. In your case, they are:

>>> [Issues rated as being “of absolute importance” displayed] <<<

Typically, political bias causes one to significantly *favor information that fits one's goals and values, and to disregard information that doesn't*.

Source:

Kahan, D. (2016). The Politically Motivated Reasoning Paradigm, *Emerging Trends in the Social and Behavioral Sciences*, 1-16.

While the message was clear as to which of the 4 issues (covered by the headlines) participants were most likely of being politically biased on (given their own responses), we eschewed specifically mentioning the sharing preference for politically congenial information in order to mitigate task demands.

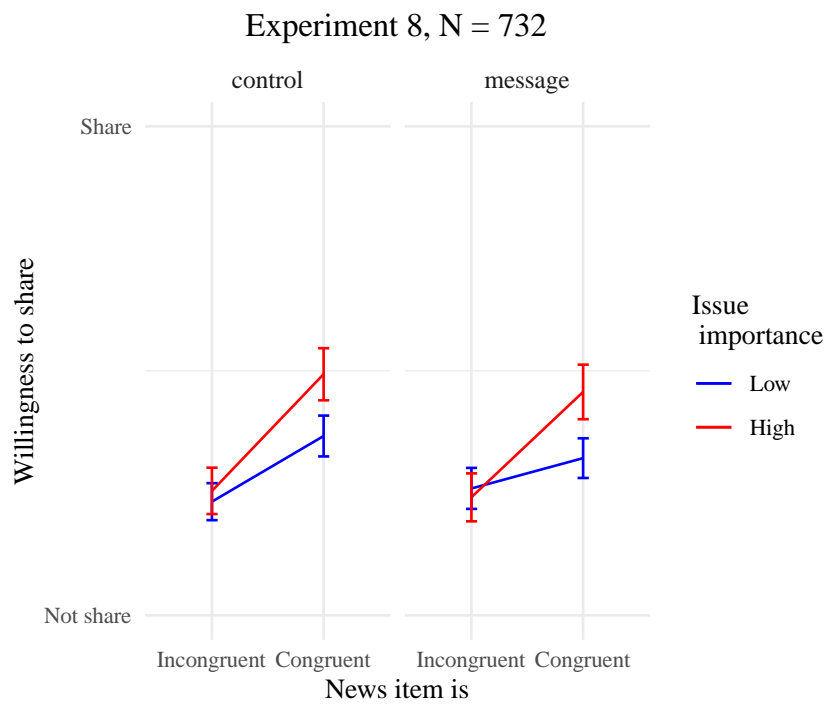
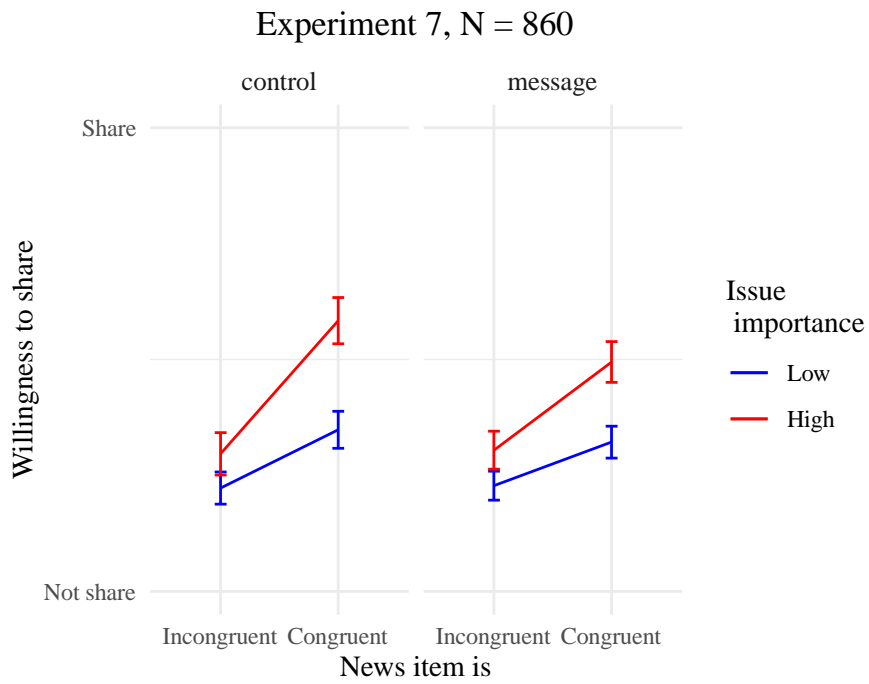
## Results

306 participants were removed from the data in the intervention experiment on real news items (Experiment 7) because they either passed the questionnaire twice (28), failed the attention check (173), had already passed some of our studies (100), or failed the English proficiency check (5). 860 participants remained in the dataset ( $M_{\text{age}} = 39$ ,  $SD_{\text{age}} = 12$ , 46% women). 331 participants were removed from the data in the intervention experiment on false news items (Experiment 8) because they either didn't complete the task (96), passed the experiment twice (42), failed the attention filter (59), had already passed some of our studies

(104), or failed the English proficiency test (30). 662 participants remained in the dataset ( $M_{age} = 38$ ,  $SD_{age} = 12$ , 42% women).

Replicating prior findings, participants were more likely to share politically congruent than incongruent headlines (main effect of news congruence: real news, Experiment 7:  $\beta = 0.31$ , [0.27,0.34],  $p < 0.001$ , false news; Experiment 8:  $\beta = 0.32$ , [0.28,0.37],  $p < 0.001$ ). They also reported greater willingness to share headlines on high importance than low importance topics (main effect of issue importance: real news:  $\beta = 0.27$ , [0.22,0.32],  $p < 0.001$ ; false news:  $\beta = 0.13$ , [0.08,0.18],  $p < 0.001$ ). The sharing preference for congruent news was strongest on high importance issues (interaction between news congruence and issue importance: real news:  $\beta = 0.24$ , [0.15,0.32],  $p < 0.001$ ; false news:  $\beta = 0.25$ , [0.15,0.34],  $p < 0.001$ ). Willingness to share in the intervention message condition did not significantly differ from sharing in the control condition (real new:  $\beta = -0.05$ , [-0.14,0.03],  $p = 0.19$ ; false news:  $\beta = -0.03$ , [-0.12,0.07],  $p = 0.58$ ).

The effect of the experimental manipulation was examined based on separate mixed-effects models for incongruent and congruent headlines (see ESM). The intervention message, to our surprise, did not affect sharing decisions very much. It did not increase sharing of true, incongruent headlines (main effect of condition:  $\beta = 0.00$ , [-0.10,0.09],  $p = 0.94$ ). The message did not increase sharing of false, incongruent headlines either (main effect of condition:  $\beta = 0.01$ , [-0.10,0.12],  $p = 0.85$ ). Regarding politically congruent items, the intervention message had a small moderating effect on willingness to share true, congruent news items (main effect of condition:  $\beta = -0.10$ , [-0.19,-0.01],  $p < 0.03$ ), an effect that seemed to be mostly occurring on high importance topics (interaction between condition and issue importance:  $\beta = -0.12$ , [-0.25,0.01],  $p = 0.07$ ). The intervention did not reduce sharing of false, congruent headlines, however (main effect of condition:  $\beta = -0.06$ , [-0.17,0.04],  $p = 0.26$ ).



**Figure 5:** Willingness to share real (Experiment 7) and false (Experiment 8) headlines as a function of news item congruence, issue importance, and experimental condition. Means are surrounded by 95% CIs.

## General Discussion

Across 8 experiments, we explored US participants' decisions to share true and false political headlines on 4 controversial issues—gun control, abortion, racial equality and sex equality—in a simulated social media context. With a novel design, our studies reproduced earlier findings that participants consistently favor politically congenial news over uncongenial ones when deciding what to share (Ekstrom & Lai, 2020; Faulmüller et al., 2012; Bakshy et al., 2015; Barberà et al., 2015; Liang, 2018; Arendt et al., 2016; Pogorelskiy & Shum, 2019; Shin & Thorson, 2017), and showed, for the first time, that this preference is magnified by participants' degree of moral conviction on the issue at stake.

The sharing preference for congenial content and its interaction with issue importance was robust, as it generalized across real and false news items alike. Contrary to the expectation that increasing the saliency of reputational concerns might moderate willingness to share politically congruent content, sharing patterns were not significantly affected by whether the audience was politically uncongenial (as opposed to congenial), nor by whether sharing was done from one's personal social media account (as opposed to an anonymous account). We also tested an intervention message spotlighting our disposition to favor information that fits one's goals and values, and to disregard information that doesn't, but the message only had the small effect of reducing willingness to share true, politically congruent headlines on high importance topics, and it did not affect sharing of false news stories.

How should one normatively regard this inclination to selectively communicate politically congenial content? Here, it is important to distinguish what is rational from the individual from what is in society's best interest. From a Bayesian standpoint, a rational individual *should* modulate the credibility she ascribes to a new piece of information based on its degree of consistency with her prior beliefs about the world (Tappin, Pennycook, Rand,

2020). To the extent, then, that politically congruent headlines fit people's priors more—which was the standard by which they were selected—and that sharing is partly motivated by accuracy—which our data, consistent with earlier work (e.g. Altay, de Araujo, Mercier, 2020), suggests is the case—it is natural that individual should privilege politically congruent content when deciding what to share.

Moreover, passing along politically congruent information at a higher rate than cross-cutting information is especially what the political rationality of coalitional management should incline us to do—in particular when the audience is one's political friends, who count as allies with whom bonds must be maintained (Osmundsen et al., 2020; Pennycook et al., 2020; Tooby & Cosmides, 2010; Tooby, 2017). In a coalitional perspective, many politically congruent headlines are relevant because they are about threatening events and persons. In circumstances of perceived group conflict, they constitute socially strategic information that can be used both to mobilize one's group to prepare actions, and to signal that one is a trustworthy coalition member. Other headlines are politically congruent because they highlight good news for one's coalition. The coalitional rationale for sharing them is that they may be a way of persuading new members to join in (proselytism), of motivating those already involved, and of showing one's devotion to the cause. Moreover, our finding that preference for congruent content increases on high importance issues suggests that the weight of coalitional motivations in people's communication may be all the more pronounced as they moralize the issue (Skitka et al., 2005; Skitka, 2010; Van Bavel & Pereira, 2018).

Selective communication on politics may be individually rational, but it is clearly harmful for society at large. Liberal and conservative partisans have predictably different beliefs about all sorts of facts, and they tend to spend most of their time with people whose views are similar to their own and who rarely challenge their representations. But if there were more convergence in citizens' beliefs on burning issues—for instance, on factual estimates of the impact on crime of the right to carry firearms in the US—it would become easier to

implement laws that are objectively better able to benefit the majority—laws banning weapons in the case where they increase the number of shootings, or laws allowing them if their propagation has no demonstrable influence on crime.

Moreover, to the extent that selective communication makes it more difficult to share the same factual reality across partisan lines, it may foster affective polarization, by making partisans on one side appear so ill-informed to the other side that their ignorance must be attributed to stupidity, insincerity, or malevolence (Iyengar et al., 2018; Levendusky, 2013; Rogowski & Sutherland, 2015). This will not help negotiation of mutually-beneficial outcomes in the policy domain.

News producers may play an important role in sustaining the status quo. Having some meta-knowledge of audiences' preconceived notions and political interests is part and parcel of both legitimate journalists and fake news manufacturers' training. As long as eyeballs, clicks and shares translate into revenue and professional success, news producers are incentivized to select and format the information they publish so as to make it fit their audience's priors and political goals, in the hope that it is virally spread in the targeted communities.

Our study designs share two limitations with most studies on news sharing online. First, the absence of a comment and share feature likely disincentivized sharing with the intention to express one's disapprobation or outrage at cross-cutting headlines, because people tend to believe their audiences will interpret sharing as some form of endorsement. How including such a feature would affect sharing decisions should be examined in future work. Second, news sharing was measured using willingness to share rather than real life sharing decisions, and on fictitious social media accounts, not real ones. Although this reduces the external validity of our findings, real-life communication may be even more selective than our online simulations based on self-report (Pronin, 2007). Besides, recent evidence from Mosleh, Pennycook & Rand (2020) suggests that self-reported willingness to share political news articles in online surveys correlates with actual sharing on Twitter ( $r = .44$ ), and that the perceived interestingness of a



piece of news in online experiments predicts its success on actual social media (Bakshy et al. 2011).

### **Authors' Contributions**

All authors contributed to the studies' design. Testing and data collection were performed by A. Marie and S. Altay under the supervision of B. Strickland, who funded the research. A. Marie performed the statistical analyses, reported the results and drafted the manuscript. S. Altay and B. Strickland provided revisions.

### **Conflict of interest**

The authors declare having no conflict of interest.

### **References**

References for this paper are reported in the general references section at the end of the dissertation.

## **Electronic Supplementary Materials (ESM)**

**All Data, scripts, and materials will be published on the Open Science Framework**

## Coding of predictors in regression tables

In regression analyses, predictor codes mean the following:

### *Experiments 1-8*

CongruentTRUE = New item is politically congruent (vs. incongruent) to the participant's attitude on the issue

CongruentNA = New item is neutral, non political (vs. incongruent)

AbsImpIssueTRUE = The issue covers by the news item is of high moral importance (vs. low importance)

### *Experiment 5*

ConditionDisagree = Politically uncongenial audience condition (vs. politically congenial)

### *Experiment 6*

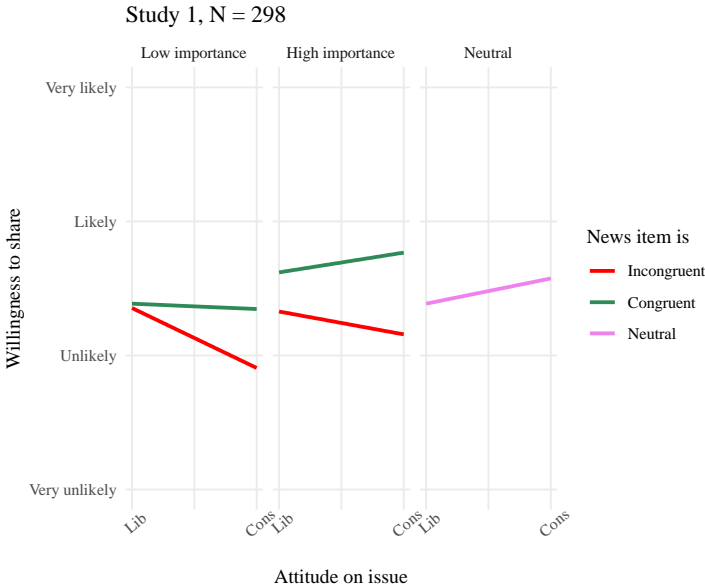
ConditionPersonal = Sharing from personal account condition (vs. from anonymous account)

### *Experiments 7-8*

ConditionMessage = Intervention message condition (vs. control, no message)

# Experiment 1: Willingness to share true headlines, 4-point DV

The following plot shows willingness to share as a function of participants' positions on the liberal-conservative attitude axis on each issue, news item's ideological congruence, and issue importance:



## Predictors of willingness to share

First model shows main effects only, second model includes the interaction:

Predictors	Sharing			Sharing		
	$\beta$	95% CI	<i>p</i>	$\beta$	95% CI	<i>p</i>
CongruentTRUE	0.26	0.19,0.32	0.00	0.19	0.11,0.27	0.00
CongruentNA	0.32	0.25,0.38	0.00	0.28	0.21,0.35	0.00
AbsImpIssueTRUE	0.17	0.10,0.24	0.00	0.07	-0.03,0.17	0.16

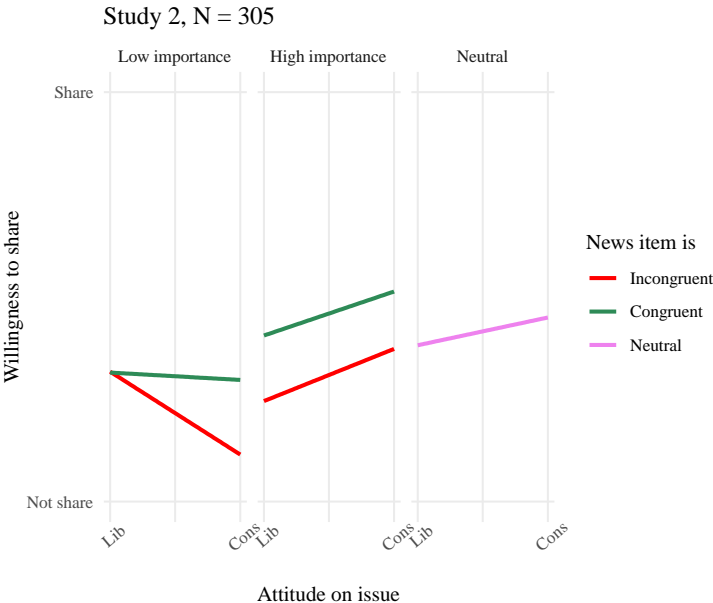
CongruentTRUE:AbsImpIssueTRUE	0.20	0.07,0.33	0.00
-------------------------------	------	-----------	------

---

Marginal R <sup>2</sup> / Conditional R <sup>2</sup>	0.019 / 0.409	0.020 / 0.410	
--	---------------	---------------	--

## Experiment 2: Willingness to share true headlines, dichotomous DV

The following plot shows willingness to share as a function of participants' positions on the liberal-conservative attitude axis on each issue, news item's ideological congruence, and issue importance:



### Predictors of willingness to share

First model shows main effects only (news item congruence and issue importance), second model includes the interaction

Predictors	Sharing			Sharing		
	$\beta$	95% CI	<i>p</i>	$\beta$	95% CI	<i>p</i>
CongruentTRUE	0.21	0.14,0.28	0.00	0.15	0.06,0.23	0.00
CongruentNA	0.40	0.32,0.47	0.00	0.36	0.29,0.44	0.00
AbsImpIssueTRUE	0.18	0.10,0.25	0.00	0.09	-0.01,0.19	0.09

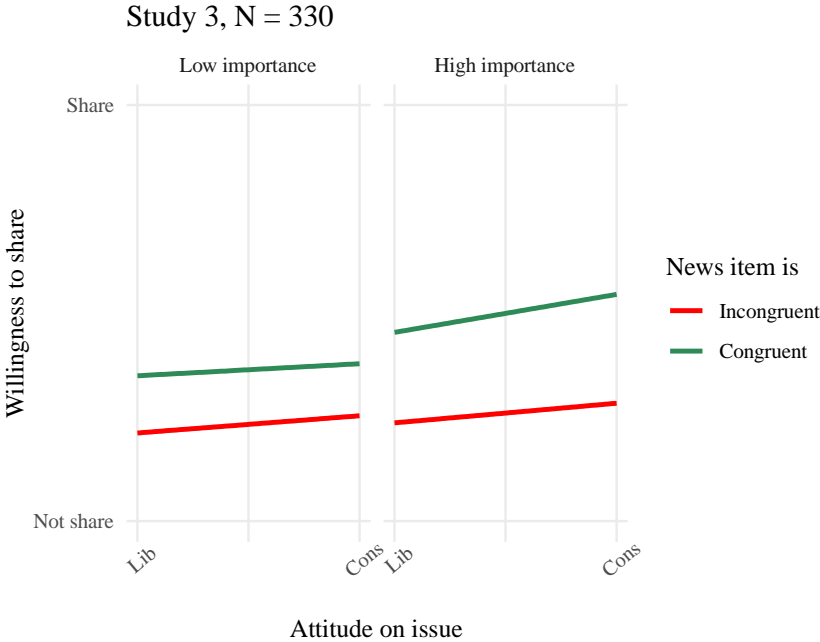
CongruentTRUE:AbsImpIssueTRUE	0.18	0.04,0.32	0.01
-------------------------------	------	-----------	------

---

Marginal R <sup>2</sup> / Conditional R <sup>2</sup>	0.024 / 0.267	0.025 / 0.268	
--	---------------	---------------	--

### Experiment 3: Willingness to share false headlines, dichotomous DV

The following plot shows willingness to share as a function of participants' positions on the liberal-conservative attitude axis on each issue, news item's ideological congruence, and issue importance:



#### Predictors of willingness to share

First model shows main effects only (news item congruence and issue importance), second model includes the interaction

Predictors	Sharing			Sharing		
	$\beta$	95% CI	<i>p</i>	$\beta$	95% CI	<i>p</i>
CongruentTRUE	0.36	0.29,0.42	0.00	0.29	0.21,0.36	0.00



AbsImpIssueTRUE	0.14	0.07,0.22	0.00	0.04	-0.05,0.14	0.38
CongruentTRUE:AbsImpIssueTRUE				0.20	0.07,0.33	0.00
<hr/>						
Marginal R <sup>2</sup> / Conditional R <sup>2</sup>	0.036 / 0.337			0.039 / 0.339		

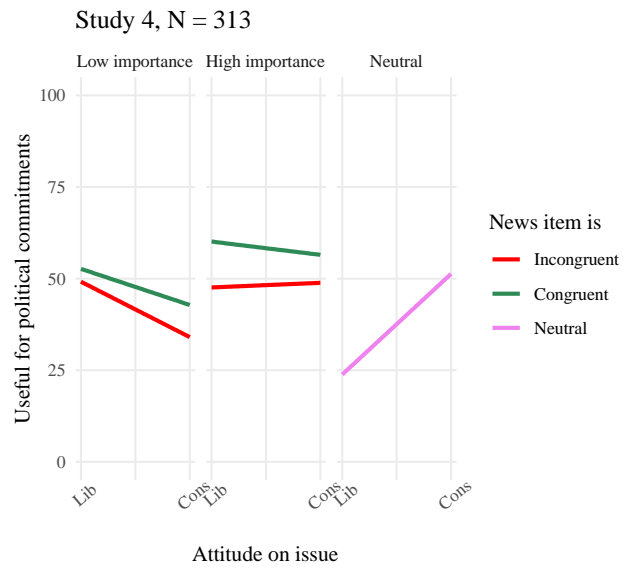
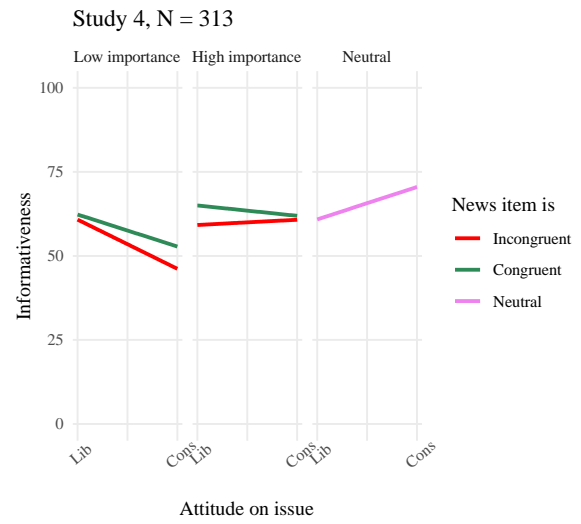
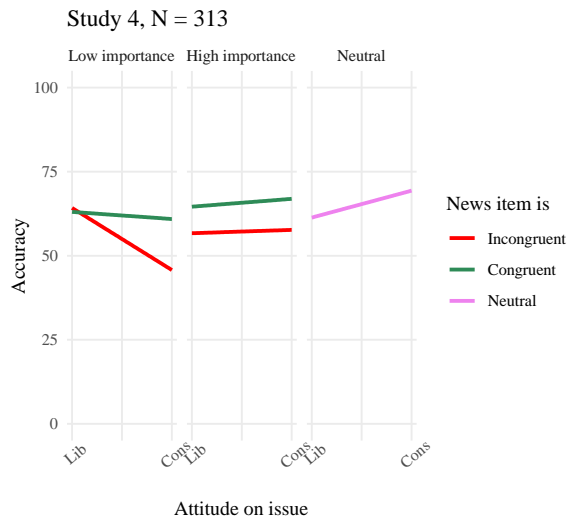
## Experiment 4: Motivations for sharing, 4-point DV

### Predictors of willingness to share

First model shows main effects only (news item congruence and issue importance), second model includes the interaction

<i>Predictors</i>	<b>Sharing</b>			<b>Sharing</b>		
	$\beta$	95% CI	<i>p</i>	$\beta$	95% CI	<i>p</i>
CongruentTRUE	0.19	0.11,0.27	0.00	0.14	0.03,0.24	0.01
CongruentNA	0.25	0.16,0.34	0.00	0.22	0.13,0.32	0.00
AbsImpIssueTRUE	0.18	0.08,0.27	0.00	0.11	-0.01,0.24	0.08
CongruentTRUE:AbsImpIssueTRUE				0.13	-0.03,0.29	0.12
Marginal R <sup>2</sup> / Conditional R <sup>2</sup>	0.012 / 0.480			0.013 / 0.481		

The following plot shows participants' reported motivations to share (Accuracy, Informativeness, Usefulness for political commitments) as a function of news item's congruence and issue importance:



**Willingness to share as a function of each motivation to share.**

First model is politically incongruent news only, second model politically congruent news, and third model neutral news.

<i>Predictors</i>	<b>Sharing (Incongruent)</b>			<b>Sharing (Congruent)</b>			<b>Sharing (Neutral)</b>		
	$\beta$	95% CI	<i>p</i>	$\beta$	95% CI	<i>p</i>	$\beta$	95% CI	<i>p</i>
Accuracy	0.14	0.07,0.21	0.00	0.17	0.10,0.24	0.00	0.19	0.12,0.26	0.00
Informative	0.24	0.17,0.32	0.00	0.22	0.14,0.29	0.00	0.24	0.16,0.31	0.00
Useful_commit	0.42	0.34,0.49	0.00	0.40	0.32,0.48	0.00	0.35	0.27,0.43	0.00
Time	0.01	-0.04,0.06	0.77	-0.00	-0.06,0.05	0.96	0.04	-0.01,0.10	0.14
Attit_Issue	-0.07	-0.13,-0.02	0.01	0.00	-0.05,0.06	0.90	-0.00	-0.08,0.07	0.91
Marginal R <sup>2</sup> / Conditional R <sup>2</sup>	0.531 / 0.655			0.472 / 0.615			0.375 / 0.617		

## Experiment 5: Sharing to politically congenial vs. uncongenial audience, dichotomous DV

### Predictors of willingness to share

First model shows main effects only (news item congruence, issue importance, experimental condition), second model includes the two-way interactions:

<i>Predictors</i>	<b>Sharing</b>			<b>Sharing</b>		
	<i>β</i>	<i>95% CI</i>	<i>p</i>	<i>β</i>	<i>95% CI</i>	<i>p</i>
ConditionDisagree	0.06	-0.09,0.22	0.44	0.18	0.00,0.37	0.05
CongruentTRUE	0.26	0.19,0.34	0.00	0.29	0.18,0.41	0.00
AbsImpIssueTRUE	0.19	0.10,0.27	0.00	0.18	0.04,0.32	0.01
AbsImpIssueTRUE:ConditionDisagree				-0.12	-0.28,0.05	0.17
CongruentTRUE:ConditionDisagree				-0.16	-0.30,-0.01	0.03
CongruentTRUE:AbsImpIssueTRUE				0.12	-0.02,0.27	0.10
Marginal R <sup>2</sup> / Conditional R <sup>2</sup>	0.026 / 0.333			0.030 / 0.336		

### Effect of condition on sharing of each type of headline

#### Incongruent headlines

	<b>Sharing</b>	<b>Sharing</b>
--	----------------	----------------

<i>Predictors</i>	$\beta$	95% CI	<i>p</i>	$\beta$	95% CI	<i>p</i>
ConditionDisagree	0.15	-0.03,0.32	0.10	0.20	0.01,0.40	0.04
AbsImpIssueTRUE	0.13	0.01,0.24	0.03	0.20	0.04,0.36	0.02
ConditionDisagree:AbsImpIssueTRUE				-0.15	-0.38,0.08	0.21
Marginal R <sup>2</sup> / Conditional R <sup>2</sup>	0.009 / 0.339			0.010 / 0.341		

### Congruent headlines

<i>Predictors</i>	Sharing			Sharing		
	$\beta$	95% CI	<i>p</i>	$\beta$	95% CI	<i>p</i>
ConditionDisagree	-0.02	-0.20,0.16	0.86	0.02	-0.18,0.21	0.88
AbsImpIssueTRUE	0.25	0.14,0.36	0.00	0.29	0.13,0.45	0.00
ConditionDisagree:AbsImpIssueTRUE				-0.08	-0.31,0.14	0.46
Marginal R <sup>2</sup> / Conditional R <sup>2</sup>	0.015 / 0.385			0.015 / 0.385		

## Experiment 6: Willingness to pay for sharing from anonymous vs. personal account, 0-100 DV

### Predictors of willingness to pay for sharing

First model shows main effects only (news item congruence, issue importance, experimental condition), second model the two-way interactions:

<i>Predictors</i>	<b>Sharing</b>			<b>Sharing</b>		
	$\beta$	95% CI	<i>p</i>	$\beta$	95% CI	<i>p</i>
ConditionPersonal	-0.03	-0.19,0.13	0.68	0.02	-0.15,0.20	0.79
CongruentTRUE	0.34	0.28,0.40	0.00	0.26	0.17,0.35	0.00
AbsImpIssueTRUE	0.20	0.13,0.27	0.00	0.05	-0.06,0.16	0.41
AbsImpIssueTRUE:ConditionPersonal				-0.04	-0.17,0.10	0.59
CongruentTRUE:ConditionPersonal				-0.09	-0.20,0.03	0.14
CongruentTRUE:AbsImpIssueTRUE				0.34	0.22,0.46	0.00
Marginal R <sup>2</sup> / Conditional R <sup>2</sup>	0.038 / 0.498			0.045 / 0.506		

**Willingness to pay for sharing incongruent and congruent headlines as a function of experimental condition and issue importance**

#### Incongruent headlines

<i>Predictors</i>	<b>Sharing</b>			<b>Sharing</b>		
	$\beta$	95% CI	<i>p</i>	$\beta$	95% CI	<i>p</i>
ConditionPersonal	0.02	-0.16,0.21	0.80	0.05	-0.14,0.25	0.59
AbsImpIssueTRUE	0.03	-0.06,0.12	0.48	0.07	-0.06,0.20	0.27
ConditionPersonal:AbsImpIssueTRUE				-0.08	-0.26,0.10	0.40
Marginal R <sup>2</sup> / Conditional R <sup>2</sup>	0.000 / 0.555			0.001 / 0.555		

### **Congruent headlines**

<i>Predictors</i>	<b>Sharing</b>			<b>Sharing</b>		
	$\beta$	95% CI	<i>p</i>	$\beta$	95% CI	<i>p</i>
ConditionPersonal	-0.08	-0.25,0.09	0.37	-0.07	-0.26,0.12	0.45
AbsImpIssueTRUE	0.37	0.28,0.47	0.00	0.38	0.25,0.51	0.00
ConditionPersonal:AbsImpIssueTRUE				-0.02	-0.21,0.17	0.84
Marginal R <sup>2</sup> / Conditional R <sup>2</sup>	0.034 / 0.502			0.034 / 0.501		



## Experiment 7: Intervention message on real news items

### Predictors of willingness to share

First model shows main effects only (news item congruence, issue importance, experimental condition), second model includes the two-way interactions:

<i>Predictors</i>	Sharing			Sharing		
	$\beta$	95% CI	<i>p</i>	$\beta$	95% CI	<i>p</i>
ConditionMessage	-0.05	-0.14,0.03	0.19	0.02	-0.08,0.11	0.76
CongruentTRUE	0.31	0.27,0.34	0.00	0.28	0.21,0.34	0.00
AbsImpIssueTRUE	0.27	0.22,0.32	0.00	0.18	0.10,0.26	0.00
AbsImpIssueTRUE:ConditionMessage				-0.05	-0.15,0.04	0.26
CongruentTRUE:ConditionMessage				-0.10	-0.18,-0.02	0.01
CongruentTRUE:AbsImpIssueTRUE				0.24	0.15,0.32	0.00
Marginal R <sup>2</sup> / Conditional R <sup>2</sup>	0.041 / 0.325			0.045 / 0.329		

### Willingness to share incongruent and congruent headlines as a function of experimental condition and issue importance

#### Incongruent headlines

<i>Predictors</i>	Sharing			Sharing		
	$\beta$	95% CI	<i>p</i>	$\beta$	95% CI	<i>p</i>
ConditionMessage	-0.00	-0.10,0.09	0.94	-0.01	-0.11,0.10	0.90

AbsImpIssueTRUE	0.16	0.09,0.22	0.00	0.15	0.06,0.25	0.00
ConditionMessage:AbsImpIssueTRUE				0.01	-0.12,0.13	0.90
<hr/>						
Marginal R <sup>2</sup> / Conditional R <sup>2</sup>	0.006 / 0.373			0.006 / 0.373		

**Congruent headlines**

<i>Predictors</i>	<b>Sharing</b>			<b>Sharing</b>		
	<i>β</i>	<i>95% CI</i>	<i>p</i>	<i>β</i>	<i>95% CI</i>	<i>p</i>
ConditionMessage	-0.10	-0.19,-0.01	0.03	-0.06	-0.16,0.04	0.26
AbsImpIssueTRUE	0.38	0.32,0.45	0.00	0.45	0.36,0.55	0.00
ConditionMessage:AbsImpIssueTRUE				-0.12	-0.25,0.01	0.07
<hr/>						
Marginal R <sup>2</sup> / Conditional R <sup>2</sup>	0.036 / 0.331			0.037 / 0.332		

## Experiment 8: Intervention message on false news items

### Predictors of willingness to share

First model shows main effects only (news item congruence, issue importance, experimental condition), second model includes the two-way interactions:

<i>Predictors</i>	Sharing			Sharing		
	$\beta$	95% CI	<i>p</i>	$\beta$	95% CI	<i>p</i>
ConditionMessage	-0.03	-0.12,0.07	0.58	0.02	-0.09,0.13	0.71
CongruentTRUE	0.32	0.28,0.37	0.00	0.27	0.20,0.34	0.00
AbsImpIssueTRUE	0.13	0.08,0.18	0.00	0.02	-0.07,0.11	0.64
AbsImpIssueTRUE:ConditionMessage				-0.03	-0.14,0.08	0.56
CongruentTRUE:ConditionMessage				-0.07	-0.16,0.02	0.11
CongruentTRUE:AbsImpIssueTRUE				0.25	0.15,0.34	0.00
Marginal R <sup>2</sup> / Conditional R <sup>2</sup>	0.030 / 0.318			0.034 / 0.322		

### Willingness to share incongruent and congruent headlines as a function of experimental condition and issue importance

#### Incongruent headlines

<i>Predictors</i>	Sharing			Sharing		
	$\beta$	95% CI	<i>p</i>	$\beta$	95% CI	<i>p</i>
ConditionMessage	0.01	-0.10,0.12	0.85	0.04	-0.08,0.17	0.48

AbsImpIssueTRUE	-0.01	-0.08,0.07	0.85	0.04	-0.06,0.14	0.46
ConditionMessage:AbsImpIssueTRUE				-0.10	-0.24,0.05	0.20
<hr/>						
Marginal R <sup>2</sup> / Conditional R <sup>2</sup>	0.000 / 0.370			0.001 / 0.372		

**Congruent headlines**

<i>Predictors</i>	<b>Sharing</b>			<b>Sharing</b>		
	<i>β</i>	<i>95% CI</i>	<i>p</i>	<i>β</i>	<i>95% CI</i>	<i>p</i>
ConditionMessage	-0.06	-0.17,0.04	0.26	-0.08	-0.19,0.04	0.21
AbsImpIssueTRUE	0.26	0.19,0.34	0.00	0.24	0.14,0.35	0.00
ConditionMessage:AbsImpIssueTRUE				0.05	-0.11,0.20	0.56
<hr/>						
Marginal R <sup>2</sup> / Conditional R <sup>2</sup>	0.017 / 0.292			0.017 / 0.292		

## **Choice of the 4 controversial issues**

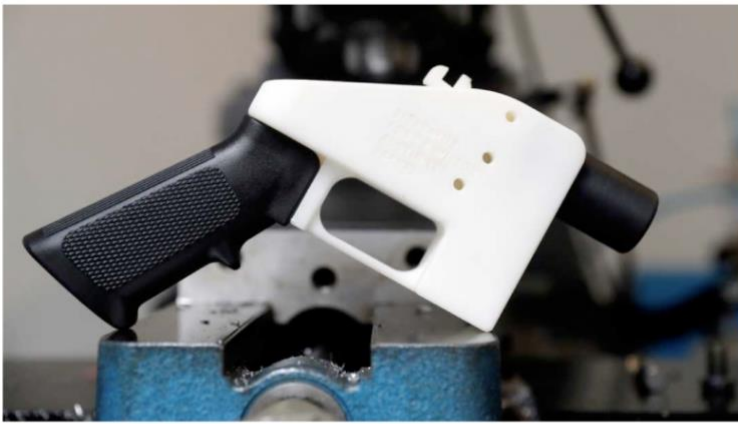
The 4 controversial issues retained for the political news items were chosen based on responses to the questions “*I could never vote for a politician if they supported ...*” and “*I could never vote for a politician if they failed to support...*” in a pilot study. These questions enabled us to know which topics U.S. MTurkers judged as being most morally important, and on which they were most likely to be polarized. Gun Control, Racial Equality, Gender Equality and Abortion were the topics that came up most frequently and were retained in all the studies.

## Real political news items

Above each headline is provided its label in the R scripts. In the label “GunCo.D”, “GunCo” specifies the issue, and “D” that it was congruent for liberals-Democrats (in contrast, “R” means that an item was congruent for conservatives-Republicans).

### Gun Control

Congruent for Liberals (GunCo.D):



The fast development of 3D gun printing may pose a major threat to public safety.

The distribution of instructions for producing plastic weapons has been authorized by a Department of Justice allowing...

Congruent for Conservatives (GunCo.R):



Guns are being used for self-defense as often as 10'000 times a year, the National Crime Victimization Survey says

Statistical evidence suggests that the possession of guns allows many by-standers to protect innocents against brutal attacks every year...

## Racial Equality

Congruent for Liberals (Race.D) :



The racist myth of the 'physical' African football team

Racist tropes continue to dominate white commentary of global football events.

Congruent for Conservatives (Race.R):



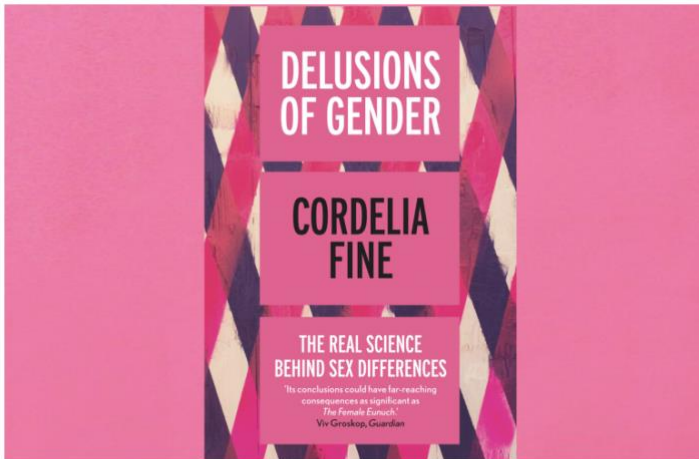
Recent genetic studies have shown genetic differences across populations

Researcher found differences across populations not just in the genetic determinants of simple traits such as skin color, but also in more complex traits like bodily dimensions and...

## **Gender Equality**

Congruent for Liberals (Gender.D)





In her new book, Cordelia Fine defends the idea that male and female brains are not significantly different.

The author debunks the idea that men and women are hardwired for different interests and cognitive abilities. Her critical review of the scientific literature...

Congruent for Conservatives (Gender.R)



A recent scientific study finds there are innate cognitive differences between men and women

According to new evidence published in a major scientific journal, sex differences in cognitive abilities may have been underestimated for many years due to...

## Abortion

Congruent for Liberals (Abor.D)



## Abortion will soon become cost-free in Ireland

Until recently, abortion was illegal in Ireland. However, according to Health Minister Simon Harris, a new law is about to be submitted to the legislature in October proposing to make abortion cost-free...

## Congruent for Conservatives (Abor.R)



## New evidence suggests a majority of abortion cases are on-demand and do not involve rape victims

One of the most common reasons given to justify abortion is that it may allow mothers to postpone childbearing to a more suitable time and provide already born children with more care and resources...

## Raters' perceptions of real political news items' slant

News items raters' political orientation was assessed using the following question:

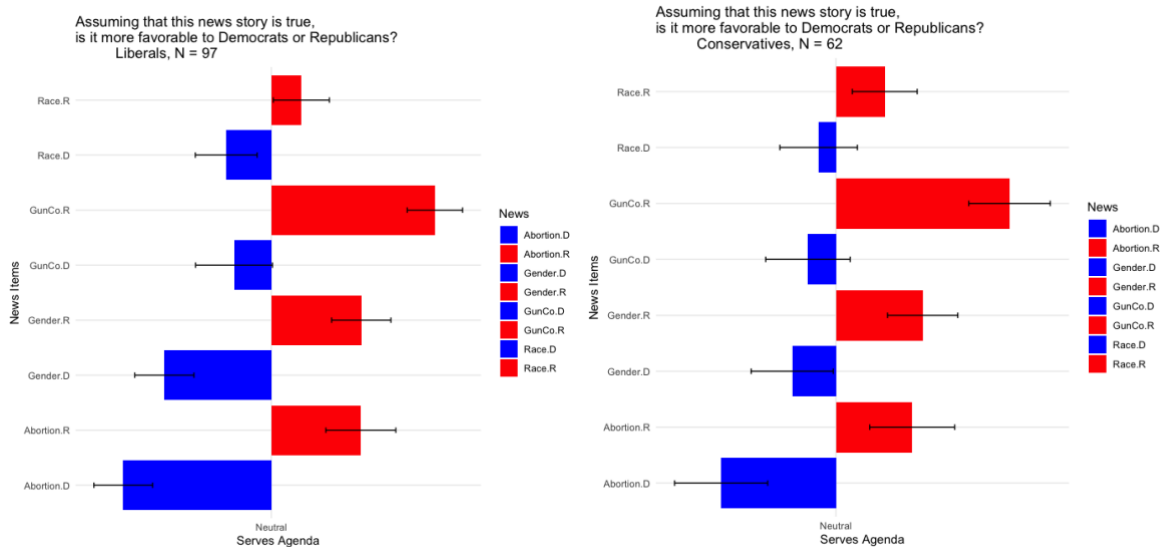
What is your political orientation?

Strongly Democrat      Strongly Republican

0 10 20 30 40 50 60 70 80 90 100



The following graphs display perceptions of true news items' political slant according to MTurk raters identifying as Democrat-Liberal (< 50) and as Republican-conservative (>= 50) to the above political orientation question:



## Real non-political news items

Neutral.1:



**New archeological evidence suggests prehistoric children as young as eight worked as brickmakers and miners**

A surge of interest in the archaeology of childhood is revealing details of the skilled and sometimes back-breaking work that youngsters performed hundreds to thousands of years ago.

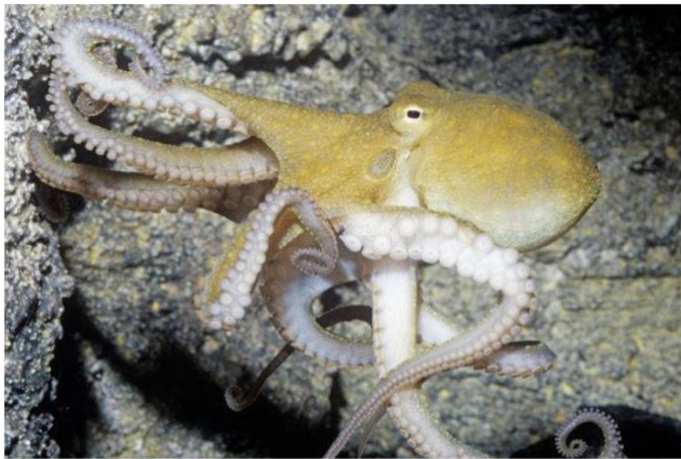
Neutral.2:



When this sea cucumber (*Pannychia moseleyi*) is stimulated, it produces a frenzy of bioluminescent light

Technological advances in video cameras and low-light sensors allowed scientists to reveal the existence of a worm-like predator living in the deep sea shooting off rings of blue light.

Neutral.3:



On Ecstasy, Octopuses Reached Out for a Hug

By administering MDMA to these sea creatures, researchers found they share parts of an ancient messaging system involved in social behaviors with humans.

Neutral.4:



## Science Is Showing How Gut Bacteria Affect The Brain, But the Efficacy of Probiotics Hasn't Been Shown Yet

Despite evidence that bacteria in our digestive systems affect our brains, we don't yet know if probiotic supplements can improve our mental health by influencing gut bacteria...

## False political news items

### Gun control

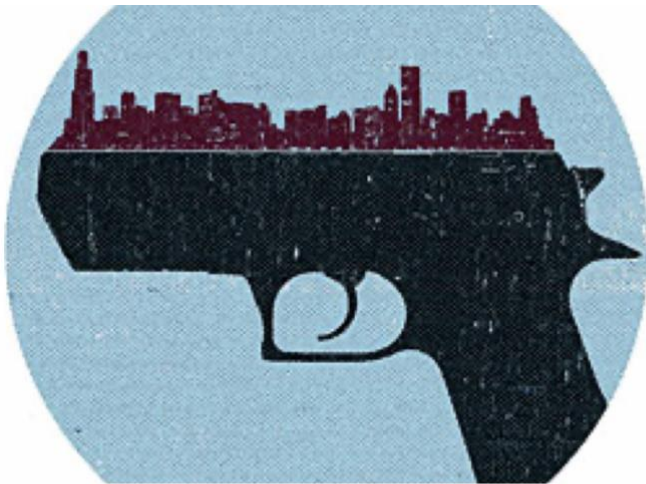
Congruent for Liberals (GunCo D\_fake/ GunCo D\_Mental):



## Republicans voted to allow people with pre-existing mental illness to buy firearms

Republican party Representatives voted to allow people with pre-existing mental illness to buy firearms, yet denied them health insurance to treat their pre-existing mental illness...

Congruent for Conservatives (GunCo.R\_fake/ GunCo.R\_Chicago):



### Chicago Shootings Prove That Gun Control Doesn't Reduce Crime

Chicago has the strongest gun laws in the U.S., and yet the total number of murders per inhabitant in Chicago is the highest in the country...

## Racial equality

Congruent for Liberals (Race.D\_FreeTuition):



### Kentucky University will offer 30% discount on tuition for Black students as an apology for slavery

Western Kentucky University's student government wants their school to pay 'reparations' to black students by...

Congruent for Conservatives (Race.R\_Mohammed):



'Mohammed' is now the most popular name for newborn boys in France, Germany, and the Netherlands because of immigration

Due to the growing number of immigrants in European countries, 'Mohammed' has been recorded as the most frequently given name, last year, to male newborns in several major European countries....

## Gender equality

Congruent for Liberals (Gender.D\_PayGap):



Women still make 60% of what men earn for same job and working time

According to the *controlled* gender pay gap — which controls for job title, years of experience, industry and location so that the only differentiation between workers is their gender — women now make \$0.60 for every dollar an equivalent man makes.



Congruent for Conservatives (Gender.R\_Hormone):



**Hundreds of deaths linked to hormonal therapy used as puberty blocker for gender-transitioning kids**

The use of drugs that block normal sexual development in people transitioning to the opposite sex has been associated with serious medical conditions, including cases of nervous system and psychiatric disorders...

## **Abortion**

Congruent for Liberals (Abor.D\_Pence):



Allowing abortion for rape victims will only incentivize women to report fictitious rapes, Pence said off the record

"The only thing allowing abortion for rapes can do is to motivate women to falsely report having been raped" said Vice President Pence in...

Congruent for Conservatives (Abor.R\_Bernie):

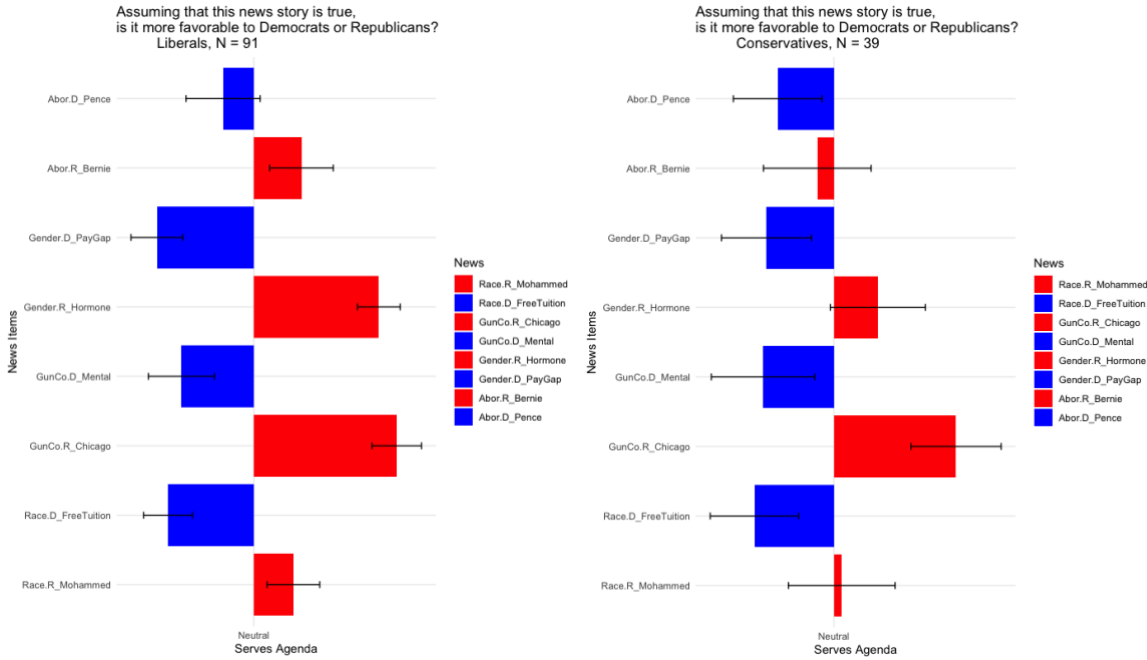


Bernie Sanders wants to set cut off date for abortions up to 7 months

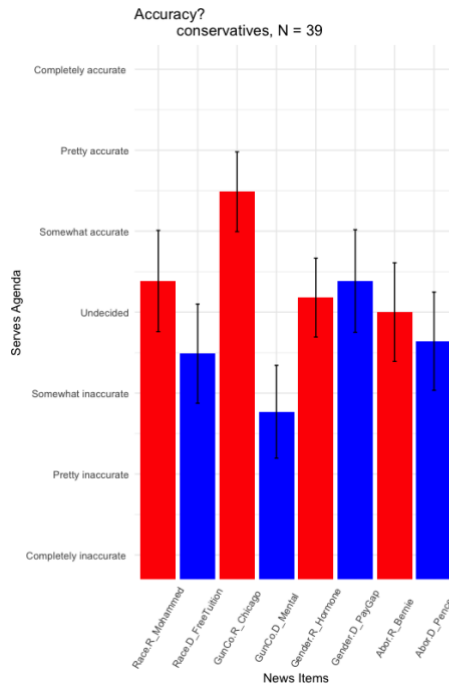
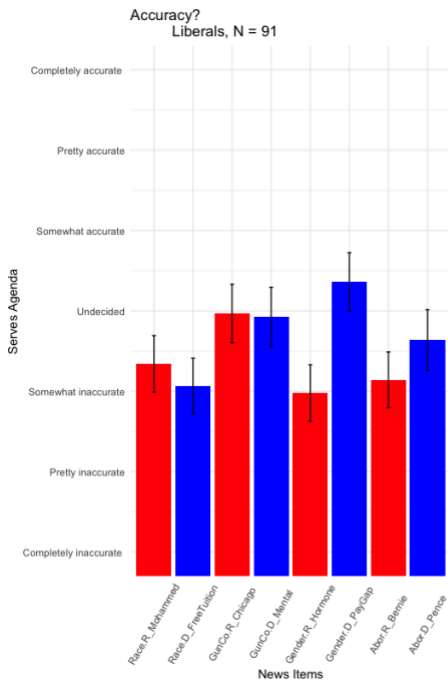
Says Bernie Sanders wants "to make the cut off date for abortions up to 28 weeks", arguing "a baby at that point feels no pain."

Raters' perceptions of false political news items' slant

False news items raters' political orientation was assessed using the same political orientation item as on true headlines. The following plot shows perceptions of false news items' political slant according to MTurk raters identifying as Democrat-Liberal (< 50) and as Republican-conservative (>= 50):



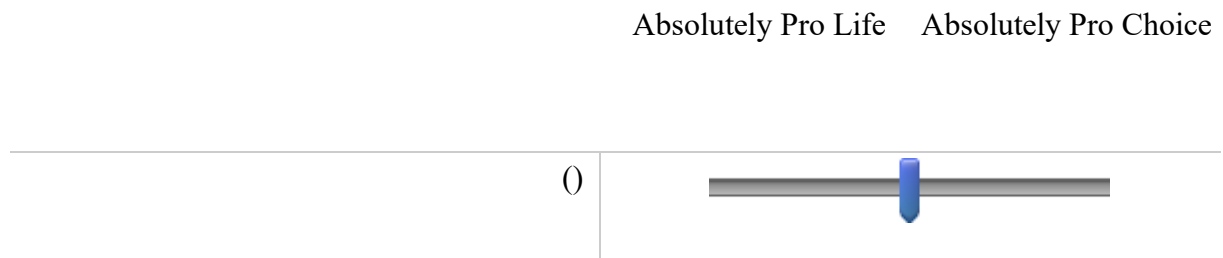
Perceptions of false news items' accuracy according to people identifying as Democrat-Liberal (< 50) and as Republican-conservative (>= 50):



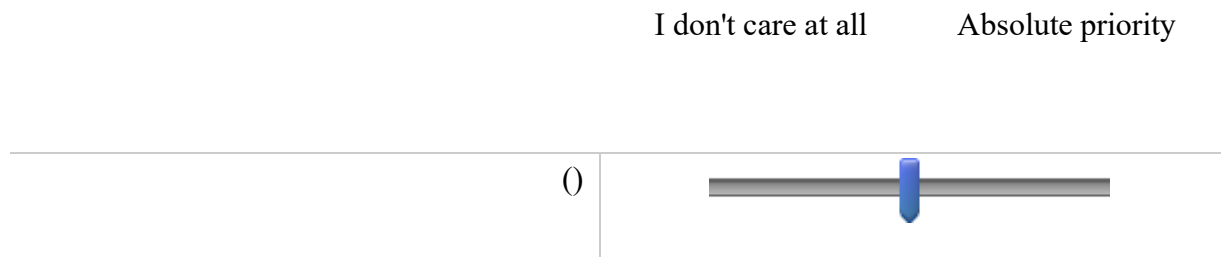
## Political attitude on issue (continuous)

These measures allowed us to determine which news story was congruent and which was incongruent for each participant in each political pair. They also allowed us to assess participants' attitudes on a liberal to conservative axis on each issue.

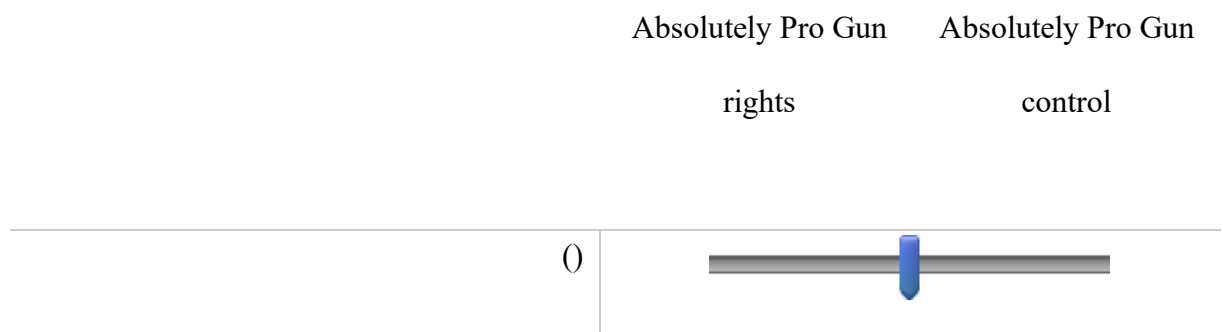
Abor\_Attit What is your position on the issue of **abortion**?



Gender\_Attit What is your position on the issue of **gender equality**?



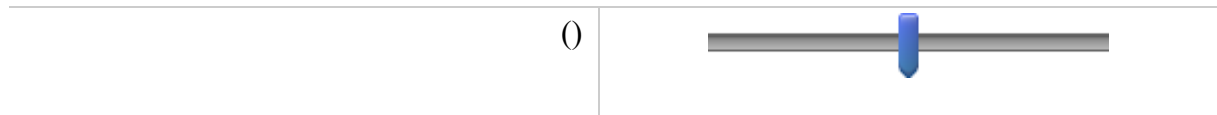
GunCo\_Attit What is your position on the issue of **guns**?



Race\_Attit What is your position on the issue of **racial equality**?

I don't care at all

Absolute priority



### Moral importance of issue (dichotomous)

Abs\_Imp Is one of the following issues of **absolute importance** to you? (several choices possible)

- Pro-life vs. Pro choice (1)
  - Racial equality (2)
  - Gender equality (3)
  - Gun control vs. Gun rights (4)
  - None of the above. Other? (5)
-

## Vignettes not provided in main text

### Experiment 6: Sharing from an anonymous vs. personal social media account

#### Anonymous account condition vignettes:

#### Page 1

*Please read very carefully the following instruction (1/2):*

Imagine you work for an association, called 'Public Support', that gives you a budget for promoting any social cause you care about through sharing news online. The budget the association gives you is limited, but allows you to have a social impact on potentially thousands of users through news sharing. To promote content you care about, you are using the **association's social media account, not your own**. This means that your promotion is done **anonymously** : no one will be able to match the content you publish to your name.



## Page 2

*Please read very carefully the following instruction (2/2):*

Currently preparing your anonymous promotion, you are about to be presented with a series of news items to potentially share online. For each of them, you are given an endowment of \$100, from which you must decide **how much money you are willing to spend in order to share the news (thereby promoting its content)**, or how much to **keep for yourself**.

For instance, if you decide to invest \$60 in order to promote a news item on a topic you care about, that means you can keep the remaining \$40 for yourself.

Note that the reliability of the news' sources is uncertain. The sources will not be made visible to you.





### Personal account condition vignettes:

#### Page 1

*Please read very carefully the following instruction (1/2):*

Imagine you work for an association that gives you a budget for promoting any social cause you care about through sharing news online. The budget the association gives you is **limited**, but allows you to have a social impact on potentially thousands of users through news sharing.

To promote content you care about, you are using **your own social media account** (but the money you are investing comes from the association, not your own pocket). This means that all your friends will be able to match the content you publish to your personal identity.

*Please read very carefully the following instruction (2/2):*

Currently preparing your nominative promotion work, you are about to be presented with a series of news items to potentially share online. For each of them, you are given an endowment of \$100, from which you must decide how much money you are willing to spend in order to share the news (thereby promoting its content), or how much to keep for yourself.

For instance, if you decide to invest \$60 in order to promote a news item on a topic you care about, that means you can keep the remaining \$40 for yourself.

Note that the reliability of the news' sources is uncertain. The sources will not be made visible to you.

### **Attention check ‘videogame’**

This attention check was applied to all datasets to delete inattentive participants.

Now imagine you are playing video games with a friend and at some point your friend says:

“I don’t want to play this game anymore! To make sure that you read the instructions, please write the three following words 'I pay attention' in the box below. I really dislike this game.”

Do you agree with your friend?

---

Participants who didn't write an expression that contained 'I pay attention' were deleted from the databases.

## Chapter 4: Intentions matter a lot, and efficiency little, in folk judgments of policy decisions

This section is based on the following manuscript:

“Intentions matter a lot, and efficiency little, in folk judgments of policy decisions”

2020 (not yet submitted)

Antoine Marie, Hugo Trad, Brent Strickland

### **ABSTRACT:**

Consequentialism contends that, when morally assessing policy decisions, the only thing that should matter is how successful they are at reaching their objective, relative to their cost. Yet, across 5 online studies (N=1515), we found that French participants regarded a policy driven by the altruistic intention to help a social issue (e.g. protect the environment), but that turned out to reach that objective very poorly at a huge cost, as being more commendable than (Experiment 1-4), and as deserving equal support as (Experiment 5), a policy decision motivated by a selfish reward, but that dramatically helped the issue while saving vast amounts of money. This preference was observed whether the decision was made by a CEO or minister and across four different issues. Independent manipulation of intent and efficiency (Experiment 5) suggested that folk judgments of policies are characterized both by low sensitivity to differences in efficiency expressed in numerical format, and to high sensitivity to actors' personal motivations. Moreover, participants' moral commitment to an issue predicted greater perceived commendability of altruistic but inefficient policies meant to tackle the issue

(Experiment 1-4), and greater support for any policy that somehow contributed to help the issue, whether their efficiency was low or high, and the intention driving them selfish or altruistic (Experiment 5).

**Keywords:** Consequences; Efficiency; Intentions; Moral commitment; Moral judgment; Policy

## Introduction

Would you prefer a public policy that is highly efficient in taming CO<sub>2</sub> emissions but turns out to be motivated primarily by the search for profit or votes, or would you prefer a policy that fares poorly in curbing emissions, but is driven by the good intention to help this issue? One may suspect that many people would spontaneously see the former as the most moral of the two options.

In the public and private organizations of democratic societies, electoral processes put vast amounts of executive power in the hands of particular individuals such as ministers and CEOs. This power enables them to make decisions that can massively help society, but also result in considerable wasted money and opportunities. From a consequentialist or pragmatic standpoint, the only thing that should matter is how successful policies are at reaching their goals—such as addressing global warming, promoting sex equality, or ensuring national defense—relative to the amounts of resources mobilized for their implementation. Assuming the ends pursued by public action enjoy wide public support, citizens' opinions on those policies should not be swayed by whether the actors that put them into effect personally benefitted, electorally or financially.

However, cost-benefit analyses, despite being central and routine for evidence-based policy makers, do not easily factor into citizens' judgments about policy in comparison to the weight spontaneously given to actors' motivations. In the domain of charity giving, donors strongly underestimate differences between charities' effectiveness (Caviola et al., 2020), and tend to care less about how much the charities to which they give actually contribute to aggregate welfare, than to whether those charities pursue their preferred causes (Berman et al., 2018). In the domain of folk economic beliefs, the intuition that the “selfish” pursuit of monetary interest on the market is morally bad and incompatible with the common good is very

widespread, making people reject Adam Smith's notion that private self-regard can be an efficient way of making others better off (Boyer & Petersen, 2018; Rubin, 2003, 2014). Likewise, we suspect that the belief that a technology is driven by the "altruistic" intention to help a social issue may be a stronger predictor of popular support for it than whether it effectively works at solving the issue. For instance, despite wide scientific agreement on the important role that GMO foods may play in the fight against world hunger (Wesseler & Zilberman 2014), part of the resistance they face seems justified by the intuition that the search for profitability is so central to their development that it makes them "impure" (of course, other intuitions, such as an essentialist folk biology, likely play an important role in causing popular distrust towards GMOs; Blancke et al., 2015). In a similar way, the prioritization of intent over efficiency may play a role in the public enthusiasm for solar or wind energies, labelled as "good", over nuclear energy, labelled as "bad", despite the latter being orders of magnitude more productive and reliable than the former for generating electricity (Cameron & Taylor, 2011; Siqueira et al., 2019).

More generally, the intuition that "it's the thought that matters" is central to the moral matrices of so-called "WEIRD" and "non-WEIRD" societies, and as such may be human morality's clearest universal (Henrich, Heine, Norenzayan, 2010; Young & Tsoi, 2013). It is recognized by almost all moral philosophical theories (Walmsley & O'Madagain, 2020), not just deontological (Kant, 1785/2005) but also virtue-ethical (Slote, 1995) and consequentialist ones (Sidgwick, 1884), and is part of a wide class of deontological deviations from consequentialism that comprises intuitions on the importance of fairness, the prohibition to sacrifice, or the preference for harmful omission over less harmful acts (Baron, 1994; Baron & Spranca 1997; Graham & Haidt 2012; Haidt 2001, 2012; Tetlock et al., 2000; Tetlock, 2003; Spranca, Minsk & Baron 1991).

Of course, we do not doubt that giving central moral importance to others' subjective motivations is often quite rational. Whenever others, such as our friends and spouses, can

physically affect our welfare, ensuring their values and goals do not conflict too much with one's own is obviously adaptive. However, the existence of organizations in modern democratic societies creates special situations in which, through policy decisions, the reins of social change are put in the hands of particular individuals with whom most people have no close relationship. In those circumstances, by consequentialist standards, what should mostly matter is the policies success' and their cost for society, not the nobility of the motivations driving the actors responsible for their design and implementation.

### **Overview of the experiments**

The present paper presents 5 online experiments meant to explore the relative weight of policy makers' intentions and policies' efficiency in folk moral judgments and preferences for policies. All experiments were run on French participants using the software Qualtrics, and recruited on Foule Factory, a French crowdsourcing platform equivalent to Amazon Mechanical Turk.

Experiments 1 to 4 tested whether participants would find an inefficient policy decision driven by an altruistic intention in the decision maker (e.g. a policy intended to protect the environment, but which does so poorly and costs a fortune), as more commendable than a policy decision that is massively efficient but turns out to be motivated by selfishness. We varied which social issue the decision maker's intention altruistically helped vs. selfishly ignored, and whether the decision maker was the CEO of a private company or a minister. Experiment 5 adopted a different dependent variable in terms of support for the policy (rather than commendability of the decision), and focused on the issue of protecting the environment. Its design was more complex, as it independently manipulated the policy's efficiency, the intention driving it, as well as the decision maker's identity, in order to parse out their relative contribution to participants' support for policies.



In all experiments, we also looked into how participants' degree of conviction that tackling the issue at stake is a moral imperative—i.e. their moral commitment to the issue—modulated their judgments of policy decisions (Skitka et al., 2005; Skitka, 2010). We expected moral commitment to increase approbation of altruistically motivated policies, and to decrease concern for their efficiency.

## **Experiments 1-4**

Experiments 1 to 4 adopted the same design. However, they touched on four distinct issues, and, for believability reasons, they involved two types of decision makers (which are specified in parentheses below). Experiment 1: protecting the environment using carbon capture technologies (CEO). Experiment 2: fighting for sex equality in the workplace (CEO). Experiment 3: promoting France's diplomatic authority in the world (minister). Experiment 4: regulating immigration (minister). For conciseness purposes, Experiments 1-4 are presented here in the same section.

## **Method**

### ***Participants***

247 participants were recruited for Experiment 1, 211 participants for Experiment 2, 202 participants for Experiment 3, and 194 participants for Experiment 4. All were paid €0.50.

### ***Materials and procedure***

Experiment 1-4 touched on distinct issues but were otherwise identical. Our description of the four studies' materials and procedure is based on Experiment 1, which touched on the issue of protecting the environment and had a CEO as decision maker. Participants had to give their informed consent to participate. Following a within-subjects design, they were randomly presented with two vignettes that narrated the discussion between a CEO and his advisor on a new policy they were considering to implement. In the Altruistic intent/Low efficiency version of the policy decision, the CEO reported being driven by the altruistic intention to help the issue when considering to launch the policy: *"I honestly care deeply about [fighting climate change]"*. However, his advisor informed him that the policy would cost 100 million euros to the company and that it would only do little to help the issue: *" +10%"* impact at the national level and *" +0.001%"* impact at the global level. In contrast, in the Selfish intent/High efficiency version of the policy decision, the decision maker reported being motivated by a selfish reward unrelated to solving the issue: *"Honestly, I really don't care about fighting climate change. What I care about is [the profits we can make and our company's success.]"* However, his advisor notified him that the policy would allow for 100 million euros to be saved by the company, and that it would be highly effective at helping the issue: *" +80%"* impact at the national level, and *" +1%"* impact at the global level. The Selfish intent/High efficiency policy was thus 8 times more impactful at a national level, and 1000 times more impactful at the global scale, than the Altruistic intent/Low efficiency policy, in addition to being vastly profitable financially as opposed to costing huge amounts of money. Below is the verbatim of the vignettes used in Experiment 1 (protecting the environment, CEO as decision maker):

### **Altruistic intent/Low efficiency**

The CEO of a global cement and concrete company is considering implementing a new policy: directly capturing CO<sub>2</sub> as it is leaving its factories' smokestacks, and then recycle it as an ingredient that makes the factories' concrete stronger.

His chief advisor told him, "If you implement this policy, you will cut our company's CO<sub>2</sub> emissions by 10%. But even with our company's size, this would make only a minor difference in the fight against climate change, as it would decrease the world's total CO<sub>2</sub> emissions by only 0.001%. This policy would only help the world a little bit."

The CEO responded to the advisor, "I honestly care deeply about fighting climate change. How much would the policy cost?"

The advisor responded, "By my calculations, our company would lose about 100 million euros with that policy, because the new technology would significantly increase production costs."

The CEO decided to implement the policy. He indeed reduced his company's CO<sub>2</sub> emissions by 10%, thus reducing the world's total net CO<sub>2</sub> emissions by .001%. However, because the CEO cared deeply about his action, he thought it was definitely worth the 100 million euros missed profits.

### **Selfish intent/High efficiency**

The CEO of a global cement and concrete company is considering implementing a new policy: directly capturing CO<sub>2</sub> as it is leaving its factories' smokestacks, and then recycle it as an ingredient that makes the factories' concrete stronger.

His chief advisor told him, "If you implement this policy, you will cut our company's CO<sub>2</sub> emissions by almost 80%. Given our company's size, the impact on fighting climate change would be massive, as it would decrease the world's total CO<sub>2</sub> emissions by 1%. This is the kind of policy that could really help the world."

The CEO responded to the advisor, "Honestly, I really don't care about fighting climate change. What I care about are the profits we can make and our company's success. How much would the policy cost?"

The advisor responded, "By my calculations, our company would *make* about 100 million euros with that policy, because the new technology would significantly decrease production costs."

The CEO decided to implement the policy. He indeed reduced his company's CO<sub>2</sub> emissions by 10%, thus reducing the world's total net CO<sub>2</sub> emissions by .001%. He indeed reduced his company's CO<sub>2</sub> emissions by 80%, thus reducing the world's total net CO<sub>2</sub> emissions by 1% and significantly contributing to the fight against climate change. However, more importantly to the CEO, he also made a nice profit of 100 million euros.

Participants were asked to morally assess each one of the two policy decisions by answering the question "*To what extent would you say that the [CEO's] decision was*

*commendable?*” displayed immediately after each vignette, on the same page. Responses were collected on a 7-point Likert scale ranging from 1, “*Not at all commendable*”, 2, “*Not commendable*”, 3, “*Somewhat not commendable*”, 4, “*Undecided*”, 5, “*Somewhat commendable*”, 6, “*Clearly commendable*”, to 7, “*Totally commendable*”.

After delivering commendability ratings, participants reported their level of moral commitment to the issue at stake by answering the question “*To what extent do you think that [protecting the environment] should be the government’s priority?*” (Marie, Altay, Strickland, 2020). Responses were collected on a slider scale ranging from 0, “*I don’t care*” to 100, “*Absolute priority*”, with 50 as default slider position. The questionnaire ended with demographic questions: political orientation on a 1-item left-right axis, sex, age, level of education.

As already mentioned, the only two difference between Experiments 1-4 were the issue at stake and the decision maker’s identity (see ESM for vignettes). On the issues of environmental protection (Experiment 1) and the pursuit of sex equality (Experiment 2), we chose the CEO of a private company as decision maker because those were policy interventions that one can easily represent as being implemented at the level of a private company. In contrast, we saw it as more believable that decisions meant to further the values of national authority (Experiment 3) and immigration regulation (Experiment 4) be taken at the governmental level, hence a minister as decision maker.

Varying the decision maker’s identity implied adapting the selfish reward in the Selfish intent/High efficiency decision. It also required adapting which entity would lose vs. gain the 100 million euros as a consequence of the policy in the Altruistic intent/Low efficiency and Selfish intent/High efficiency decisions respectively. When the decision maker was a CEO, their selfish reward was financial (“*What I care about are the profits we can make and our company’s success.*”), and the policy was described as being costly, or beneficial, to the company. In contrast, when the decision maker was a minister, the selfish reward was electoral

(“*What I care about is to save public money so as to increase chances that our party wins the next elections.*”), and the policy as being costly, or beneficial, to the state.

## Results

All analyses in this paper were run in R (version 3.6.2) using R Studio (Version 1.2.5033). No participant was excluded from Experiments 1 to 4. Participants’ demographic characteristics were as follows. Experiment 1:  $M_{age} = 44.2$  ;  $SD_{age} = 13.5$ ; 61.5% female. Experiment 2:  $M_{age} = 43.8$  ;  $SD_{age} = 12.7$ ; 60% female. Experiment 3:  $M_{age} = 42.5$  ;  $SD_{age} = 13.6$ ; 61% female. Experiment 4:  $M_{age} = 43.8$  ;  $SD_{age} = 13.1$ ; 64% female.

### *Effect of decision type on commendability judgments*

Our primary analyses were based on a dataset aggregating Experiments 1-4 (N = 854). Commendability of the policy decision was regressed in two linear mixed-effects models (using the “lmerTest” R package) with decision type, issue, and moral commitment as fixed effects and participant as random effect (see “main models” in Electronic Supplementary Materials after the References section). Main effects are reported from the first model containing the main effects only (decision type, issue, moral commitment), and two-way interactions are reported from the second model containing both the main effects and their two-way interactions. We report regression coefficients, 95% CIs between brackets, and *ps* from these models.

Commendability ratings were highest regardless of decision type on the issue of environmental protection, followed by sex equality (main effect of issue:  $\beta = -0.17$ , [-0.29,-0.05],  $p < 0.001$ ), promoting France’s authority ( $\beta = -0.52$ , [-0.65,-0.39],  $p < 0.001$ ), and regulating immigration ( $\beta = -0.63$ , [-0.76,-0.50],  $p < 0.001$ ). In analyzing across all issues,

commendability judgments were significantly lower in the Selfish intent/High efficiency condition than in the Altruistic intent/Low efficiency condition (main effect of decision type:  $\beta = -0.54$ ,  $[-0.62, -0.46]$ ,  $p < 0.001$ ). We recognized that our within-subjects design could have incentivized participants to exaggerate the difference between their two judgments compared to a between-subjects design. Yet, the Selfish intent/High efficiency decision was still rated as being less commendable than the Altruistic intent/Low efficiency decision when subsetting the data to the first condition participants had been exposed to only (main effect of decision type:  $\beta = -0.56$ ,  $[-0.67, -0.44]$ ,  $p < 0.001$ ). Importantly, the difference between the two decision types was moderated on the issues of France's authority (interaction between decision type and issue:  $\beta = 0.57$ ,  $[0.34, 0.81]$ ,  $p < 0.001$ ) and regulating immigration ( $\beta = 0.50$ ,  $[0.27, 0.74]$ ,  $p < 0.001$ ). Note that on those two latter issues, the decision maker was a minister rather than a CEO (See figure 1a).

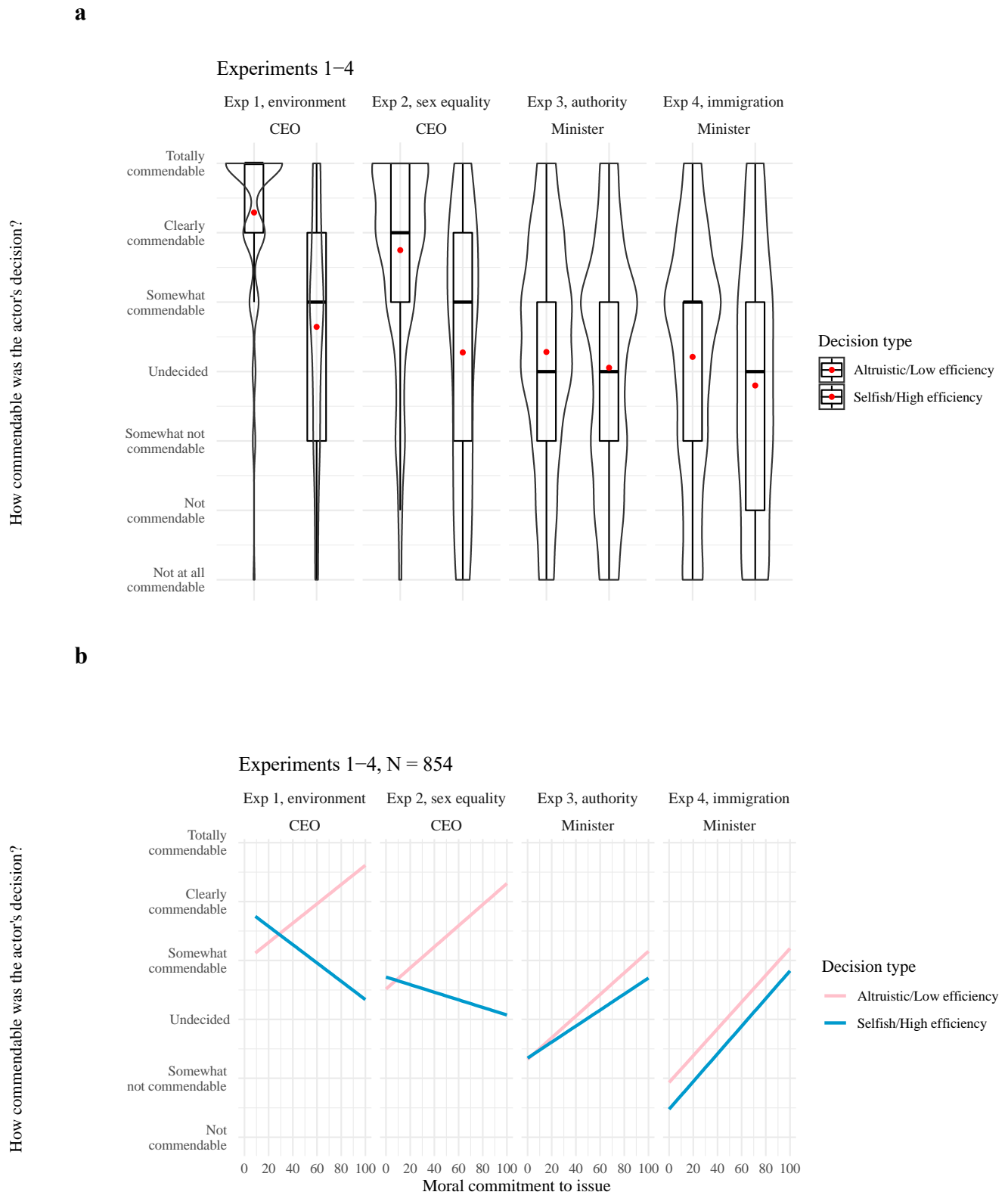
As *post hoc* analyses, we also compared commendability judgments of the two decision types in each experiment or issue separately. On the issue of environmental protection (Experiment 1), participants rated the CEO's decision in the Altruistic intent/Low efficiency condition ( $M = 6.29$ ,  $Median = 7$ ,  $SD = 1.14$ ) as being significantly more commendable than in the Selfish intent/High efficiency condition ( $M = 4.64$ ,  $Median = 5$ ,  $SD = 1.68$ ),  $t(246) = 11.87$ ,  $p < 0.001$  (paired), Cohen's  $d = 1.15$  (large effect). The difference on the issue of sex equality (Experiment 2) had a similar size, with mean commendability of the CEO's decision in the Altruistic intent/Low efficiency condition ( $M = 5.75$ ,  $Median = 6$ ,  $SD = 1.40$ ) being significantly higher than in the Selfish intent/High efficiency condition ( $M = 4.27$ ,  $Median = 5$ ,  $SD = 1.87$ ),  $t(210) = 8.75$ ,  $p < 0.001$  (paired),  $d = 0.89$  (large effect). Differences between decision types on the issues of enhancing France's authority and regulating immigration were smaller, but went in the same direction. As regards enhancing France's authority, participants judged the minister's decision in the Altruistic intent/Low efficiency condition ( $M = 4.28$ ,  $Median = 4$ ,  $SD = 1.64$ ) as slightly more commendable than in the Selfish intent/High efficiency condition, but

the difference was not significant ( $M = 4.05$ ,  $Median = 4$ ,  $SD = 1.73$ ),  $t(201) = 1.27$ ,  $p = 0.2$  (paired),  $d = 0.13$  (negligible effect). On the issue of immigration regulation, participants also rated the minister's decision in the Altruistic intent/Low efficiency condition ( $M = 4.21$ ,  $Median = 5$ ,  $SD = 1.73$ ) as more commendable than in the Selfish intent/High efficiency condition ( $M = 3.80$ ,  $Median = 4$ ,  $SD = 1.86$ ),  $t(193) = 2.68$ ,  $p < 0.01$  (paired),  $d = 0.23$  (small effect).

### ***Effect of moral commitment on commendability judgments***

Participants reported average levels of moral commitment that were highest on the issue of environmental protection ( $M = 80.2$ ,  $Median = 82$ ,  $SD = 18.8$ ), followed by sex equality ( $M = 68.9$ ,  $Median = 71$ ,  $SD = 23.4$ ), France's authority ( $M = 56.4$ ,  $Median = 60$ ,  $SD = 28.3$ ) and immigration ( $M = 52.4$ ,  $Median = 54.5$ ,  $SD = 21.8$ ). For ease of interpretation, the effect of commitment on commendability judgments on each issue was assessed using simple linear regressions rather than three-way interactions between commitment, decision type, and issue (see ESM).

On all four topics, greater moral commitment to the issue was consistently associated with more praise of the Altruistic intent/Low efficiency decision: Environmental protection:  $\beta = 0.27$ ,  $[0.15, 0.39]$ ,  $p < .001$ ; sex equality:  $\beta = 0.30$ ,  $[0.17, 0.43]$ ,  $p < .001$ ; France's authority:  $\beta = 0.24$ ,  $[0.11, 0.38]$ ,  $p < .001$ ; regulating immigration:  $\beta = 0.37$ ,  $[0.24, 0.50]$ ,  $p < .001$  (see ESM). In contrast, greater moral commitment to the issue had an unpredictable effect on commendability judgments of the Selfish intent/High efficiency decision. It was negative on issues of environmental protection ( $\beta = -0.17$ ,  $[-0.30, -0.05]$ ,  $p < 0.01$ ) and negative but non-significant on promoting sex equality ( $\beta = -0.08$ ,  $[-0.22, 0.06]$ ,  $p = 0.24$ ), where the decision maker was a CEO. However, it was positive on issues of France's authority ( $\beta = 0.17$ ,  $[0.03, 0.31]$ ,  $p < 0.01$ ) and regulating immigration ( $\beta = 0.36$ ,  $[0.23, 0.49]$ ,  $p < 0.001$ ), where the decision maker was a minister.



**Figure 1:** Commendability judgments of policy decisions as a function of decision type in Experiments 1-4: a, main effect, black lines are medians and red dots means; b: linear regression on moral commitment to issue.



## Discussion

Experiments 1 to 4 found that a policy decision that only helped a societal issue a little bit while coming at a huge cost, but was motivated by an altruistic intention, was judged as more commendable than a decision at least 8 times more impactful, and financially profitable, but which happened to be driven by a selfish motive. This result generalized across four distinct issues—protecting the environment, fighting for sex equality, promoting France’s authority in the world, and regulating immigration—and was observed whether the decision maker was a CEO or a minister. We also consistently found that the more participants were morally committed to an issue, the more praiseworthy they saw an altruistically motivated policy despite its poor efficiency.

Since Experiments 1-4 did not independently manipulate issue and decision maker identity, it is unclear why the Altruistic intent/Low efficiency policy decision was rated higher in Experiments 1-2 than in Experiments 3-4. This may be due to people expecting greater exertions for the common good from ministers than from CEOs, because the former are state servants while the latter are typically seen as working for private interests only (Rubin, 2014). This assumption may have caused people to think that ministers (Exp. 3-4) deserve less praise than CEOs (Exp. 1-2) for the same kind of altruistic decision. A second and complementary explanation of the Altruistic intent/Low efficiency decision being rated higher on environmental protection (Exp. 1) and sex equality (Exp. 2) than on France’s authority (Exp. 3) and immigration regulation (Exp. 4) is that participants were on average more morally committed to the former two issues than to the latter two. Between-participants analyses within each issue also suggested that greater commitment is associated with more praise of the Altruistic intent/Low efficiency decision (*cf.* Figure 1b).

## Experiment 5

The design employed thus far showed that a combination of good intent but low efficiency tended to be seen as more commendable than a combination of bad intent but high efficiency. While those results unambiguously show that laypeople's judgments of policy decisions strongly deviate from consequentialism, it is unclear whether they are due to participants giving considerable importance to the decision maker's motivations, being insensitive to vast differences in efficiency, or a combination of both mechanisms.

Answering this question requires examining whether judgments track differences in efficiency in the absence of any intention-relevant information, and independently manipulating intent and efficiency. To do this, Experiment 5, which focused on the issue of protecting the environment, adopted a full factorial design to examine how support for a low *vs.* high efficiency policy decision without intention specified would be affected by introducing cues to the decision maker's altruistic *vs.* selfish intention. We meant the no intentions baseline to gauge how consequentialist participants would spontaneously be in the absence of any cue to the decision maker's subjective motivations. We expected to reproduce our finding that an altruistic but low efficiency decision would be rated higher than a selfish but high efficiency decision. Experiment 5 also varied the decision maker's identity (CEO *vs.* minister). Finally, we were interested to examine the effect of moral commitment on support for policy in each combination of efficiency, intent, and decision maker's identity.

### Method

#### *Participants*

661 French participants were recruited on Foule Factory for Experiment 5 in exchange for €0.30.  $M_{age} = 39$ ;  $SD_{age} = 12.4$ ; 48% female. No participant was excluded from the data.

### ***Materials and procedure***

Experiment 5 focused on the issue of protecting the environment only. The vignettes and procedure were the same as in Experiment 1 except for the following changes. Instead of comparing two decision types only, we independently varied the decision maker's intention across three levels (no intention specified, altruistic, and selfish), the policy's efficiency across two levels (low vs. high), and the decision maker's identity across two levels (CEO vs. Minister). The design was now a 3x2x2 design with 12 conditions, between-subjects in order to avoid that repeated measures artificially inflate differences in responses between conditions.

Altruistic and selfish intentions reported by the decision maker were the same as in Experiments 1, and we suppressed the sentences providing information on the decision maker's intentions in the two relevant paragraphs of the vignettes in the no intention specified conditions. The selfish motivation of the decision maker where slightly changed to "*improving the company's image*" when they were a CEO, and to "*improving the [minister's] political party's image in order to increase chances of winning the presidential elections*" when they were a minister.

We also increased the contrast between the low and high levels of the efficiency factor compared to Experiments 1-4 in order to make it more salient to participants, even in a between-subjects design. In the low efficiency conditions, the policy was described as reducing carbon emissions by only 2% at the national level (instead of 10% in Experiment 1-4) and 0.001% at the global level, and as costing 100 million euros in the long run (to the state when it was a minister vs. to the company when it was a CEO). In contrast, in the high efficiency conditions, the policy was described as contributing 80% reduction of CO<sub>2</sub> emissions at the national level and 10% reduction at a global scale (instead of 1% in Experiment 1-4), and as allowing 100

million euros to be saved in the long run. The high efficiency policy was thus 40 times more impactful at a national level, and 10 000 times more impactful at the global scale, than the low efficiency policy, in addition to being vastly financially profitable as opposed to costing a lot of money.

We realized that our dependent variable “*To what extent would you say that the [e.g. CEO's] decision was commendable?*” could implicitly prime participants with a virtue-ethics mindset by making them focus on the moral character of the actor’s decision to the detriment of the policy’s efficiency. This could artificially increase chances their responses deviate from consequentialism. We therefore adopted a more neutral formulation: “*To what extent do you support this policy?*”, with responses collected on a 7-point scale (1, “*Not at all*”, 2, “*Clearly not*”, 3, “*Somewhat not*”, 4, “*Undecided*”, 5, “*Somewhat*”, 6, “*Clearly*”, 7, “*Totally*”).

Participants then reported their level of moral commitment to environmental protection on a 3-item scale: “*Protecting nature is an absolute moral imperative*”, “*The conviction that one must fight to protect the environment is central to my identity*” (inspired from Skitka et al., 2005; Stahl, Zaal, Skitka, 2016); “*Protecting the environment should be the government’s priority*” (Marie, Altay, Strickland, 2020). Responses were collected on 0-100 slider scales anchored in 0, “*Totally disagree*”, and 100, “*Totally agree*”, with 50, “*I don’t know*” as default position. The experiment ended with demographic questions: political orientation, sex, age, and education.

## Results

### *Effect of efficiency, intent and decision maker identity on support for policy*

Results were analyzed using two multiple regression models of support for policy judgments on policy efficiency (low vs. high), decision maker intent (unspecified vs. altruistic vs. selfish) and decision maker identity (CEO vs. minister) (see “main models” in ESM). Levels

listed first in parentheses were set as baselines for each factor in the regression. Main effects are reported from a first model containing only the main effects, and the two-way interactions between those factors are reported from a second model containing both the main effects and the two-way interactions. Welch's *t*-tests and Cohen's *d*s were also performed as *post hoc* analyses; all are non-paired.

The high efficiency policies ( $M = 5.87$ ,  $Median = 6$ ,  $SD = 1.42$ ) garnered more support than the low efficiency policies ( $M = 4.65$ ,  $Median = 5$ ,  $SD = 1.85$ ), regardless of intention and decision maker identity (main effect of efficiency:  $\beta = 0.66$ ,  $[0.53, 0.79]$ ,  $p < 0.001$ ). This difference was greater when the decision maker was a minister than a CEO (interaction between efficiency and decision maker identity:  $\beta = 0.61$ ,  $[0.36, 0.86]$ ,  $p < 0.001$ ), an effect driven mainly by low efficiency policies being supported less when a minister was implementing them ( $M = 3.98$ ,  $Median = 4$ ,  $SD = 1.85$ ) than when it was a CEO ( $M = 5.33$ ,  $Median = 6$ ,  $SD = 1.58$ ),  $t(322) = 7.17$ ,  $p < 0.001$ ,  $d = 0.7$ . Focusing on the data corresponding to conditions that did not ascribe any intention to the decision maker, high efficiency policies ( $M = 6.16$ ,  $Median = 7$ ,  $SD = 1.09$ ) were more supported than low efficiency policies ( $M = 4.58$ ,  $Median = 5$ ,  $SD = 1.67$ ),  $t(193) = -8.37$ ,  $p < 0.001$ ,  $d = 1.12$ ). Although this latter effect was, statistically speaking, quite clear, it is striking that participants' judgments showed so little responsiveness to a difference in efficiency described as huge in the vignettes. While support for high efficiency policies was around "Clearly", support dropped by only 1,5 scale points to "Undecided-Somewhat" when efficiency was low, despite their gigantic cost and tiny impact.

Attributing a selfish motivation to the decision maker decreased support for the decision compared to when no intention was specified (main effect of selfish intent:  $\beta = -0.47$ ,  $[-0.62, -0.31]$ ,  $p < 0.001$ ), both on low and high efficiency policies. In contrast, ascribing an altruistic intent to the decision maker increased support for the decision (main effect of altruistic intent:  $\beta = 0.17$ ,  $[0.01, 0.33]$ ,  $p < 0.05$ ), except on high efficiency policies, presumably because of a

ceiling effect on the dependent variable (interaction between efficiency and altruistic intent:  $\beta = -0.39, [-0.70, -0.08], p < 0.01$ ).

In line with results from Experiments 1-4, participants in Experiment 5 supported the Altruistic intent/Low efficiency policies at least to the same extent ( $M = 5.31, Median = 6, SD = 1.66$ ) as the Selfish intent/High efficiency policies ( $M = 5.33, Median = 6, SD = 1.67$ ),  $t(218) = -0.08, p = 0.93$ . Breaking down by decision maker identity, support for the Altruistic intent/Low efficiency policy was slightly higher ( $M = 6.04, Median = 6, SD = 1.24$ ) than for the Selfish intent/High efficiency policy ( $M = 5.6, Median = 6, SD = 1.56$ ) when a CEO was implementing the policy, but the difference was not significant:  $t(103) = 1.67, p = 0.1$ . When a minister was making the decision, support for the Altruistic intent/Low efficiency policy ( $M = 4.62, Median = 5, SD = 1.72$ ) was slightly lower than for the Selfish intent/High efficiency policy ( $M = 5.07, Median = 5, SD = 1.75$ ), but the difference was not significant:  $t(109) = -1.36, p = 0.18$ . Let us highlight again that the equal support for the Altruistic intent/Low efficiency and the Selfish intent/High efficiency policies were found in a context where the discrepancy between the low and high efficiency levels was even greater than in Experiments 1-4, and where the new dependent variable focused on “*the policy*”, not “*the [CEO’s] decision*”, which should have elicited more consequentialist judgments (*cf.* Materials and procedure of Experiment 5).

There also was a main effect of decision maker identity, such that policies implemented by a minister garnered less support across the board than those made by a CEO ( $\beta = -0.48, [-0.61, -0.35], p < 0.001$ ).

### ***Effect of moral commitment on support for policy***

The effect of participants’ degree of moral commitment to environmental protection on support for the policies was overall positive. It was first assessed using Experiment 5’s main regression models (see “main models” in ESM), in which the main effect of commitment was:  $\beta = 0.26, [0.20, 0.33], p < 0.001$ . The effect of commitment on policy support was marginally smaller on

high than low efficiency policies, presumably due to a ceiling effect on the policy support scale (interaction between commitment and high efficiency:  $\beta = -0.12$ ,  $[-0.25, 0.01]$ , but the interaction was not significant:  $p = 0.07$ ). Consistent with Experiments 1-4, the positive effect of commitment on support for policy seemed greater when the decision maker was motivated by altruism than when no intention was specified, but the interaction was not significant (interaction between commitment and altruistic intent:  $\beta=0.12$ ,  $[-0.02, 0.27]$ ,  $p=0.10$ ). The positive relationship between commitment and support for policy was not moderated by the decision maker being driven by selfishness compared to when no intention was specified (interaction between commitment and selfish intent:  $\beta=0.03$ ,  $[-0.14, 0.19]$ ,  $p=0.75$ ).

Figure 2c below shows that the relationship between moral commitment and support for policy was positive in almost all conditions. The only exception to this trend was when the policy was simultaneously driven by a selfish intention and of high efficiency. To get a detailed understanding of this relationship, we complemented our main analysis with simple linear regressions of policy support on commitment in each one of the 12 conditions:

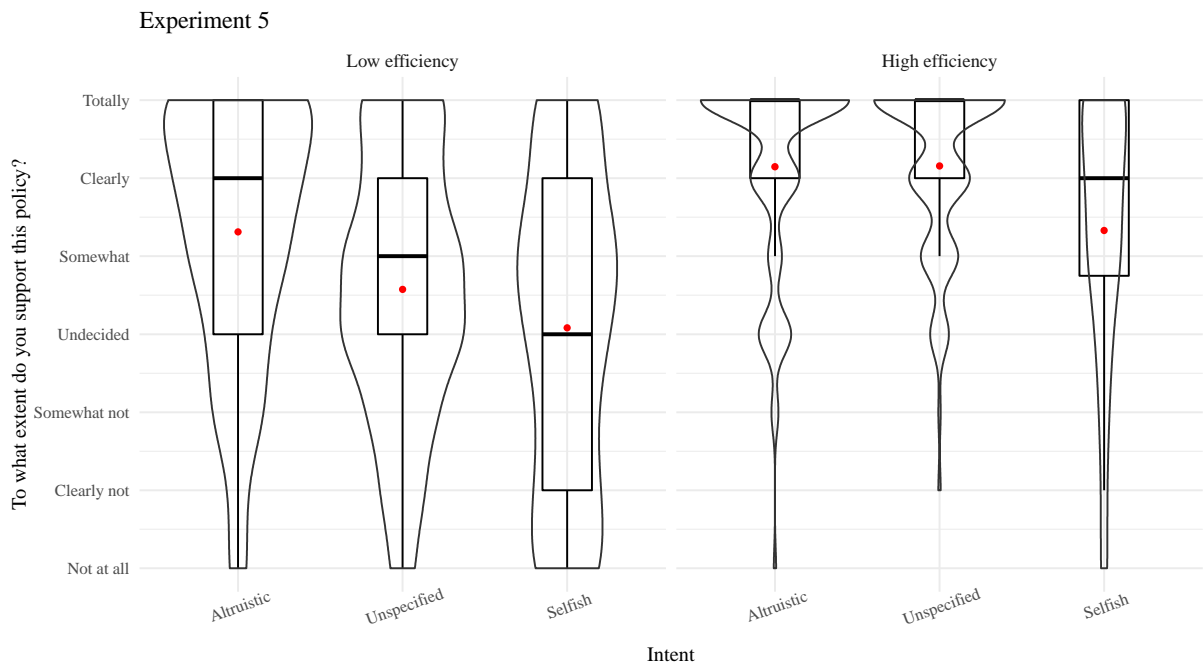
Low efficiency, CEO: Altruistic intent:  $\beta=0.53$ ,  $[0.29, 0.77]$ ,  $p<0.001$ ; No intent:  $\beta=0.39$ ,  $[0.15, 0.64]$ ,  $p<0.001$ ; Selfish intent:  $\beta=0.29$ ,  $[0.03, 0.56]$ ,  $p<0.001$ .

High efficiency, CEO: Altruistic intent:  $\beta=0.43$ ,  $[0.18, 0.68]$ ,  $p<0.001$ ; No intent:  $\beta=0.30$ ,  $[0.04, 0.56]$ ,  $p<0.05$ ; Selfish intent:  $\beta=0.04$ ,  $[-0.23, 0.31]$ ,  $p=0.76$ .

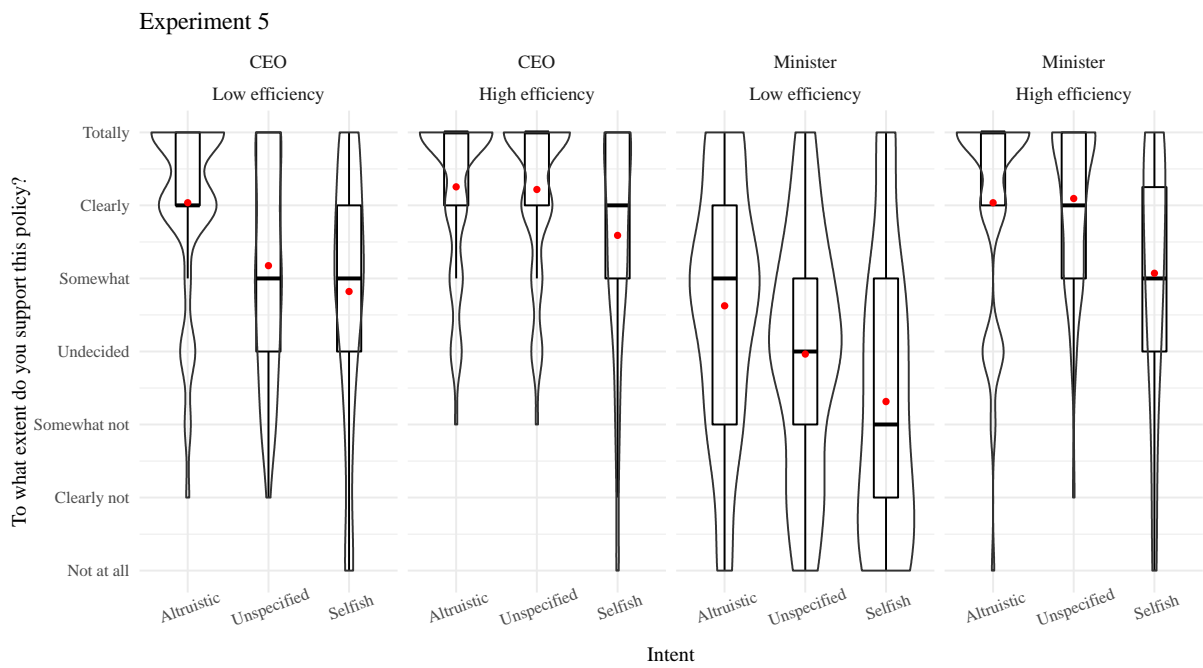
Low efficiency, Minister: Altruistic intent:  $\beta=0.37$ ,  $[0.11, 0.62]$ ,  $p<0.01$ ; No intent:  $\beta=0.26$ ,  $[-0.01, 0.52]$ ,  $p<0.05$ ; Selfish intent:  $\beta=0.35$ ,  $[0.09, 0.61]$ ,  $p<0.01$ .

High efficiency, Minister: Altruistic intent:  $\beta=0.49$ ,  $[0.25, 0.73]$ ,  $p<0.001$ , No intent:  $\beta=0.29$ ,  $[0.03, 0.56]$ ,  $p<0.05$ ; Selfish intent:  $\beta=0.08$ ,  $[-0.19, 0.36]$ ,  $p=0.54$ .

**a**

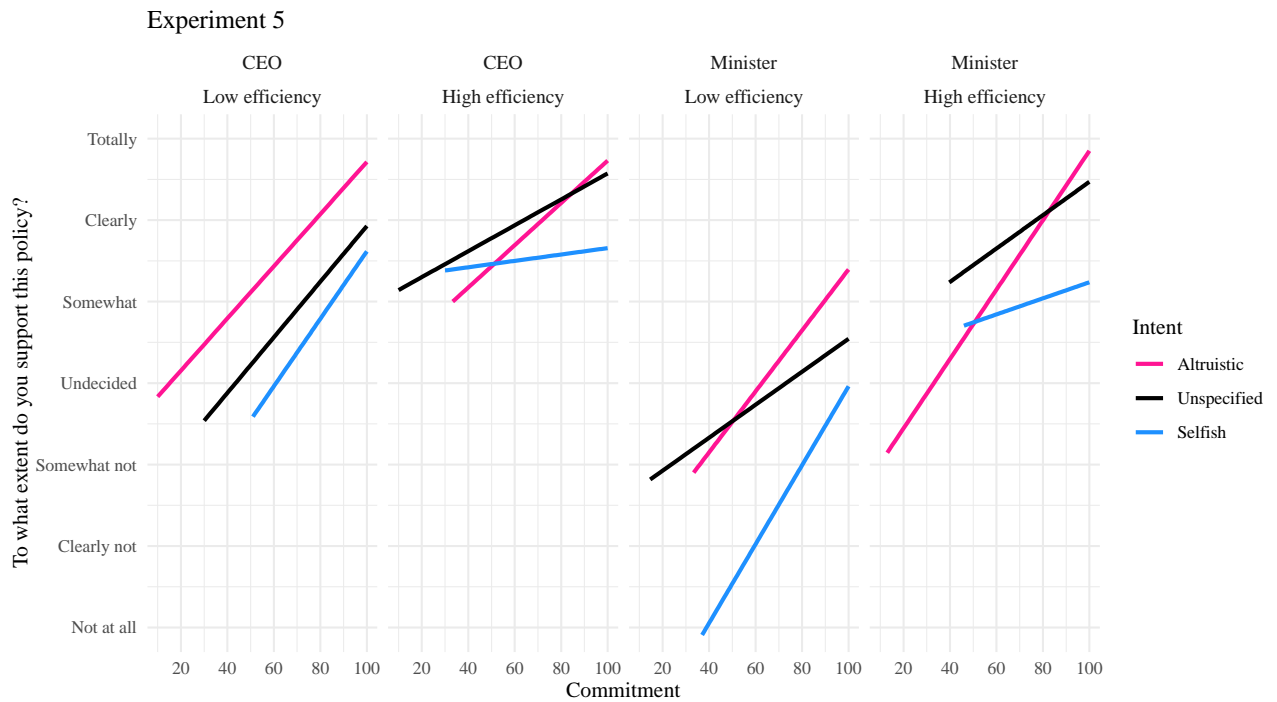


**b**



**c**





**Figure 2:** Support for policy judgments in Experiment 5 as a function of intent and efficiency: a, regardless of decision maker identity; b: breaking down by decision maker identity; black lines are medians and red dots means. c: simple linear regressions on moral commitment to protecting the environment.

## General Discussion

Ministers and CEOs sit at the top of powerful organizational chains of command that allow them to make a big difference on whatever societal and economic issue they are touching on. Moreover, they are officials with whom most citizens will never have any personal relationship. By consequentialist standards—and contrary to relationships we have with friends and relatives, with whom physical interdependence is higher—whether ministers and CEOs, as particular individuals, share our values and interests or not should be of little import. What

citizens should mostly care about is how successful policies are at reaching their goals, while balancing that against their cost for society.

*Contra* consequentialism, our experiments found that laypeople's intuitive judgments of policy decisions were only little responsive to huge differences in efficiency expressed in numerical format. In Experiments 1 to 4, the fact that high efficiency policy decisions were at least 8 times more impactful than low efficiency policies, and that they would make society save rather than lose millions, never sufficed to make participants judge high efficiency decisions as more commendable than low efficiency ones. Experiment 5 showed that in the absence of any information on the decision maker's motivations in implementing the policy, low efficiency policies at least 40 times less impactful than high efficiency policies, in addition to being hugely costly rather than profitable, received only about 1,5 scale points less support than high efficiency policies ( $d = 1.12$ ).

Against consequentialist principles again, we also found that cues to the involvement of altruistic vs. selfish intentions in policies' implementation have substantial influence on how commendable they are perceived (Exp. 1-4), and how much popular support they get (Exp. 5). When the policy's efficiency was low—i.e. when support for it was not already near the scale's maximum—adding information on the decision maker's altruistic intentions had a moderate increasing effect on support compared to when no intention-relevant information was provided. Conversely, portraying the policy as motivated by selfish reward—whether financial or electoral—decreased support for it compared to when information on intentions was absent, regardless of how efficient the policy was.

Expert discussions of policies' costs and benefits typically rely on figures which are then communicated by the media and appropriated by the public in informal conversations (Katz & Lazarsfeld, 1955). Ordinary citizens should be able to take into account those figures if they are to form opinions and vote in pragmatic ways. To the extent policy makers' decisions are influenced by their citizens and employees' preferences, the combination of low

responsiveness to efficiency and high sensitivity to intentions have the potential to lead to considerable missed opportunities and wasted money for society. They may bring citizens to support highly expensive political programs that achieve little so long as they pursue values they hold dear, and to overlook vastly efficacious and profitable ones just because the reason for their implementation is perceived as lacking nobility.

What is worse, people's capacity to engage in cost/benefit thinking on politics appears further lowered when they moralize the issue more. Experiments 1 to 4 consistently showed that increased moral commitment to a societal value predicted greater support for well-intended but highly inefficacious and expensive policies aimed to promote the value. Experiment 5 suggested that greater moral commitment to an issue may increase support for *any* policy that somehow contributed to address the issue. The fact that moral commitment, in this latter study, had a positive effect on support for policy whether the policy sprung from altruism or selfishness is good news for the consequentialist. But the finding that the positive influence of commitment on policy support was *not* affected by efficiency, meaning that it extended to hugely costly programs that only make a tiny positive difference, is worrisome. It indicates that those who care the most about solving a societal issue may also be the ones most likely to relate deontologically to it—the ones least concerned to proportion efforts and resources to the targeted benefit (Atran, 2016; Skitka et al., 2005; Skitka 2010; Marie 2019, 2020).

An evolutionary approach to moral judgments can shed light on those results. Humans spent most of their evolution in small-scale societies in which the need to carry out joint ventures would have required them to constantly gauge other individuals' trustworthiness as potential partners (Baumard, André, Sperber 2013; Tooby & Cosmides, 1990). As a result, they evolved intuitive inference systems for tracking cues diagnostic of a person's value as a cooperator, and moral character (Uhlmann, Pizarro, Diermeier, 2015).

Among other algorithms, a cooperator-reliability barometer would be expected to contain a simple heuristic tracking whether, in a given domain of cooperation (e.g. a societal

issue), an individual did, or did not, do their part (e.g. Yamagishi et al., 2007). Throughout our experiments, what our participants mostly seemed to care about is whether the CEO or minister *did something* to reduce CO<sub>2</sub> emissions rather than nothing, not how effective their program was in doing so. Use of a simple “*Did they do their part?*” heuristic, that outputs a “yes”/”no” dichotomous outcome, may contribute to explain this myopia to differences in policy efficiency. Moreover, the strong relationship between increased commitment and support for all versions of a policy, whatever its level of efficiency, suggests that the contribution of this simple heuristic to moral judgment may be proportional to a person’s degree of moral commitment to the issue.

Note that people may have a myopic demand for cooperation in others while also expecting certain categories of actors to cooperate more than others. In Experiments 1-4, participants supported the same altruistic but low efficiency policy more if a CEO was implementing it than if it was a minister. In Experiment 5, low efficiency policies, regardless of the intentions driving them, garnered more support if they were implemented by a CEO than a minister. This may be due to French people having lower expectations towards private actors serving the common good than towards state servants (Rubin, 2014), and as a consequence, thinking that CEOs deserve more praise than ministers for identical, beneficial actions.

Moreover, among the cues an evolved cooperator-reliability detector would be expected to attend in priority are the *intentions* of whomever might become one’s cooperation partner. As a window into how a person orders her priorities, intentions provide crucial estimates of the future costs and benefits of personally collaborating with them, regardless of which particular situation one finds oneself in. Our brains are now thrown in a modern world in which a new and counter-intuitive moral standard has emerged—e.g. consequentialism—which requires them to turn down information to which they evolved to give instinctive importance: agents’ trustworthiness as cooperators—e.g. CEOs’ and ministers’ trustworthiness as individuals—*inferred from their intentions*. Meanwhile, this counter-intuitive consequentialist standard

requires them to focus on information they did not evolve to process: policies' abstract costs and benefits, expressed in multiple figures and percentages. Numerical notation is indeed a recent cultural innovation (Dehaene, 2009), and quantities are notoriously easier to represent visually as lengths or areas than in numbers (Kirk, 2016). The propensity of folk moral judgments of policies to depart from consequentialism may thus be explained in terms of a mismatch between our evolved psychology and the biologically novel cognitive demands of cost-benefit thinking.

### **Authors' Contributions**

All authors contributed to the studies' design. B. Strickland funded the research. A. Marie collected the data, performed the statistical analyses, reported the results and drafted the manuscript. H. Trad and B. Strickland provided revisions.

### **Conflict of interest**

The authors declare having no conflict of interest.

### **References**

References for this paper are reported in the general references section at the end of the dissertation.

### **Electronic Supplementary Materials (ESM)**

## Experiments 1-4

### Vignettes

The following displays vignettes used in Experiments 2 to 4. Paragraphs were separated by a line break in the original display. Translation was made from French using DeepL.

#### Experiment 2: Pursuing sex equality in the workplace

##### **Altruistic intent/Low efficiency**

The CEO of a major communications agency is considering a new policy: to change the recruitment and compensation procedures for its employees in order to increase parity in the company and reduce the wage gap between men and women.

His advisor says: "By implementing this policy, you increase gender equality by 10%. But, despite the size of our company, the impact in terms of promoting parity would be minor, as it would only reduce gender inequality in the workplace by 0.001% worldwide. This policy could help only a little bit in the fight against gender inequality. »

The CEO said: "I honestly care deeply about fighting against gender pay inequalities and striving for parity. How much would the policy cost us? »

The advisor replied: "According to my calculations, our company would lose a total of about 100 million euros with this policy, because it would significantly increase production costs. »

The CEO decided to implement the policy. He effectively managed to reduce gender inequality in his company by 10%, contributing to a minimal decrease of

##### **Selfish intent/High efficiency**

The CEO of a major communications agency is considering a new policy: to change the recruitment and compensation procedures for its employees in order to increase parity in the company and reduce the wage gap between men and women.

His advisor says: "By implementing this policy, you increase gender equality by 80%. Given the size of our company, the impact in terms of promoting parity would be massive, as it would reduce gender inequality in the workplace by 1% globally. That's the kind of policy that could really help combat gender inequality. "

The CEO said: "Honestly, I really don't care about fighting gender pay inequalities and striving for parity. What I care about is the profits we can make and the success of our business. How much would the policy cost us? »

The advisor replied: "According to my calculations, our company would gain a total of around 100 million euros with this policy, as it would significantly reduce production costs. "

The CEO decided to implement the policy. He effectively reduced gender inequality in his company by 80%, contributing to a significant decrease of 1% in the international business world. But more importantly for the CEO, he also made a profit of 100 million euros.

0.001% in the international business world.  
But because he was personally very keen on  
this gesture, the CEO considered it well  
worth the 100 million euros in lost profits.

### **Experiment 3: Promoting France's diplomatic authority in the world**

#### **Altruistic intent/Low efficiency**

The Minister of Defense is considering a new policy: to restore France's authority and influence in the world by investing more in its diplomatic corps and military infrastructure.

His adviser told him: "By implementing this policy, you increase France's geopolitical influence by 10%. But despite the size of our administration, the impact in terms of increased authority and influence abroad would be minor, as it would only reduce the problem of international political instability by 0.001%. Such a policy could help only a tiny bit in advancing France's interests and values in the world. "

The Minister declared: "I honestly care deeply about restoring France's influence and authority internationally. How much would the policy cost us? "

The adviser replied: "According to my calculations, the State would lose about 100 million euros in total with this policy, because it would significantly increase the wage bill. "

The minister decided to implement the policy. He actually managed to increase France's influence and authority abroad by 10%, thus contributing to a 0.001% decrease in political instability in the world. But because he was personally very keen on this gesture, the minister considered it well worth the 100 million euros lost by the state.

#### **Selfish intent/High efficiency**

The Minister of Defense is considering a new policy: to restore France's authority and influence in the world by investing more in its diplomatic corps and military infrastructure.

His adviser told him: "By implementing this policy, you increase France's geopolitical influence by 80%. Given the size of our administration, the impact in terms of increased authority and influence abroad would be massive, as it would reduce the problem of international political instability by 1%. This is the kind of policy that could really help advance France's interests and values in the world. »

The minister declared: "Honestly, I really don't care about restoring France's influence and authority internationally. What I care about is making budget savings that could increase the chances that our party will win the next election. How much would the policy cost us? »

The adviser replied: "According to my calculations, the State would save about 100 million euros in total with this policy, because it would significantly increase our bargaining power. »

The minister decided to implement the policy. He effectively increased France's influence and authority abroad by 80%, thus contributing to a 1% decline in political instability in the world. More importantly for the minister, he also achieved a budget saving of 100 million euros, which directly benefited his party.



## Experiment 4: Regulating immigration

### Altruistic intent/Low efficiency

The Minister of the Interior is considering a new policy: to make French migration policy more restrictive by making the granting of French nationality and the right of residence conditional on passing tests of knowledge of republican values (universalism and secularism), and of French culture, history and language.

His adviser told him: "By implementing this policy, you eliminate unwanted immigration and Islamism by 10%. But, despite the size of our administration, the impact in terms of immigration regulation would be minor, as it would only reduce the problem of unwanted immigration and Islamism by 0.001% at the global level. This policy could only help a little bit in the fight against the influence of people refusing our values. »

The Minister said, "I honestly care deeply about restricting immigration rights to those who know and accept our values and culture. How much would the policy cost us? »

The adviser replied: "According to my calculations, the State would lose about 100 million euros in total with this policy, because it would significantly increase the wage bill. »

The minister decided to implement the policy. He effectively managed to reduce undesirable immigration and Islamism by 10%, contributing to a minimal decrease of 0.001% in the world. However, because he was personally very keen on this gesture, the minister considered it well worth the 100 million euros lost by the state.

### Selfish intent/High efficiency

The Minister of the Interior is considering a new policy: to make French migration policy more restrictive by making the granting of French nationality and the right of residence conditional on passing tests of knowledge of republican values (universalism and secularism), and of French culture, history and language.

His adviser told him: "By implementing this policy, you eliminate undesirable immigration and Islamism in France by 80%. Given the size of our administration, the impact in terms of regulating immigration would be massive, as it would reduce the problem of unwanted immigration and Islamism by 1% globally. This is the kind of policy that could really help to fight the influence of people who refuse our values. »

The Minister said, "Honestly, I really don't care about restricting immigration rights to those who know and accept our values and culture. What I care about is making budget savings that could increase the chances of our party winning the next election. How much would the policy cost us? »

The adviser replied: "According to my calculations, the State would save about 100 million euros in total with this policy, because it would significantly reduce the wage bill. »

The minister decided to implement the policy. He effectively reduced undesirable immigration and Islamism in France by 80%, thus contributing to a significant decrease of 1% in the world. But more importantly for the minister, he also achieved a budget saving of 100 million euros which directly benefited his party.

## Main multiple regression models on aggregated data

The first model (left) contains the main effects only (Issue, Decision type, commitment to issue), the second model (right) contains the main effects and the two way interactions:

**Exp. 1-4: Multiple linear regression on aggregated data**

<i>Predictors</i>	Commendable				Commendable			
	$\beta$	95% CI	Statistic	<i>p</i>	$\beta$	95% CI	Statistic	<i>p</i>
IssueExp 2, sex equality	-0.17	-0.29,-0.05	-2.83	0.00	-0.27	-0.44,-0.10	-3.10	0.00
IssueExp 3, authority	-0.52	-0.65,-0.39	-7.75	0.00	-0.88	-1.07,-0.70	-9.38	0.00
IssueExp 4, immigration	-0.63	-0.76,-0.50	-9.59	0.00	-0.93	-1.11,-0.75	-10.21	0.00
commitment	0.17	0.12,0.22	6.96	0.00	0.09	-0.03,0.20	1.48	0.14
ConditionSelfish/High	-0.54	-0.62,-0.46	-12.93	0.00	-0.79	-0.95,-0.64	-10.04	0.00
ConditionSelfish/High:IssueExp 2, sex equality					0.01	-0.20,0.23	0.13	0.90
ConditionSelfish/High:IssueExp 3, authority					0.57	0.34,0.81	4.74	0.00
ConditionSelfish/High:IssueExp 4, immigration					0.50	0.27,0.74	4.20	0.00
IssueExp 2, sex equality:commitment					0.08	-0.06,0.22	1.10	0.27
IssueExp 3, authority:commitment					0.21	0.07,0.36	2.83	0.00
IssueExp 4, immigration:commitment					0.32	0.19,0.46	4.70	0.00
ConditionSelfish/High:commitment					-0.18	-0.27,-0.09	-4.09	0.00
Marginal R <sup>2</sup> / Conditional R <sup>2</sup>	0.206 / 0.261				0.254 / 0.317			

NB:

Levels defined as baselines of each factor:

Issue: environment

Condition: Altruistic intent/Low efficiency

Selfish/High means Selfish intent/High efficiency

Commitment means moral commitment

Main multiple regression models on aggregated data, data from 1st condition only

**Exp. 1-4: Multiple linear regression on aggregated data, data from 1st condition only**

<i>Predictors</i>	Commendable				Commendable			
	$\beta$	95% CI	Statistic	<i>p</i>	$\beta$	95% CI	Statistic	<i>p</i>
IssueExp 4, immigration	-0.61	-0.78,-0.43	-6.89	0.00	-0.83	-1.09,-0.58	-6.39	0.00
commitment	0.23	0.16,0.30	6.87	0.00	0.15	-0.00,0.31	1.91	0.06
Conditionselfish_high	-0.56	-0.67,-0.44	-9.43	0.00	-0.77	-0.99,-0.54	-6.75	0.00
IssueExp 2, sex equality	-0.14	-0.30,0.02	-1.67	0.09	-0.13	-0.36,0.10	-1.09	0.27
IssueExp 3, authority	-0.40	-0.57,-0.22	-4.48	0.00	-0.79	-1.04,-0.53	-6.07	0.00
Conditionselfish_high:IssueExp 3, authority					0.61	0.27,0.95	3.51	0.00
Conditionselfish_high:IssueExp 4, immigration					0.40	0.06,0.74	2.32	0.02
IssueExp 2, sex equality:commitment					0.02	-0.16,0.21	0.25	0.81
IssueExp 3, authority:commitment					0.19	-0.00,0.39	1.93	0.05
Conditionselfish_high:commitment					-0.17	-0.30,-0.05	-2.68	0.01
Conditionselfish_high:IssueExp 2, sex equality					-0.18	-0.49,0.13	-1.13	0.26
IssueExp 4, immigration:commitment					0.32	0.14,0.50	3.48	0.00
Marginal R <sup>2</sup> / Conditional R <sup>2</sup>	0.237 / 0.365				0.292 / 0.415			

Additional simple regressions models of commendability on moral commitment:

**Experiment 1: Protecting the environment**

<i>Predictors</i>	<b>Commendable (Altruistic intent/Low efficiency)</b>			<b>Commendable (Selfish intent/High efficiency)</b>		
	$\beta$	<i>95% CI</i>	<i>p</i>	$\beta$	<i>95% CI</i>	<i>p</i>
commitment	0.27	0.15,0.39	0.00	-0.17	-0.30,-0.05	0.01
$R^2 / R^2$ adjusted	0.073 / 0.069			0.030 / 0.026		

**Experiment 2: Pursuing sex equality in the workplace**

<i>Predictors</i>	<b>Commendable (Altruistic intent/Low efficiency)</b>			<b>Commendable (Selfish intent/High efficiency)</b>		
	$\beta$	<i>95% CI</i>	<i>p</i>	$\beta$	<i>95% CI</i>	<i>p</i>
commitment	0.30	0.17,0.43	0.00	-0.08	-0.22,0.06	0.24
$R^2 / R^2$ adjusted	0.089 / 0.085			0.007 / 0.002		

**Experiment 3: Promoting France’s diplomatic authority in the world**

<i>Predictors</i>	<b>Commendable (Altruistic intent/Low efficiency)</b>			<b>Commendable (Selfish intent/High efficiency)</b>		
	$\beta$	<i>95% CI</i>	<i>p</i>	$\beta$	<i>95% CI</i>	<i>p</i>

<i>Predictors</i>	$\beta$	<i>95% CI</i>	<i>p</i>	$\beta$	<i>95% CI</i>	<i>p</i>
commitment	0.24	0.11,0.38	0.00	0.17	0.03,0.31	0.01
$R^2 / R^2$ adjusted	0.060 / 0.055			0.029 / 0.024		

#### **Experiment 4: Regulating immigration**

	<b>Commendable (Altruistic intent/Low efficiency)</b>			<b>Commendable (Selfish intent/High efficiency)</b>		
<i>Predictors</i>	$\beta$	<i>95% CI</i>	<i>p</i>	$\beta$	<i>95% CI</i>	<i>p</i>
commitment	0.37	0.24,0.50	0.00	0.36	0.23,0.49	0.00
$R^2 / R^2$ adjusted	0.139 / 0.134			0.129 / 0.124		

#### **Experiment 5**

Main multiple regression models of support for policy judgments

The first model (left) contains the main effects only (Efficiency, Intent, Decision maker identity, and commitment to issue), the second model (right) contains the main effects and the two way interactions:

## Experiment 5

<i>Predictors</i>	<b>Support</b>				<b>Support</b>			
	$\beta$	95% CI	Statistic	<i>p</i>	$\beta$	95% CI	Statistic	<i>p</i>
EfficiencyHigh	0.66	0.53,0.79	10.14	0.00	0.51	0.26,0.76	4.00	0.00
IntentAltruistic	0.17	0.01,0.33	2.09	0.04	0.43	0.16,0.70	3.16	0.00
IntentSelfish	-0.47	-0.62,-0.31	-5.78	0.00	-0.29	-0.55,-0.02	-2.13	0.03
ActorMin	-0.48	-0.61,-0.35	-7.38	0.00	-0.64	-0.90,-0.39	-5.02	0.00
commitment_scale	0.26	0.20,0.33	7.92	0.00	0.25	0.11,0.39	3.60	0.00
EfficiencyHigh:IntentAltruistic					-0.39	-0.70,-0.08	-2.49	0.01
EfficiencyHigh:IntentSelfish					-0.07	-0.38,0.25	-0.41	0.68
EfficiencyHigh:ActorMin					0.61	0.36,0.86	4.76	0.00
EfficiencyHigh:commitment_scale					-0.12	-0.25,0.01	-1.83	0.07
IntentAltruistic:ActorMin					-0.13	-0.44,0.18	-0.80	0.42
IntentSelfish:ActorMin					-0.30	-0.61,0.02	-1.86	0.06
IntentAltruistic:commitment_scale					0.12	-0.02,0.27	1.67	0.10
IntentSelfish:commitment_scale					0.03	-0.14,0.19	0.31	0.75
ActorMin:commitment_scale					0.03	-0.10,0.16	0.51	0.61
$R^2$ / $R^2$ adjusted	0.302 / 0.297				0.341 / 0.327			

NB:

Levels defined as baselines of each factor:

Efficiency: Low

Intent: Unspecified

Commitment\_scale means moral commitment

Additional simple regression models of support on moral commitment

### Low efficiency, CEO

<i>Predictors</i>	<b>Support (Altruistic)</b>			<b>Support (No intent)</b>			<b>Support (Selfish)</b>		
	$\beta$	95% CI	<i>p</i>	$\beta$	95% CI	<i>p</i>	$\beta$	95% CI	<i>p</i>
commitment_scale	0.53	0.29,0.77	0.00	0.39	0.15,0.64	0.00	0.29	0.03,0.56	0.03
R <sup>2</sup> / R <sup>2</sup> adjusted	0.283 / 0.269			0.156 / 0.140			0.087 / 0.070		

### High efficiency, CEO

<i>Predictors</i>	<b>Support (Altruistic)</b>			<b>Support (No intent)</b>			<b>Support (Selfish)</b>		
	$\beta$	95% CI	<i>p</i>	$\beta$	95% CI	<i>p</i>	$\beta$	95% CI	<i>p</i>
commitment_scale	0.43	0.18,0.68	0.00	0.30	0.04,0.56	0.03	0.04	-0.23,0.31	0.76
R <sup>2</sup> / R <sup>2</sup> adjusted	0.186 / 0.170			0.090 / 0.073			0.002 / -0.017		

### Low efficiency, Minister



<i>Predictors</i>	<b>Support (Altruistic)</b>			<b>Support (No intent)</b>			<b>Support (Selfish)</b>		
	$\beta$	95% CI	<i>p</i>	$\beta$	95% CI	<i>p</i>	$\beta$	95% CI	<i>p</i>
commitment_scale	0.37	0.11,0.62	0.01	0.26	-0.01,0.52	0.05	0.35	0.09,0.61	0.01
R <sup>2</sup> / R <sup>2</sup> adjusted	0.135 / 0.119			0.067 / 0.049			0.125 / 0.108		

### High efficiency, Minister

<i>Predictors</i>	<b>Support (Altruistic)</b>			<b>Support (No intent)</b>			<b>Support (Selfish)</b>		
	$\beta$	95% CI	<i>p</i>	$\beta$	95% CI	<i>p</i>	$\beta$	95% CI	<i>p</i>
commitment_scale	0.49	0.25,0.73	0.00	0.29	0.03,0.56	0.03	0.08	-0.19,0.36	0.54
R <sup>2</sup> / R <sup>2</sup> adjusted	0.238 / 0.224			0.086 / 0.069			0.007 / -0.011		

## Chapter 5: Conclusion and future avenues

### **Conclusion**

Facts don't care about our feelings, so our attitudes and choices in the policy domain should be guided by assessments of risks and benefits based on the best scientific evidence available. A pragmatic moral mindset would command to follow the experts on the issues on which the science is clear—e.g. that human activities make the planet warmer, that commercialized GMOs are safe to consume, or that there is no massive voter fraud in the US—and to resist the temptation of holding extreme opinions on the issues on which the evidence is equivocal. Seeking and deciding based on reliable knowledge is a moral responsibility for policy makers given they are in a unique position to increase collective welfare or to waste society's resources and money. But developing an *ethos* of truth-seeking and pragmatism is also morally important for ordinary citizens, whose votes and beliefs can end up having significant influence at an aggregate level.

### **Misleading intuitions**

The truth of the matter on complex issues is never directly observable, and can only be arrived at only through systematic examination of the evidence, following the painstaking, laborious, and counter-intuitive methods of science. However, the human mind is populated by numerous cognitive inclinations, some of which may have evolved by natural selection, that make it difficult to access reliable information of policy-relevant facts, and to find the motivation to let that information guide our choices when it is made available to us.

As argued in Chapter 2, a first important hindrance is that people lack knowledge of what makes scientific expertise vastly more trustworthy than their intuitions, and those of their peers (Levy, 2017; Mercier & Heintz, 2013; Morin et al., 2020; Pinker, 2018, 2019). We acquire many of our beliefs through the word of others, to which we are generally receptive when cues of benevolence (“why would they try to deceive us?”) or affiliation (“they’re like us”) are present—whether they are family, friends, colleagues, or congenial journalists or politicians—or when an opinion appears particularly widespread (Mercier, 2017; Origg, 2004; Sperber et al. 2010; Sperber, 2009). However, we are ill-equipped to see that those others we trust generally lack competence on the issues of which they speak. The crushing majority of people have little knowledge of the methodological requirements that would make an empirical belief truly reliable—such as the experimental method, controlled statistical inference, peer-review, meta-analysis, etc.. What most people take as “evidence” is often little more than salient anecdotes and images of dubious representativeness (Pinker 2018, 2019). They also spontaneously tend to trust their own intuitions, falling prey to an illusion of understanding on most complex issues (Fernbach et al., 2013, 2019; Keil, 2003). Considered from an evolutionary perspective, our lack of intuitions on what grounds reliable expertise should not come as a surprise. Our minds evolved in traditional, tight-knit communities in which nobody possessed advanced theoretical expertise, and at a time when we had only limited technical control over nature, so no selection pressures were exerted by the consequences of complex, large-scale economic or energy policies.

Most of the claims we are exposed to by the sources we trust are semantically vague and piecemeal, containing no quantitative specification of the phenomena they portray (“GMOs are dangerous”; “Guns kill people”; “Vaccines can cause cancer”). But even when information is precise and accurate at some point of the communication chain (e.g. in a newspaper article we read, or a presentation by an expert we hear), it is selectively attended, remembered, reconstructed, and retransmitted to others in ways that follow the natural slopes

of our intuitions. While some of those content biases likely exist because having them would have been useful in the ancestral environment, they sometimes make us form and convey representations of risks and benefits in ways that are unwarranted by the facts.

As argued in Chapter 2, certain types of threat-relevant information are very attractive to the mind. While an over-sensitivity to threat makes a lot of adaptive sense from an individual standpoint (the only one relevant for natural selection), it may contribute to focus disproportionate attention on policies' and technologies' risks only, and to have those risks exaggerated in communication (Moussaïd, Brighton, & Gaissmaier, 2015). This may be the case for instance of information eliciting fear of contamination, or disgust. Plausibly anchored into a pathogen-avoidance system (Tybur et al., 2013), those instincts may play a role in fueling fear of nuclear energy, with radiation being intuitively represented as a potentially contaminating pathogen (Hacquin et al., 2020). Triggered by violation of our essentialist biological intuitions (Gelman, 2004), emotions of disgust may also contribute to the public's distrust of GMO foods, despite the well-established safety of those currently commercialized, with people thinking that it is against the proper order of things, and depraved, to modify a "natural" species' genome (Blancke et al. 2015).

Information pertaining to social threats is particularly attractive too. Conspiracy theories, for instance, may surf on evolved equipment for detecting coordinated action from hostile coalitions, which plausibly were an evolutionary recurrent challenge (van Prooijen & van Vugt, 2018). Whipped up by knowledge of authentic conspiracies and public lies, which of course historically existed<sup>7</sup>, our instinct to over-detect hostile intent leads to often disproportionate suspicions about industrialists and scientists' intentions and activities, undermining public trust in them (Barclay & Benard, 2020; Fiske & Dupree, 2014).

---

<sup>7</sup> Recent examples of true conspiracies likely to have influenced Western opinion include for instance, in the scientific domain, the tobacco industry's efforts to conceal the link between smoking and lung cancer (Brandt, 2012), and in international relations, G. W. Bush's administration's lies about Iraq possessing mass destruction weapons to justify the 2003 invasion of the country (Draper, 2020).

Similarly, the studies presented in Chapter 4 suggest that information relevant to political actors' intentions, which arguably are of little importance from a pragmatic point of view, significantly weigh on moral judgment. By contrast, information pertaining to policies' efficacy and financial cost expressed in numeric format, which should be the crux of the matter of political deliberation, only weighed little in people's judgments. We also found that insensitivity to cost/benefit calculations was all the greater as participants were morally committed to the issue the policy meant to tackle, corroborating the notion that moral convictions make us more deontological (Atran, 2010, 2016; Tetlock, 2003; Skitka & Mullen, 2002). If they transfer to the real world, the combination of those effects may lead people to support inefficient policies driven by noble intentions over vastly more efficient ones if they appear to be guided by selfishness, and to pass along political information in ways that prioritize values and symbols but neglect impact.

Findings from Chapter 4 are consistent with a view of intuitive moral judgment as wired for attending to others' moral character, intentions, and principles, more than for pragmatism. Accumulating evidence suggests our moral minds evolved primarily to see and judge others as potential cooperation partners (Baumard, André, Sperber 2013; Tooby & Cosmides, 1990; Uhlmann, Pizarro, Diermeier, 2015; Sperber & Baumard, 2012). We are comparatively unequipped to represent and react to information on policies' efficiency, particularly when it is expressed in abstract, numeric format. Large-scale public policies and modern organizations, by definition, didn't exist during most of our evolution, so they could not constitute selection pressures. Likewise, numeric systems are a recent cultural innovation (Dehaene, 2009). So, despite being a concise format in which to convey policy-relevant information (which is why it is privileged by economists and other experts), the human mind likely lacks the intuitive capacities to represent it, and be swayed by it (Kirk, 2016).

## Political polarization of factual beliefs

Sometimes people are wrong or misguided in ways that are roughly universal. But all too often, the direction in which they err from the truth can be predicted from their politics. For instance, liberals tend to believe that humans make the planet warmer, that nuclear power is not a safe and efficient source of energy, that our societies are systematically racist and sexist, while conservatives are disproportionately more likely to believe the opposite is true (Kahan et al., 2011a, 2011b, 2016; Nuclear Energy Agency, 2010; Van Bavel & Pereira, 2018). The views people have on otherwise causally unrelated issues tend to cluster together, delineating political subcultures.

This polarization of factual beliefs starts when differences in geographical residence, class, or mere historical traditions, expose people to different claims from the onset of their lives. Later, people's preference for homophily pushes them (often unconsciously) to surround themselves with people who share their background beliefs, personality traits, and views of what society should look like (Alford et al. 2011; McPherson, Smith-Lovin & Cook, 2001). It feels good to hang out and "connect" with people who share very similar view to one's own. So, given all the options we have in mass modern societies and cities, why take the pain of hanging out with folks who (we think) are obviously so wrong, and who sound like they are constantly trying to annoy? (Note, incidentally, that the ideological diversity of the contents and sources people get exposed to seems *greater* on social media than in real life, according to recent work: Acerbi, 2019; Gentzkow & Shapiro, 2011; Fletcher & Nielsen, 2017).

And when they do encounter cross-cutting information from politically uncongenial people and media, individuals tend to apply higher standards of skepticism to it than information congruent to their expectations and convictions, such that they rarely get serious

chances of correcting their worldview—whether this stems from culpable motivated reasoning (Kahan 2013, 2016; Kunda 1990; Van Bavel & Pereira, 2018), or reflects subjectively optimal Bayesian updating (Baratgin & Politzer, 2006, 2010; Pennycook, 2020; Tappin, Pennycook, Rand, 2020).

Those like-minded others we spend most of our lives with are also unlikely to correct one's beliefs with new evidence and challenging arguments because what they tell us and share with us by and large comforts the narratives they we, together, subscribe to. The studies presented in Chapter 3 documented a tendency for partisan selective communication in the form of headlines sharing online, and suggested that the more partisan or morally convinced individuals were, the more they were prone to it. This propensity to selective sharing on politics seemed robust, as it was little affected by who the audience is (politically congenial *vs.* uncongenial), whether one's reputation is at stake (anonymous *vs.* personal account), and whether or not people are being warned of the possibility that they may be politically biased in the first place. Probably the most parsimonious way of accounting for partisan selective sharing and its robustness across contextual manipulations—an hypothesis corroborated by investigation of the reasons for sharing in Chapter 3—is that participants share what they think is *true*—despite having different prior beliefs on many issues. This means that misinformation and “fake news” may travel far and wide among networks of like-minded peers if they surf on shared preconceived notions. (Truth be told, however, studies on Facebook data suggest that people share very little fake news overall: Guess, Nagler & Tucker, 2019b).

Claims that people live in “echo chambers” or “house of mirrors” are certainly overblown (Guess et al., 2019a; Mercier, 2020). But most people develop and spend their lives in networks that are characterized by at least some level of political bias with respect to the type of narratives and information sources that dominate their political conversations and news diets. Knowing this, news producers and political leaders are financially and electorally

incentivized to select and format what they publish and say in ways that exploit their readership's prior beliefs. This will not contribute to make people's views converge on the factual aspects of society's burning issues, and what could be done to effectively tackle them.

## **Social motivations to communicate and believe**

What is more, the more societies become divided into rival political factions, the more certain claims are made to signal one's political membership, in ways un conducive to truth and nuance. In the studies presented in Chapter 3, for instance, the strongest predictor of participants' sharing decisions was a headline's "usefulness for [one's] political commitments". Given the polarized nature of the topics covered (abortion, racial equality, gun control and sex equality), this likely included motives such as disparaging political opponents and expressing support for one's coalition, so as to signal how one stands politically. In a similar vein, recent studies on Twitter data also suggest that signaling one's political commitments (along with trying to cause outrage or discredit the out-party) may be one of the primary reasons why some people share fake news online (Osmundsen et al., 2020).

But to complicate things a little further, in contexts of political conflict, not just verbal claims and shared articles but also *beliefs*, including the most "sincere" ones, may be guided by (unconscious) signaling strategies. Despite any logical connection between facts and values (Hume, 1739/2000), people intuitively attribute normative "implications" to facts, and cherry-pick the facts that best corroborate their political interests. In parallel, we also have social signaling instincts, inherited from an evolution of group life and group conflict, which, if they are to be most persuasive, should be expected to influence beliefs (and not just say, verbal claims or clothing; von Hippel & Trivers, 2011; Mercier, 2017, 2020). In those conditions, certain factual beliefs pertaining to controversial political issues may function as



markers of coalitional membership, i.e. as signs of the kind of people individuals want to be seen as, and who they are fighting against (Kahan et al. 2011a; Levy, 2017; Mercier, 2017, 2020; Pinker 2019; Tooby & Cosmides, 2010). And as Pinker (2018:701) nicely puts it, commenting on Kahan’s work on “cultural cognition”, incentives to hold certain beliefs as signals should *not* be expected to incentivize people to cultivate virtues of nuance and truth-seeking:

[A] person’s opinions on climate change or evolution are astronomically unlikely to make a difference to the world at large. But they make an enormous difference to the respect the person commands in his or her social circle. To express the wrong opinion on a politicized issue can make one an oddball at best—someone who “doesn’t get it”—and a traitor at worst. The pressure to conform becomes all the greater as people live and work with others who are like them and as academic, business, or religious cliques brand themselves with left-wing or right-wing causes. For pundits and politicians with a reputation for championing their faction, coming out on the wrong side of an issue would be career suicide. Given these payoffs, endorsing a belief that hasn’t passed muster with science and fact-checking isn’t so irrational after all—at least, not by the criterion of the immediate effects on the believer. The effects on the society and planet are another matter.”

### **From diverging priors to political hostility**

Finally, contexts of political polarization, beyond incentivizing people to be less nuanced and more wrong than they could, can be harmful in another way: through their fostering affective polarization, i.e. political hostility between groups (Haidt 2012; Iyengar et

al., 2018). When partisans on one side are informationally isolated (because of who they hang out with, where they get their news from, etc.), and have low reputational incentives to develop more nuanced views, the impression they leave on the other side is not just that of incompetence, but of hypocrisy, bad faith, or malevolence. When this impression is reciprocal, partisans on both sides feel they are dealing with uncooperative, alien, immoral individuals, leaving room for automatic feelings of hostility to creep in—hence the notion of *affective* polarization (or “political sectarianism”: Finkel et al., 2020). As a wealth of psychological research has noted, the problem of affective polarization has become fairly serious between Democrats and Republicans in the US, where hostility towards the other side has come to overtake affection for one’s co-partisans (Finkel et al., 2020; Pew Research Center, 2014). According to recent polls, for instance, on a feeling thermometer anchored in cold (0°) to warm (100°), while sentiments towards in-party members has remained around 70°-75° for several decades, affect towards out-party members has fallen from 48° in the 1970s to only 20° in 2020 (American National Election Study data, reported in Iyengar et al., 2018).

Admittedly, France has been relatively preserved by the scourge of affective polarization, and academics are incentivized to exaggerate the problem to attract readers (Mercier, 2020). But international comparisons suggest the phenomenon may be on the rise across the world, from Brazil to Turkey to Kenya (Carothers & O’Donohue, 2019). Any increase, even small, in mutual hostility between members of competing political tribes must be taken seriously, as it can worsen our (already fragile) ability to open ourselves to alternative points of view, converge on the truth, and avoid legislative gridlocks. In one word, to be collectively intelligent.

\*\*

The remaining pages of this thesis propose some future avenues for the empirical work presented in Chapters 3 and 4, and end with a quick summary of the political psychology project I will be working on at Michael Bang Petersen's lab in Denmark during my postdoc, starting April 2021.

### **Moral convictions and selective communication: future avenues**

In “Moral conviction predicts sharing preference for politically congruent headlines” (Chapter 3), participants were not given the possibility of adding a comment, in particular a critical comment, when sharing a headline. This may seem like a design detail but I realized that absence of a comment and share feature, as Twitter allows to do when making a “quote tweet”, might have de-incentivized sharing of incongruent content—content participants would have wanted to denounce if they could. When publishing something online, the default assumption in one's audience is indeed likely to be that, in the absence of any specification to the contrary, people *endorse* political content they share, at least when said content is in the vicinity of what the person might believe—which of course, will depend on the person's prior beliefs and ideology. So, extreme content put aside, it seems reasonable to assume that individuals spontaneously see a share as implying some form of endorsement, and may refrain from sharing an extreme or incongruent headline if they don't have the possibility of specifying what they meant when sharing it.

In fairness, I know of no experimental study demonstrating that the default attitude is to interpret sharing of believable content as endorsement, a hypothesis that definitely deserves its own experimental study. I must rely on intuitions, which I acquired in part through discussions with colleagues. At any rate, if the hypothesis that sharing of plausible content

tends to be interpreted by one's audiences as endorsement, I suspect that, in our experiments on headlines sharing, absence of a comment and share feature may have artificially decreased sharing of incongruent headlines in participants afraid of coming across as subscribing to the narrative upheld by the rival political side. This risk was likely present not only with respect to the real news items—which, since they were true, were quite believable—but also to the false news items—which were selected so as not to be perceived as being too implausible. An important next step in this series of experiments (most likely in another paper) would be to replicate our willingness to share studies on true and false headlines while allowing participants to leave a comment. The question could look like Figure 1a below. I would expect this feature to *increase* participants' sharing preference for incongruent content, at least on high importance topics (if not regardless of issue importance) (see Figure 1b below for predicted data).

a

---

Would you share this news story to friends or people you like? Note that you can leave a comment explaining your decision.

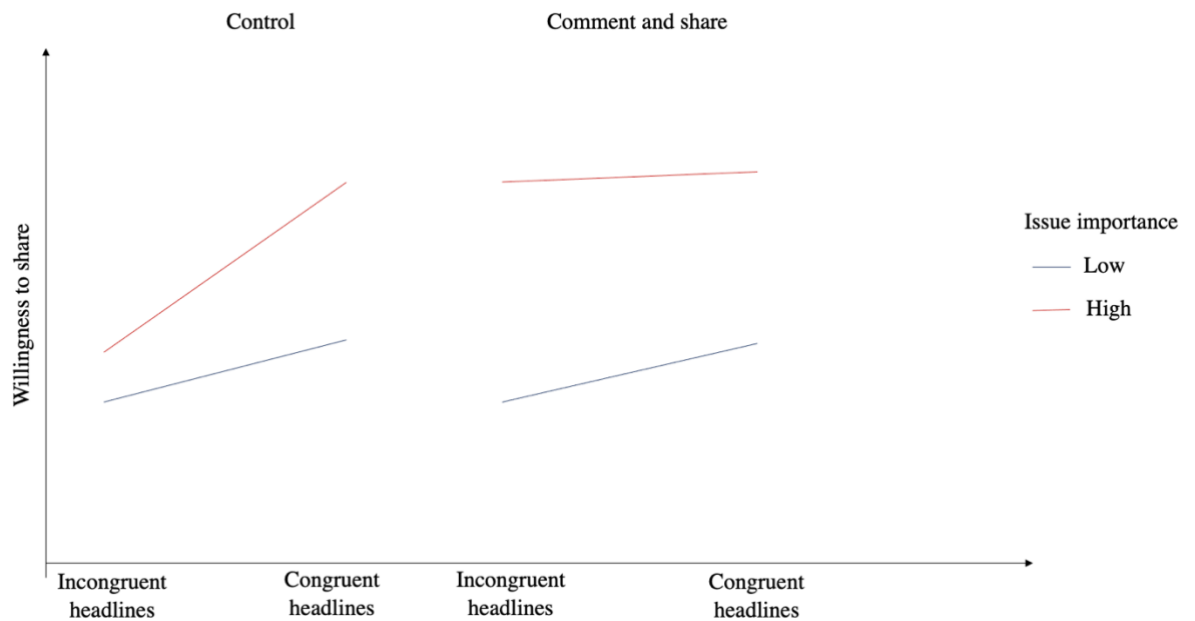
Not Share  Share

---

Comment:

---

b



**Figure 1.** a: Possible formulation of a willingness to share question that would allow participants to comment on shared content. b: Predicted willingness to share political headlines (whether true or false) as a function of news congruence, issue importance, and possibility *vs.* impossibility of leaving a comment when sharing.

Another direction for the studies on willingness to share political headlines would be to test new types of intervention messages. Studies 7 and 8 of the paper “Moral convictions predicts sharing preference for politically congruent headlines” (Chapter 3) tested whether reminding participants of their tendency to give more weight to information that fits their goals and values than information that doesn’t could mitigate their sharing preference for politically congruent headlines<sup>8</sup>. Overall, we found that this intervention had little to no

<sup>8</sup> The instructions were the following:

Previous research has shown that the issue(s) you judge as having absolute importance are the ones on which you are *most likely to be politically biased*. In your case, they are:

>>> [Issues rated as being “of absolute importance” displayed] <<<

Typically, political bias causes one to significantly *favor information that fits one’s goals and values, and to disregard information that doesn’t*.

effect—a small reducing effect of true, congruent headlines on high importance issues at most. This result may be parsimoniously explained by the fact that many news we categorized as “politically congruent” actually fit partisans’ priors. Since people tend to spontaneously trust their beliefs, it is not incredibly surprising, with hindsight, that an instruction telling them to distrust those beliefs without any argument other than a few scientific references did not change their intuitions.

As an alternative to trying to reduce the sharing difference between congruent and incongruent headlines, I and Sacha Altay thought of an intervention that would target people’s propensity to share “fake news” in general. To do so, we thought that a message that would highlight the *reputational costs* of sharing false information could work. Altay & Mercier (2020) showed that while people may be reputationally incentivized to pass along to their networks information that is relevant—e.g. shocking or unbelievable—there is a trade-off between relevance and accuracy in how they are seen by others. In their studies, the authors found that sharing claims that appear relevant but which others discover were false has the effect of lowering trust in the sender in ways that are difficult to compensate in the future. Inspired from those findings, S. Altay and I would be curious to see if an intervention message like on Figure 2 below, highlighting the reputational damages of sharing fake news, could reduce people’s willingness to share them compared to a control condition with no message:

---

Source:  
Kahan, D. (2016). The Politically Motivated Reasoning Paradigm, *Emerging Trends in the Social and Behavioral Sciences*, 1-16.

---

Here is something you should know:

Psychological research has shown that **most people around you highly value information accuracy 🧐 and do not share "fake news" online.**

... This is because spreading "fake news" online can **damage your reputation, and make a fool of yourself.** 🤡

Studies have also found that this risk is highest regarding news stories that touch on issues you see as being of **absolute moral importance to you** (😡, 😞 or 😠).

In case you've forgotten, here are the issues you rated as being of absolute importance to you — and on which **you're therefore most likely to forget about critical thinking:**

>>> Racial equality, Gender equality <<<

Source:

Altay, S. & Mercier, H. (2020). Relevance Is Socially Rewarded, But Not at the Price of Accuracy. *Evolutionary Psychology*.

Guess, A., Nagler, J. and Tucker, J.. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science advances* 5(1)

Pennycook et al. (2020). Understanding and reducing the spread of misinformation online. Preprint.

**Figure 2:** Projected intervention message meant to remind participants of the reputational costs of sharing “fake news”.

## **Intentions and efficiency in judgments of policies: future avenues**

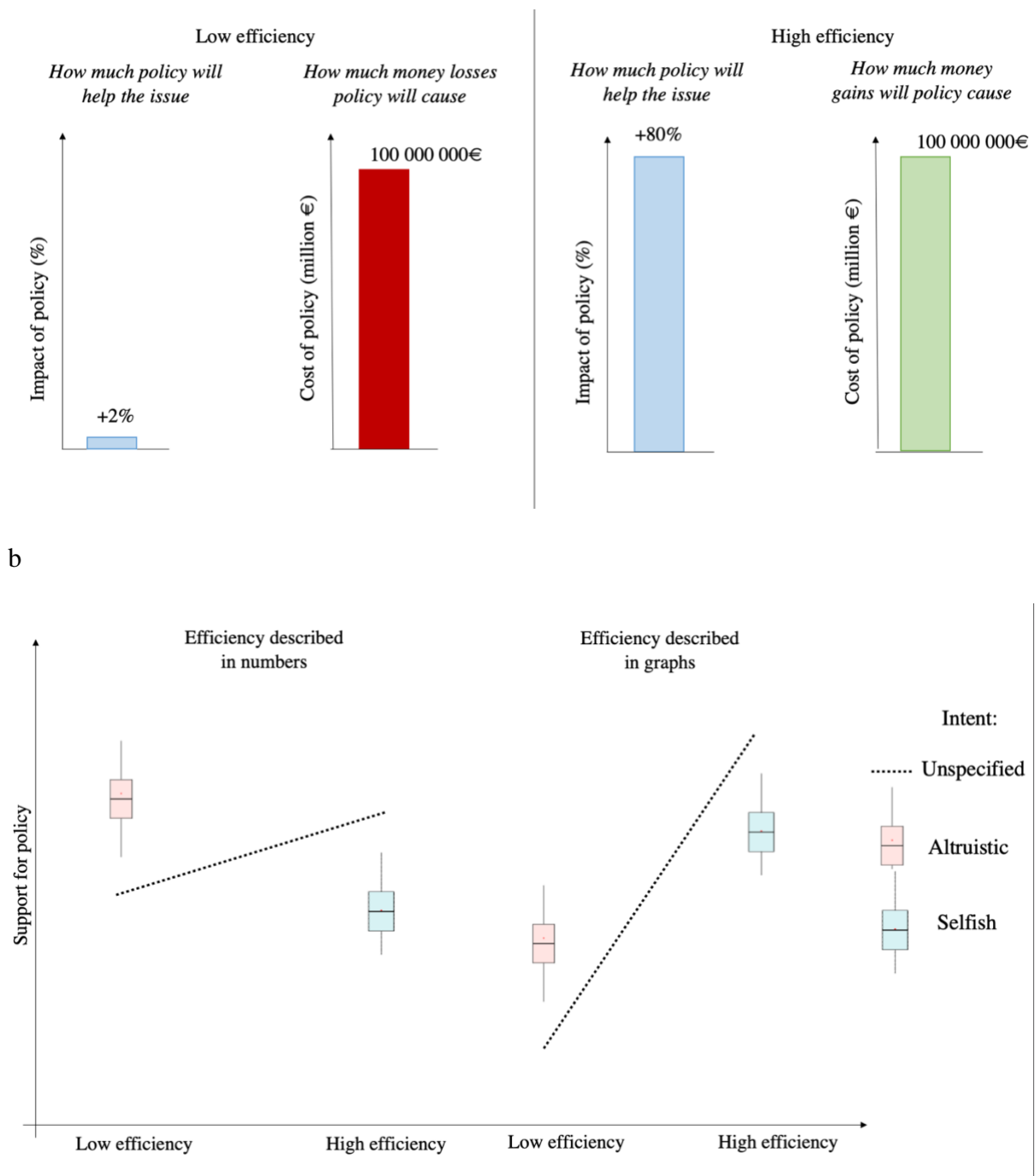
Experiments 1-5 of “Intentions matter a lot, and efficiency little, in folk judgments of policy decisions” (Chapter 4) found that participants were little responsive to huge differences in efficiency when they were expressed in numerical format. As argued in the discussion however, we know that differences between quantities are not intuitively grasped when expressed in numbers (in fact, one of the advantages of having a numerical notation system is precisely to be able to synthetically represent information without taking up space). An image is worth a thousand words, so visualization is key to represent numbers and proportions. We

scientists are constantly reminded of this fact when examining our own experimental data and drafting our papers.

An interesting next step could thus be to compare participants' sensitivity to differences in efficiency when they are cashed out in graphical format *vs.* in numbers. If this contrast in formatting increases people's sensitivity to differences in efficiency by a large margin, then it makes for an interesting intervention which could be used to make people more consequentialist in their judgments of policies. Figure 3a shows a possible graphic way of conveying information on policy efficiency (both low and high) in addition to the text vignettes we used in the paper. Figure 3b shows projected data that would correspond to the scenario in which adding information on the policy's efficiency in graphic format to the vignette does indeed increase participants' sensitivity to efficiency (compared to when efficiency is described in numbers only in the vignette). If, however, providing information on policy efficiency in graphic format does *not* increase sensitivity to efficiency to a degree that would be politically meaningful, then the claim that "people show little responsiveness" to differences in efficiency, which we are currently making, would garner further support.

a





**Figure 3.** a: Possible way of conveying information on policy efficiency in graphic format, to complement the text vignette. b: Predictions of support for policy judgments in a design varying whether the policy’s efficiency is presented in numeric format in vignettes only (left panel) vs. in graphic format accompanying the vignettes (right

panel). Here I assume that this intervention would succeed at increasing participants' sensitivity to efficiency.

In “Intentions matter a lot, and efficiency little, in folk judgments of policy decisions” (Chapter 4), our findings that folk moral judgments showed little responsiveness to efficiency, and significant sensitivity to intentions, relied entirely on judgments recorded using continuous scales (commendability of policy in Experiments 1-4, and support for policy in Experiment 5). Those findings allow us to expect that over the course of cultural transmission—as people pass along political information to one another in discussions—information about policies' efficiency will tend to be forgotten or omitted, and that information on actor's intentions will comparatively be better memorized and more reliably transmitted. Demonstrating this hypothesis directly, however, would require using the method of transmission chains (Mesoudi & Whiten 2008; Aarøe & Petersen 2018). Transmission chains experiments ask participants to recall some material after they have read or heard it and the material has disappeared. The operation is repeated multiple times or “generations” through consecutive online surveys in which each participant inherits the material recollected by the previous participant. By measuring the rate at which certain types of information get faithfully conserved or exaggerated, or on the contrary lost in transmission, using automated content analysis, transmission chains are a powerful way of examining the mind's biases in favor of certain types of information. I would be excited to use this method to examine whether, as I would expect based on our findings, information on policies' efficiency tends to disappear at a higher rate than information about policy makers' intentions (and in particular, information about negative intentions, if it is true that negative information is on average more attention-grabbing than positive information; e.g. Moussaïd, Brighton, & Gaissmaier, 2015).

## Exploring the functions of demonizing political opponents

In December 2020, I was honored to get a postdoctoral position offered by Professor Michael Bang Petersen and the Department of Political Science at Aarhus University in Denmark. The position is linked to a project called Research on Online Political Hostility (ROPH) and is funded by the Carlsberg Foundation (yes, a non-profit that belongs to the company that produces beer). The goal of the ROPH project is to investigate the risk factors facilitating “ordinary” political hostility, the hostility shown by people like you and me when they talk about political topics on which they have strong convictions.

As a contribution to their research program, I proposed to focus on partisan individuals’ propensity to ascribe *negative traits and intentions* to political opponents—such as selfishness, hypocrisy, depravedness or hatred—as an operationalization of political hostility, and to explore the adaptive functions this “demonization” of opponents may play in contexts of political polarization around controversial issues. I came up with several experimental designs meant to test three evolutionary hypotheses on the role played by negative attributions in intergroup conflict. The first evolutionary hypothesis (H1) is inspired from the ROPH team’s work and is that negative attributions emerge more easily in online than real life contexts because people lack cues to the physical presence of their addresses online, in particular facial cues, which would undermine their emotion regulation capacities. I derived the other two hypotheses from my reading of coalitional psychology theory (Mercier, 2020; Petersen, Osmundsen, Tooby, 2020; Tooby & Cosmides 2010). The second hypothesis (H2) is that attributions of negative traits and intentions to one’s political opponents might be a way of triggering group coordination and cooperation, through the sending of an alarm call about an outgroup threat, in order to ultimately benefit from the protection of one’s allies. The third hypothesis (H3) is that negative traits ascriptions to political opponents may be used as a way

of signaling one's devotion to one's group: of signaling that one really cares to belong to *this* group, not other groups, by "burning bridges" with rival groups by demonizing them (Mercier 2020).

To take but one example, I proposed to test H1 and H3 using the following design (see Figure 4a below). We would start by recruiting politically progressive students on campus who would be interested to become part of a student political association, for instance a feminist association (this might require to find some way of incentivizing participants to become members, e.g. through course credit). In Western universities, many students mobilize for sex equality and feminism, and against patriarchy or "men's rights activists"<sup>9</sup>, the latter representing out-groups from their perspective. We would then tell participants that their fit with the student association's values would have to be assessed by its recruitment team using data collected on an online questionnaire (e.g. a Qualtrics questionnaire). The questionnaire would in fact be bogus, but believable. After having taken the questionnaire, the recruitment would give feedback to the participant applicants. We would manipulate whether this feedback is highly inclusive and encouraging ("100% fit") vs. not so inclusive and encouraging ("50% fit"). Drawing on social signaling and coalitional psychology theory (Tooby & Cosmides 2010), I hypothesize that participants who are given feedback that they only are a "50% fit" would feel less welcome, less desired, and therefore would feel a greater need to signal their motivation to become members than those told they are a "100% fit" (H3).<sup>10</sup>

The next phase would test H1, the hypothesis that absence of facial cues increases chances of ascribing negative traits to one's political opponent. After feedback has been given

---

<sup>9</sup> Men's rights activists are an international social movement driven by the conviction that the movement for female emancipation has led public attention to neglect the rights of men and boys (for instance, that family law privileges women on the issue of child custody, that suicide is a bigger issue for men, etc.). It is also animated by the idea that manliness must be cultivated and protected, and mobilizes against what its members see as the "excesses" of the feminists movements (e.g. hatred of men).

<sup>10</sup> A challenge of this design would be to avoid completely discouraging the candidates when they are told they only are a "50% fit". So we would need to find a way of making sure they have a high baseline motivation to become members, that they are strongly committed to feminism.

to the participants on their degree of fit with the association, members of the recruitment team would provide participants with information about a rival student association, for instance a “men’s rights activists” association, which would constitute the outgroup from the perspective of the feminist association. We would vary whether the outgroup is presented in a highly humanizing way with facial cues, for instance in a video clip in which one of its members express their motivations, or in a less humanizing way, using a text with no facial cues in which the same motivations are spelled out.

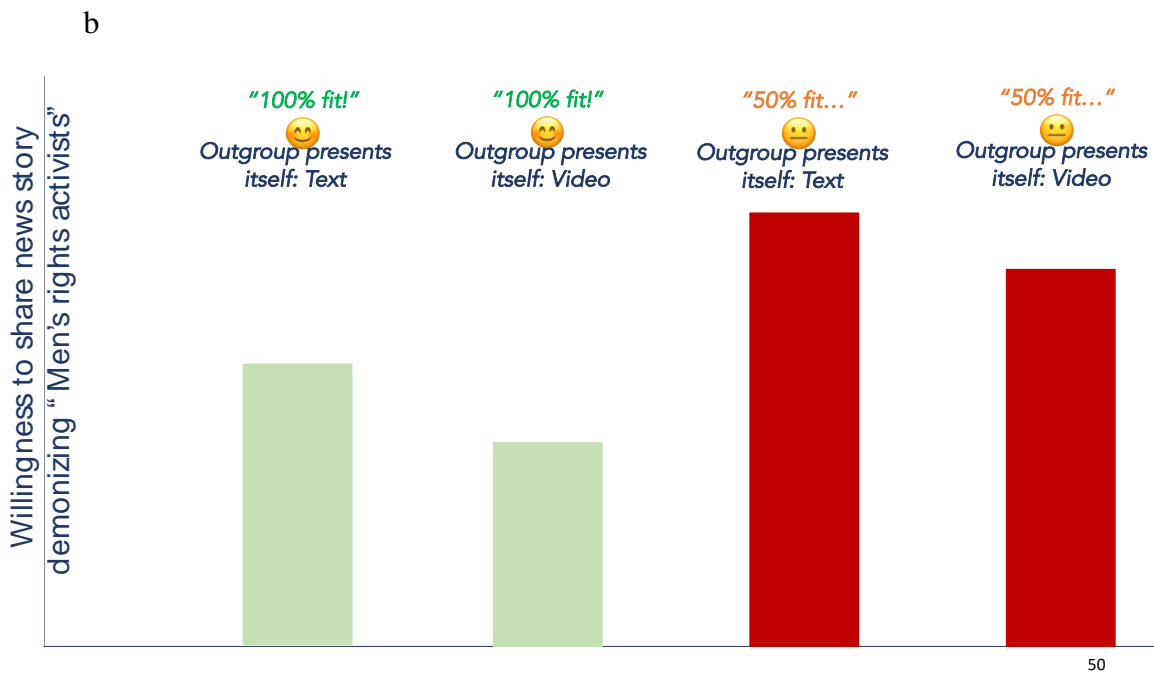
Participants’ propensity to ascribe negative traits to the outgroup, the dependent variable, could be measured in the form of participants’ willingness to share with the recruitment team an article (e.g. from a student’s newspaper) demonizing the “men’s rights activists”.

I would predict willingness to share the demonizing article to be higher when participants are told they were only a “50% fit” for the association than when they are told they were a “100% fit”, and this inclusivity factor to interact with the type of information the participants have been provided about the outgroup (such that candidates would be less likely to pass along the demonizing article when they have seen the faces of the men’s rights activists than when they haven’t; see Figure 4b below).

a



DV: "How likely would you be to share this news story with the recruitment team?"



**Figure 4:** a: Experimental design meant to test H1 and H3. b: Predicted willingness to share an article demonizing the outgroup as a function of whether the participant received a highly inclusive vs. low inclusive feedback from the recruitment team, and the type of information they received about the outgroup (video: facial cues vs. text: no facial cues).



# Appendix

## **Moral rigidity as a proximate facilitator of group cohesion and combativeness**

This commentary article was published in *Behavioral and brain sciences*:

Marie (2019). Moral rigidity as a proximate facilitator of group cohesion and combativeness. Commentary on Carsten K. W. De Dreu and Jörg Gross: Revisiting the form and function of conflict: Neurobiological, psychological, and cultural mechanisms for attack and defense within and between groups (2019). *Behavioral and brain sciences*, 42, e116

### **Abstract of De Dreu and Gross (2019)**

Conflict can profoundly affect individuals and their groups. Oftentimes, conflict involves a clash between one side seeking change and increased gains through victory and the other side defending the status quo and protecting against loss and defeat. However, theory and empirical research largely neglected these conflicts between attackers and defenders, and the strategic, social, and psychological consequences of attack and defense remain poorly understood. To fill this void, we model (1) the clashing of attack and defense as games of strategy and reveal that (2) attack benefits from mismatching its target's level of defense, whereas defense benefits from matching the attacker's competitiveness. This suggests that (3) attack recruits neuroendocrine pathways underlying behavioral activation and overconfidence, whereas defense invokes neural networks for behavioral inhibition, vigilant scanning,



and hostile attributions; and that (4) people invest less in attack than defense, and attack often fails. Finally, we propose that (5) in intergroup conflict, out-group attack needs institutional arrangements that motivate and coordinate collective action, whereas in-group defense benefits from endogenously emerging in-group identification. We discuss how games of attack and defense may have shaped human capacities for prosociality and aggression, and how third parties can regulate such conflicts and reduce their waste.

**Abstract of commentary article (Marie 2019):**

De Dreu and Gross's description of the proximate mechanisms conditioning success in intergroup conflict omits humans' deontological morality. Drawing on research on sacralization and moral objectivism, I show how "moral rigidity" may have evolved through partner selection mechanisms to foster coalitions' cohesion and combativeness in intergroup conflict.

De Dreu and Gross's argument that attack and defense are distinct strategies underpinned by different neuropsychological circuitries is an original refinement of the theory of conflict. However, their description of the proximate mechanisms facilitating success in intergroup competition (sect. 4, target article) omits humans' deontological moral intuitions. In interaction with overconfidence biases, hostile attributions, and the enforcement of "cultural rituals and sanctioning systems" (sect. 4, para. 1), what may crucially help groups of individuals cohere and prevail in conflict are high levels of "moral rigidity" in their tribal members, that is, of inflexible loyalty to their interpersonal commitments within the group.

**Proximate processes of moral rigidity.** *Sacralization* is the valuation of core social obligations (e.g., loyalty to comrades, individual freedom), symbols (the Koran, the flag), or resources (the Holy City, Hindus' cows) far beyond their practical utility. Individuals who sacralize tend to disregard consequences, opportunity costs, and the maximization of aggregate welfare, and to perceive trade-offs with secular values as taboo and morally contaminating (Atran 2010; Baron & Spranca 1997; Graham & Haidt 2012; Tetlock 2003; Tetlock et al. 2000). *Folk moral objectivism* is the intuitive propensity to represent some moral obligations as “facts” of nature, externally imposed on human wills and irreducible to mere subjective preferences (Beebe et al. 2015; Goodwin & Darley 2010, 2012; Stanford 2018).

Sacralization as a motivational process, and folk objectivism as an intuitive ontological commitment (Quine 1948), are low-level cognitive mechanisms. Both tendencies underlie what I propose to call “moral rigidity”, the inclination to treat key interpersonal obligations as non-negotiable and to reify them.

Individuals and subcultures vary as to which moral norms, “foundations”, and resources they sacralize (Graham et al. 2013; Haidt 2012; Atran 2010). Moreover, recent studies on Western samples found variability in the degree to which individuals provide deontological (*vs.* consequentialist) justifications for their moral judgments (Piazza & Sousa 2013) and objectify moral prescriptions (Goodwin & Darley 2010, 2012). Still, evidence suggests that children are predisposed to moral objectivism (Wainryb et al. 2004), and that most normally developing individuals, regardless of political orientation, have some sacred values (Frimer et al. 2016; Haidt 2012). Taken together, advances support the hypothesis that human minds may be innately prepared for moral rigidity, understood as a set of domain-specific, yet culturally flexible, cognitive adaptations to the demands of social life.

Why would the cognitive adaptations underlying moral rigidity have evolved? And what role do they play in intergroup conflict? I defend the hypothesis that humans evolved tendencies to sacralize and reify moral obligations so as to behave and be seen as devoted partners, and avoid threats on their coalition's cooperative and competitive potential. These tendencies, which proximately manifest as moral rigidity, would have helped maintain costly investment in risky coalitionary ventures, from collaborative hunting to violent intergroup conflict.

**From interdependence to a proportionality-based morality.** Much of hominin life has been about achieving social integration in fluid groups by (1) reliably signaling one's willingness to respect others' welfare, while simultaneously (2) protecting oneself from exploitation. Evolutionary theorists (Alexander 1987; Baumard et al. 2013; Frank 1988; Trivers 1971) have argued that selection pressures resulting from competitive altruism (Barclay & Willer 2007), and increasingly refined abilities to sanction insufficient dedication in potential partners (e.g., gossip and shunning), have selected for moral preferences calibrated for securing conditional cooperation (in hunting, gathering, shared parenting, combat, etc.), by sharing its costs and benefits equally among stakeholders. As a result, humans evolved intuitions that individuals with whom they interact are persons with inalienable rights, and that one's commitment to their interests should be proportionate to the amount of benefits one can expect to secure (or losses one can hope to avoid) through them (Aktipis et al. 2018; Baumard et al. 2013).

**Moral rigidity as an error management “bias”.** The logic of partner choice mutualism implies that to any type of joint venture corresponds a minimal level of cooperative engagement that each team member would obligatorily demand of his partners.

As long as (i) groups cohabitated peacefully or natural resources abounded, for instance, one can expect within-group interdependence to have been moderate. Social selection pressures would have mainly been on individuals' ability to demonstrate unconditional respect for their in-groups' physical integrity and property, and the community's resources, in order to avoid being left out of everyday ventures like hunting and gathering. However, (ii) ancestral hominins also faced recurrent intergroup competition and warfare (LeBlanc & Register 2003; Tooby & Cosmides 2010). In such circumstances, lethal threats coming from outside the group and the possibility of losing reproductive resources (such as territory or women) to rival coalitions would have driven costs of insufficient commitment to skyrocket, thereby dramatically elevating the level of prosocial engagement demanded of each group member.

Insufficient dedication to one's partners' interests, whether in situations (i) or (ii), would on average have led to reproductive impasses. In terms of error management theory (Haselton et al. 2015), evolution should therefore have selected for moral heuristics that are "biased" in favor of the most adaptive strategy: (1) making absolutely sure the individual will display the minimal level of prosocial commitment demanded by the contextual level of interdependence, while (2) showing automatic aversion to acts that could endanger the coalition's cooperative potential and combativeness in intergroup conflict. In this respect, intuitively sacralizing and objectifying core interpersonal obligations may constitute the most cost-effective computational way of completely isolating them from the temptation to trade them off with immediate self-interest (Marie & Fitouchi in preparation).

In line with this hypothesis, experiments find that potential partners expressing deontological moral judgments in everyday settings – a proxy of underlying moral rigidity – are judged more trustworthy and are more likely to be chosen (Everett et al. 2016). Furthermore, the anthropology and psychology of contemporary parochial altruism in military, insurgent, and terrorist behavior suggest that moral rigidity may have been shaped by ancestral warfare to

bolster defense of in-groups, territory, and communal resources. The values and goals for which contemporary attackers and defenders fight (e.g., the Reich, the Caliphate) often undergo ritualistic processes of sacralization and objectification (Atran 2010, 2016), up to a point where activists end up representing them as absolute coordinates of reality, and become blinded to exit strategies. By turning fighters into “devoted actors”, activation of high levels of moral rigidity – often complemented by “identity fusion” (Whitehouse 2019) – predicts individuals’ willingness to engage in costly sacrifice, including death, for their comrades and cause, and *ceteris paribus* significantly increases fighters’ chances’ to prevail in combat (Atran 2016; Gomez et al. 2017).

#### **Acknowledgments:**

I thank Léo Fitouchi, to whom my thought on “moral rigidity” is heavily indebted, and Manvir Singh, for his useful remarks on earlier drafts.

#### **References**

References are in the general references section at the end of the dissertation.

## General references

- Aarøe, L., & Petersen, M. B. (2018). Cognitive Biases and Communication Strength in Social Networks: The Case of Episodic Frames. *British Journal of Political Science*, 50(4), 1561–1581. <https://doi.org/10.1017/s0007123418000273>
- Acerbi, A. (2019). *Cultural evolution in the digital age*. Oxford University Press.
- Ainslie, G. (2001). *Breakdown of will*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139164191>
- Aktipis, A., Cronk, L., Alcock, J., Ayers, J. D., Baciú, C., Balliet, D., ... Winfrey, P. (2018). Understanding cooperation through fitness interdependence. *Nature Human Behaviour*, 2(7), 429–431. <https://doi.org/10.1038/s41562-018-0378-4>
- Al-Shawaf, L., Lewis, D. M. G., Wehbe, Y. S., & Buss, D. M. (2018). Context, Environment, and Learning in Evolutionary Psychology. In *Encyclopedia of Evolutionary Psychological Science* (pp. 1–12). Springer International Publishing. [https://doi.org/10.1007/978-3-319-16999-6\\_227-1](https://doi.org/10.1007/978-3-319-16999-6_227-1)
- Alexander, R. (1987) *The biology of moral systems (foundations of human behavior)*. Aldine de Gruyter.
- Alford, J. R., Hatemi, P. K., Hibbing, J. R., Martin, N. G., & Eaves, L. J. (2011). The Politics of Mate Choice. *The Journal of Politics*, 73(2), 362–379. doi:[10.1017/S0022381611000016](https://doi.org/10.1017/S0022381611000016)
- Allcott, H., Braghieri, L., Eichmeyer, S., & Gentzkow, M. (2020). The Welfare Effects of Social Media. *American Economic Review*, 110(3), 629–676. doi:[10.1257/aer.20190658](https://doi.org/10.1257/aer.20190658)
- Altay, S., & Mercier, H. (2020). Relevance Is Socially Rewarded, But Not at the Price of Accuracy. *Evolutionary Psychology*, 18(1), 147470492091264. doi:[10.1177/1474704920912640](https://doi.org/10.1177/1474704920912640)

- Altay, S., de Araujo, E., Mercier, H. (2020). “If this account is true, it is most enormously wonderful”: Interestingness-if-true and the sharing of true and false news. OSF: <https://osf.io/9ujq6/>
- An, J., Quercia, D. & Crowcroft, J. (2014). Partisan Sharing: Facebook Evidence and Societal Consequences. [COSN '14: Proceedings of the second ACM conference on Online social networks](#), 13-24, DOI: [10.1145/2660460.2660469](https://doi.org/10.1145/2660460.2660469)
- An, J., Quercia, D., Cha, M., Gummadi, K., & Crowcroft, J. (2014). Sharing political news: the balancing act of intimacy and socialization in selective exposure. *EPJ Data Science*, 3(1). <https://doi.org/10.1140/epjds/s13688-014-0012-2>
- Arendt, F., Steindl, N., & Kümpel, A. (2016). Implicit and explicit attitudes as predictors of gatekeeping, selective exposure, and news sharing: Testing a general model of media- related selection. *Journal of Communication*, 66, 717-740. <https://doi.org/10.1111/jcom.12256>
- Atran, S. (2016). The devoted actor: Unconditional commitment and intractable conflict across cultures. *Current Anthropology*, 57(S13), S192–S203. <https://doi.org/10.1086/685495>
- Atran, S. (2010) *Talking to the enemy: Violent extremism, sacred values, and what it means to be human*. Penguin.
- Ausubel, J. H. (2007). Renewable and nuclear heresies. *International Journal of Nuclear Governance, Economy and Ecology*, 1(3), 229. <https://doi.org/10.1504/ijngee.2007.014671>
- Bago, B., Rand, D. G. & Pennycook, G. (2020). Reasoning about climate change. *PsyArXiv*, preprint DOI: [10.31234/osf.io/vcpcb](https://doi.org/10.31234/osf.io/vcpcb)
- Bakshy, E., Hofman, J. M., Mason, W. A., & Watts, D. J. (2011). Everyone’s an influencer: Quantifying influence on Twitter. *Proceedings of the Fourth ACM International*

Conference on Web Search and Data Mining, 65–74.

<https://doi.org/10.1145/1935826.1935845>

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348, 1130-1132. DOI: 10.1126/science.aaa1160

Baratgin, J., & Politzer, G. (2010). Updating: A psychologically basic situation of probability revision. *Thinking & Reasoning*, 16(4), 253–287.

<https://doi.org/10.1080/13546783.2010.519564>

Baratgin, J., Politzer, G. (2006). Is the mind Bayesian? The case for agnosticism. *Mind & Society*, 5, 1–38. <https://doi.org/10.1007/s11299-006-0007-1>

Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science*, 26, 1531-1542. <https://doi.org/10.1177/0956797615594620>

Barclay, P., & Benard, S. (2020). The effects of social vs. asocial threats on group cooperation and manipulation of perceived threats. *Evolutionary Human Sciences*, 2. <https://doi.org/10.1017/ehs.2020.48>

Barclay, P. & Willer, R. (2007) Partner choice creates competitive altruism in humans. *Proceedings of the Royal Society of London B: Biological Sciences* 274(1610):749-53.

Barkow, J. H., Cosmides, L. & Tooby, J. (1992). *The adapted mind: Evolutionary psychology and the generation of culture*. New York, NY: Oxford University Press

Baron, J. (1994). Nonconsequentialist decisions. *Behavioral and Brain Sciences*, 17(1), 1–10. <https://doi.org/10.1017/s0140525x0003301x>

Baron, J., & Leshner, S. (2000). How serious are expressions of protected values? *Journal of Experimental Psychology: Applied*, 6(3), 183–194. <https://doi.org/10.1037/1076-898x.6.3.183>



- Baron, J., & Spranca, M. (1997). Protected values. *Organizational Behavior and Human Decision Processes*, 70, 1-16.
- Barrett, H. C., Bolyanatz, A., Crittenden, A. N., Fessler, D. M. T., Fitzpatrick, S., Gurven, M., Henrich, J., Kanovsky, M., Kushnick, G., Pisor, A., Scelza, B. A., Stich, S., von Rueden, C., Zhao, W., & Laurence, S. (2016). Small-scale societies exhibit fundamental variation in the role of intentions in moral judgment. *Proceedings of the National Academy of Sciences*, 113(17), 4688–4693.  
<https://doi.org/10.1073/pnas.1522070113>
- Baumard, N., André, J.-B., & Sperber, D. (2013). A mutualistic approach to morality: The evolution of fairness by partner choice. *Behavioral and Brain Sciences*, 36(1), 59–78.  
<https://doi.org/10.1017/s0140525x11002202>
- Beebe, J., Qiaoan, R., Endara, M. A., & Wysocki, T. (2015). Moral Objectivism in Cross-Cultural Perspective. *Journal of Cognition and Culture*, 15(3–4), 386–401.  
<https://doi.org/10.1163/15685373-12342157>
- Bentham, J. (1983). *The collected works of Jeremy Bentham: Deontology, together with a table of the springs of action; and the article on utilitarianism*. Oxford, UK: Oxford University Press. (Original work published 1879)
- Berman, J. Z., Barasch, A., Levine, E. E., & Small, D. A. (2018). Impediments to Effective Altruism: The Role of Subjective Preferences in Charitable Giving. *Psychological Science*, 29(5), 834–844. doi:[10.1177/0956797617747648](https://doi.org/10.1177/0956797617747648)
- Blancke, S., Van Breusegem, F., De Jaeger, G., Braeckman, J. & Van Montagu, M. (2015). Fatal attraction: the intuitive appeal of GMO opposition. *Trends in Plant Science*. 20, 414–418. <https://doi.org/10.1016/j.tplants.2015.03.011>
- Bloom, P. (2017). Empathy and Its Discontents. *Trends in Cognitive Sciences*, 21(1), 24–31.  
<https://doi.org/10.1016/j.tics.2016.11.004>

- Boghossian, Peter, & Lindsay, James (2019). How to have impossible conversations: A very practical guide. Boston, MA:Da Capo Press.
- Boyer, P. (2018). Minds Make Societies : How Cognition Explains the World Humans Create. Yale University Press
- Boyer, P. & Petersen, M. B. (2018). Folk-economic beliefs: An evolutionary cognitive model. Behavioral and Brain Sciences, 41, e158.  
<https://doi.org/10.1017/S0140525X17001960>
- Brady, W. J., Crockett, M. J., & Van Bavel, J. J. (2020). The MAD Model of Moral Contagion: The Role of Motivation, Attention, and Design in the Spread of Moralized Content Online. Perspectives on Psychological Science, 15(4), 978–1010.  
<https://doi.org/10.1177/1745691620917336>
- Brand, S. (2009). Whole Earth discipline: Why dense cities, nuclear power, transgenic crops, restored wildlands, and geoengineering are necessary. New York: Penguin.
- Brandt, A. (2012). Inventing conflicts of interest: a history of tobacco industry tactics. American Journal of Public Health, 102(1), 63-71. doi: 10.2105/AJPH.2011.300292
- Bronner, G. (2011). Invariants mentaux et variables sociales. In Clément, F., Kaufmann, L., La sociologie cognitive (eds.). Cogniprisme.
- Brumbach B. H., Figueredo A. J. & Ellis B. J. (2009). Effects of harsh and unpredictable environments in adolescence on development of life history strategies. Human Nature 20(1):25–51.
- Callebaut, W. & Rasskin-Gutman, D. (Eds.) (2005). Modularity: Understanding the development and evolution of natural complex systems. Cambridge, MA: MIT Press.
- Cameron, R., & Taylor, M. (2011). The 2050 roadmap for nuclear: Making a global difference. Energy and Environment, 22(1–2), 1–15. <https://doi.org/10.1260/0958-305X.22.1-2.1>

- Carothers, T. & O'Donohue, A. (2019). *Democracies Divided: The Global Challenge of Political Polarization*. Washington, D.C.: Brookings Institution Press.
- Caviola, L., Schubert, S., Teperman, E., Moss, D., Greenberg, S., F., Nadira S. (2020). Donors vastly underestimate differences in charities' effectiveness. *Judgment and Decision Making*. 15(4), 509–516.
- Clune, J., Mouret, J.-B. & Lipson, H. (2013). The evolutionary origins of modularity. *Proceedings of the Royal Society B*, 280(1755), 20122863.  
<https://doi.org/10.1098/rspb.2012.2863>
- Coppini, D., Duncan, M. A., McLeod, D. M., Wise, D. A., Bialik, K. E., & Wu, Y. (2017). When the whole world is watching: A motivations-based account of selective expression and exposure. *Computers in Human Behavior*, 75, 766-774.  
<https://doi.org/10.1016/j.chb.2017.04.020>
- Cordonier, L. (2018). *La nature du social*. Paris : Presses Universitaires de France.
- Cushman, F., Young, L., & Greene, J. D. (2010). Multi-system Moral Psychology. In *The Moral Psychology Handbook* (pp. 47–71). Oxford University Press.  
<https://doi.org/10.1093/acprof:oso/9780199582143.003.0003>
- Damasio, A. (1994). *Descartes's error: emotion, reason, and the human brain*. Grosset/Putnam, New York
- Dehaene, S. (2009). *Reading in the Brain: The New Science of How We Read*. Penguin.
- De Dreu, C. K. W., & Gross, J. (2018). Revisiting the form and function of conflict: Neurobiological, psychological, and cultural mechanisms for attack and defense within and between groups. *Behavioral and Brain Sciences*, 42.  
<https://doi.org/10.1017/s0140525x18002170>
- Ditto, P. H., Liu, B. S., Clark, C. J., Wojcik, S. P., Chen, E. E., Grady, R. H., ... Zinger, J. F. (2018). At Least Bias Is Bipartisan: A Meta-Analytic Comparison of Partisan Bias in

Liberals and Conservatives. *Perspectives on Psychological Science*,

174569161774679. doi:[10.1177/1745691617746796](https://doi.org/10.1177/1745691617746796)

Draper, R. (July 16th, 2020). Colin Powell Still Wants Answers. *The New York Times*.

<https://www.nytimes.com/2020/07/16/magazine/colin-powell-iraq-war.html>

Ederly, S., Mourão, R. R., Thorson, E., & Tham, S. M. (2019). When Do Audiences Verify?

How Perceptions About Message and Source Influence Audience Verification of

News Headlines. *Journalism & Mass Communication Quarterly*, 97(1), 52–71.

<https://doi.org/10.1177/1077699019864680>

Ekstrom, D. & Calvin K. Lai. (2020). The Selective Communication of Political Information.

PsyArXiv: <https://psyarxiv.com/pnr9u/>

Everett, J.A., Pizarro, D.A. & Crockett, M. J. (2016) Inference of trustworthiness from

intuitive moral judgments. *Journal of Experimental Psychology: General* 145(6):772.

Faulmüller, N., Mojzisch, A., Kerschreiter, R., & Schulz-Hardt, S. (2012). Do you want to

convince me or to be understood? Preference-consistent information sharing and its

motivational determinants. *Personality and Social Psychology Bulletin*, 38, 1684-

1696. DOI: [10.1177/0146167212458707](https://doi.org/10.1177/0146167212458707)

Fernbach, P. M., Light, N., Scott, S. E., Inbar, Y., & Rozin, P. (2019). Extreme opponents of

genetically modified foods know the least but think they know the most. *Nature*

*Human Behaviour*, 3(3), 251–256. <https://doi.org/10.1038/s41562-018-0520-3>

Fernbach, P. M., Rogers, T., Fox, C. R., & Sloman, S. A. (2013). Political Extremism Is

Supported by an Illusion of Understanding. *Psychological Science*, 24(6), 939–946.

<https://doi.org/10.1177/0956797612464058>

Finkel, E.J., Bail, C. A., Cikara, M., Ditto, P. H, Iyengar, S., Klar, S. Mason, L., McGrath, M.

C. Nyhan, B., Rand, D. G., Skitka, L. J., Tucker, J., Van Bavel, J. J., Wang, C. S.

Druckman, J. N. (2020). Political sectarianism in America. *Science*, 370(6516), 533-

536, DOI: [10.1126/science.abe1715](https://doi.org/10.1126/science.abe1715)

- Fisher, M., Knobe, J., Strickland, B., & Keil, F. C. (2016). The Influence of Social Interaction on Intuitions of Objectivity and Subjectivity. *Cognitive Science*, 41(4), 1119–1134. doi:[10.1111/cogs.12380](https://doi.org/10.1111/cogs.12380)
- Fiske, S. T., & Dupree, C. (2014). Gaining trust as well as respect in communicating to motivated audiences about science topics. *Proceedings of the National Academy of Sciences*, 111(Supplement\_4), 13593–13597. <https://doi.org/10.1073/pnas.1317505111>
- Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: warmth and competence. *Trends in Cognitive Sciences*, 11(2), 77–83. DOI:[10.1016/j.tics.2006.11.005](https://doi.org/10.1016/j.tics.2006.11.005)
- Fletcher, R., & Nielsen, R. K. (2017). Are news audiences increasingly fragmented? A cross-national comparative analysis of cross-platform news audience fragmentation and duplication. *Journal of Communication*, 67(4), 476–498.
- Frank, R. (1988) *Passions within reason: The strategic role of the emotions*, vol. 1. Norton.
- Frenda, S. J., Knowles, E. D., Saletan, W., & Loftus, E. F. (2013). False memories of fabricated political events. *Journal of Experimental Social Psychology*, 49(2), 280–286. <https://doi.org/10.1016/j.jesp.2012.10.013>
- Frimer, J. A., Skitka, L. J., & Motyl, M. (2017). Liberals and conservatives are similarly motivated to avoid exposure to one another’s opinions. *Journal of Experimental Social Psychology*, 72, 1–12. <https://doi.org/10.1016/j.jesp.2017.04.003>
- Frimer, J., Motyl, M., & Tell, C. (2016). Sacralizing liberals and fair-minded conservatives: Ideological symmetry in the moral motives in the culture war. *Analyses of Social Issues and Public Policy*.
- Gelman, S. (2004). Psychological essentialism in children. *Trends in Cognitive Sciences*, 8(9), 404–409. <https://doi.org/10.1016/j.tics.2004.07.001>

- Gentzkow, M., & Shapiro, J. M. (2011). Ideological segregation online and offline. *Quarterly Journal of Economics*, 126(4), 1799–1839.
- Gigerenzer, G. (2018). *Review of Behavioral Economics*. 5, 303–336.
- Gómez, Á., López-Rodríguez, L., Sheikh, H., Ginges, J., Wilson, L., Waziri, H., ... Atran, S. (2017). The devoted actor's will to fight and the spiritual dimension of human conflict. *Nature Human Behaviour*, 1(9), 673–679. <https://doi.org/10.1038/s41562-017-0193-3>
- Goodwin, G. P. & Darley, J. M. (2010) The perceived objectivity of ethical beliefs: Psychological findings and implications for public policy. *Review of Philosophy and Psychology* 1:161–88. Available at: <http://dx.doi.org/10.1007/s13164-009-0013-4>.
- Goodwin, G. P., & Darley, J. M. (2012). Why are some moral beliefs perceived to be more objective than others? *Journal of Experimental Social Psychology*, 48(1), 250–256. <https://doi.org/10.1016/j.jesp.2011.08.006>
- Graham, J., & Haidt, J. (2012). Sacred values and evil adversaries: A moral foundations approach. In M. Mikulincer & P. R. Shaver (Eds.), *Herzliya series on personality and social psychology. The social psychology of morality: Exploring the causes of good and evil* (p. 11–31). American Psychological Association. <https://doi.org/10.1037/13091-001>
- Graham, J., Haidt, J., Koleva, S. P., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. *Advances in Experimental Social Psychology*, 47, 55–130.
- Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. *Advances in Experimental Social Psychology*, 47, 55–130.
- Gray, K., Young, L., & Waytz, A. (2012). Mind Perception Is the Essence of Morality. *Psychological Inquiry*, 23(2), 101–124. doi:[10.1080/1047840X.2012.651387](https://doi.org/10.1080/1047840X.2012.651387)

- Guess, A., Nagler, J., & Tucker, J. (2019a). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances*, 5(1), eaau4586.  
<https://doi.org/10.1126/sciadv.aau4586>
- Guess, A., Nyhan, B., Lyons, B., & Reifler, J. (2019b). Avoiding the echo chamber about echo chambers: Why selective exposure to like-minded political news is less prevalent than you think. Knight Foundation. [https://kf-site-production.s3.amazonaws.com/media\\_elements/files/000/000/133/original/Topos\\_KF\\_White-Paper\\_Nyhan\\_V1.pdf](https://kf-site-production.s3.amazonaws.com/media_elements/files/000/000/133/original/Topos_KF_White-Paper_Nyhan_V1.pdf)
- Guess, A. M., Lerner, M., Lyons, B., Montgomery, J. M., Nyhan, B., Reifler, J., & Sircar, N. (2020). A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. *Proceedings of the National Academy of Sciences*, 201920498.  
<https://doi.org/10.1073/pnas.1920498117>
- Hacquín, A.-S., Altay, S., Aaroe, L., Mercier, H. (2020). Fear of contamination and public opinion on nuclear energy. *PsyArXiv*. [10.31234/osf.io/wdvuq](https://doi.org/10.31234/osf.io/wdvuq)
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814–834.  
<https://doi.org/10.1037/0033-295x.108.4.814>
- Haidt, J. (2012) *The Righteous Mind: Why Good People Are Divided by Politics and Religion*: Vintage Books.
- Haselton, M. G., & Nettle, D. (2006). The Paranoid Optimist: An Integrative Evolutionary Model of Cognitive Biases. *Personality and Social Psychology Review*, 10(1), 47–66. [https://doi.org/10.1207/s15327957pspr1001\\_3](https://doi.org/10.1207/s15327957pspr1001_3)
- Haselton, M. G., Nettle, D., & Andrews, P. W. (2015). The Evolution of Cognitive Bias. In D. M. Buss (Ed.), *The Handbook of Evolutionary Psychology* (pp. 724–746). Hoboken, NJ, USA: John Wiley & Sons, Inc. <https://doi.org/10.1002/9780470939376.ch25>

- Hatemi, P. K., Medland, S. E., Klemmensen, R., Oskarsson, S., Littvay, L., Dawes, C. T., Verhulst, B., McDermott, R., Nørgaard, A. S., Klofstad, C. A., Christensen, K., Johannesson, M., Magnusson, P. K. E., Eaves, L. J., & Martin, N. G. (2014). Genetic Influences on Political Ideologies: Twin Analyses of 19 Measures of Political Ideologies from Five Democracies and Genome-Wide Findings from Three Populations. *Behavior Genetics*, 44(3), 282–294. <https://doi.org/10.1007/s10519-014-9648-8>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *The Behavioral and Brain Sciences*, 33, 61–83; 83–135.  
doi:10.1017/S0140525X0999152X
- Hibbing, J. R., Smith, K. B., & Alford, J. R. (2014). Differences in negativity bias underlie variations in political ideology. *Behavioral and Brain Sciences*, 37(3), 297–307.  
<https://doi.org/10.1017/s0140525x13001192>
- Hume, D. (2000). *A Treatise of Human Nature*. Oxford, Clarendon Press. (Original work published in 1739).
- Iyengar, S., & Hahn, K. S. (2009). Red media, blue media: Evidence of ideological selectivity in media use. *Journal of Communication*, 59, 19–39. doi:10.1111/j.1460-2466.2008.01402.x
- Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2018). The Origins and Consequences of Affective Polarization in the United States. *Annual Review of Political Science*, 22(1). doi:[10.1146/annurev-polisci-051117-073034](https://doi.org/10.1146/annurev-polisci-051117-073034)
- Jenkins-Smith, H. C., Silva, C. L., Nowlin, M. C., & deLozier, G. (2010). Reversing Nuclear Opposition: Evolving Public Acceptance of a Permanent Nuclear Waste Disposal Facility. *Risk Analysis*, 31(4), 629–644. <https://doi.org/10.1111/j.1539-6924.2010.01543.x>



- Kahan, D. M. (2016). The Politically Motivated Reasoning Paradigm, Part 1: What Politically Motivated Reasoning Is and How to Measure It. *Emerging Trends in the Social and Behavioral Sciences*, 1-16. doi:[10.1002/9781118900772.etrds0417](https://doi.org/10.1002/9781118900772.etrds0417)
- Kahan, D. M. (2013). Ideology, motivated reasoning, and cognitive reflection. *Judgment and Decision Making*, 8(4), 407–424.
- Kahan, D. M., Jenkins-Smith, H., & Braman, D. (2011a). Cultural cognition of scientific consensus. *Journal of Risk Research*, 14(2), 147–174.  
<https://doi.org/10.1080/13669877.2010.511246>
- Kahan, D. M., Wittlin, M., Peters, E., Slovic, P., Ouellette, L. L., et al. (2011b). The tragedy of the risk-perception commons: Culture conflict, rationality conflict, and climate change. *Cultural Cognition Working Paper 89*.  
[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1871503](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1871503).
- Kahneman, D., & Tversky, A. (1988). Prospect theory: An analysis of decision under risk. In P. Gärdenfors & N.-E. Sahlin (Eds.), *Decision, probability, and utility: Selected readings* (p. 183–214). (Reprinted from "Econometrica," 47 (1979), pp. 263-291)Cambridge University Press. <https://doi.org/10.1017/CBO9780511609220.014>
- Kant, I. (1785/2005). *Groundwork for the metaphysics of morals* (ed. L. Denis). Toronto, Canada: Broadview Press.
- Katz, E., Lazarsfeld, P. F. (1955). *Personal Influence*. New York: Free Press.
- Keil, F. C. (2003). Folkscience: coarse interpretations of a complex reality. *Trends in Cognitive Sciences*, 7(8), 368–373. doi:[10.1016/S1364-6613\(03\)00158-X](https://doi.org/10.1016/S1364-6613(03)00158-X)
- Kharecha, P. A., & Sato, M. (2019). Implications of energy and CO2 emission changes in Japan and Germany after the Fukushima accident. *Energy Policy*, 132, 647–653.  
<https://doi.org/10.1016/j.enpol.2019.05.057>
- Kirk, A. (2016). *Data visualization. A Handbook for data driven design*. London: Sage.

- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3), 480–498.  
<https://doi.org/10.1037/0033-2909.108.3.480>
- Kurzban, R. (2010) *Why everyone (else) is a hypocrite: Evolutionary psychology and the modular mind*. Princeton University Press.
- LeBlanc, S.A. & Register, K. (2003). *Constant battles: The myth of the peaceful, noble savage*. St. Martins Press.
- Leeper, Thomas J and Rune Slothuus (2014). Political parties, motivated reasoning, and public opinion formation. *Political Psychology* 35:129–156.  
<https://doi.org/10.1111/pops.12164>
- Lees, J., Cikara, M. (2019). Inaccurate group meta-perceptions drive negative out-group attributions in competitive contexts. *Nature human behavior*, 4(3), 279-286.  
<https://doi.org/10.1038/s41562-019-0766-4>
- Levendusky, M. (2013). Why Do Partisan Media Polarize Viewers? *American Journal of Political Science*, 57(3), 611-623. Retrieved January 28, 2021, from  
<http://www.jstor.org/stable/23496642>
- Levy, J. S., & Thompson, W. R. (2011). *The arc of war: Origins, escalation, and transformation*. Chicago: University of Chicago Press.
- Levy, N. (2017). Due deference to denialism: explaining ordinary people’s rejection of established scientific findings. *Synthese*, 196(1), 313–327.  
<https://doi.org/10.1007/s11229-017-1477-x>
- Liang, H. (2018). Broadcast versus viral spreading: The structure of diffusion cascades and selective sharing on social media. *Journal of Communication*, 68, 525-546.  
<https://doi.org/10.1093/joc/jqy006>
- Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of*

Personality and Social Psychology, 37, 2098-2109. <https://doi.org/10.1037/0022-3514.37.11.2098>

Lukianoff, G. and Haidt, J. (2018). *The Coddling of the American Mind. How Good Intentions and Bad Ideas Are Setting Up a Generation for Failure*. Penguin Random House.

Marie, A. (2019). Moral rigidity as a proximate facilitator of group cohesion and combativeness, commentary on De Dreu, C. K. W. & Gross, J. (2019). Revisiting the form and function of conflict: Neurobiological, psychological, and cultural mechanisms for attack and defense within and between groups, p. 28, *Behavioral and Brain Sciences*, 42, e116: 1–66. [DOI:10.1017/S0140525X18002170](https://doi.org/10.1017/S0140525X18002170)

Marie, A. (2020). Moral rigidity evolved to strengthen bonds within groups. "What is extremism?" issue of *This View of Life* magazine, The evolution institute. Editors: Anthony C. Lopez, Hammad Sheikh, David Sloan Wilson. <https://thisviewoflife.com/moral-rigidity-evolved-to-strengthen-bonds-within-groups/>

Marie, A. & Fitouchi, L. (in preparation). The Evolution of Moral Rigidity. In De Smedt, J. & De Cruz, H. (eds.), *The nuts and bolts of evolutionary ethics* (forthcoming).

Marie, A., Altay, S., Strickland, B. (2020). Moral conviction predicts sharing preference for politically congruent headlines. OSF Preprint: [10.31219/osf.io/twq3y](https://doi.org/10.31219/osf.io/twq3y)

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco, CA: Freeman.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review Sociology*, 27, 415-444. <https://doi.org/10.1146/annurev.soc.27.1.415>

McPhetres, J., Rutjens, B. T., Weinstein, N., & Brisson, J. A. (2019). Modifying attitudes about modified foods: Increased knowledge leads to more positive attitudes. *Journal*

of Environmental Psychology, 64, 21–29.

<https://doi.org/10.1016/j.jenvp.2019.04.012>

Mell, H., Safra, L., Algan, Y., Baumard, N., & Chevallier, C. (2018). Childhood environmental harshness predicts coordinated health and reproductive strategies: A cross-sectional study of a nationally representative sample from France. *Evolution and Human Behavior*, 39(1), 1–8.

<https://doi.org/10.1016/j.evolhumbehav.2017.08.006>

Mercier H., Sperber D., (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34(2):57–74.

<https://doi.org/10.1017/S0140525X10000968>

Mercier, H. (2017). How Gullible are We? A Review of the Evidence from Psychology and Social Science. *Review of General Psychology*, 21(2), 103–122.

<https://doi.org/10.1037/gpr0000111>

Mercier, H. (2020) *Not Born Yesterday. The Science of Who We Trust and What We Believe*. Princeton University Press.

Mercier, H., & Heintz, C. (2013). Scientists' Argumentative Reasoning. *Topoi*, 33(2), 513–524. <https://doi.org/10.1007/s11245-013-9217-4>

Mercier, H., & Sperber, D. (2017). *The Enigma of Reason*. Cambridge, MA: Harvard University Press.

Mesoudi, A., & Whiten, A. (2008). The multiple roles of cultural transmission experiments in understanding human cultural evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1509), 3489–3501.

<https://doi.org/10.1098/rstb.2008.0129>

Morin, O., Jacquet, P.-O., Vaesen, K., Acerbi, A. (2020). Social information use and social information waste. *PsyArXiv*. Doi: [10.31235/osf.io/rqcdf](https://doi.org/10.31235/osf.io/rqcdf)

- Mosleh, M., Pennycook, G., & Rand, D. (2020). Self-reported willingness to share political news articles in online surveys correlates with actual sharing on Twitter. *PLOS One*, 15, e0228882. <https://doi.org/10.1371/journal.pone.0228882>
- Moussaïd, M., Brighton, H., & Gaissmaier, W. (2015). The amplification of risk in experimental diffusion chains. *Proceedings of the National Academy of Sciences*, 112(18), 5631–5636. <https://doi.org/10.1073/pnas.1421883112>
- National Academies of Sciences, Engineering, and Medicine. (2016). *Genetically Engineered Crops: Experiences and Prospects*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/23395>.
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomena in many guises. *Review of General Psychology* 2(2):175–220. <https://doi.org/10.1037/1089-2680.2.2.175>
- Nuclear Energy Agency (2010). *Public Attitudes to Nuclear Power*. [https://www.oecd-neo.org/jcms/pl\\_14534/public-attitudes-to-nuclear-power](https://www.oecd-neo.org/jcms/pl_14534/public-attitudes-to-nuclear-power)
- Origi, G. (2004). Is Trust an Epistemological Notion? *Episteme*, 1(1), 61–72. <https://doi.org/10.3366/epi.2004.1.1.61>
- Osmundsen, M., Bor, A., Vahlstrup, P. B., Bechmann, A., and Petersen, M. P. (2020). Partisan polarization is the primary psychological motivation behind “fake news” sharing on Twitter. *PsyArXiv Preprints*: <https://psyarxiv.com/v45bk/>
- Pennycook, G. (2020). Belief Bias and Its Significance for Modern Social Science. *Psychological Inquiry*, 31(1), 57–60. <https://doi.org/10.1080/1047840x.2020.1722577>
- Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. (2020). Understanding and reducing the spread of misinformation online. *PsyArXiv Working Paper*, 1–63. <https://doi.org/10.31234/OSF.IO/3N9U8>

- Pepper, G. V., & Nettle, D. (2017). The behavioural constellation of deprivation: Causes and consequences. *Behavioral and Brain Sciences*, 40.  
<https://doi.org/10.1017/s0140525x1600234x>
- Pereira, A., Van Bavel, J. J., & Harris, E. (2020). Identity concerns drive belief: The impact of partisan identity on the belief and dissemination of true and false news. PsyArXiv Working Paper. <https://doi.org/10.31234/OSF.IO/7VC5D>
- Petersen, M. B., Osmundsen, Tooby, J. (2020). The Evolutionary Psychology of Conflict and the Functions of Falsehood. PsyArXiv: <https://psyarxiv.com/kaby9>
- Pew Research Center (2014, June 12th). Political Polarization in the American Public, <https://www.pewresearch.org/politics/2014/06/12/political-polarization-in-the-american-public/>
- Pinker, S., (June 21, 2019). One thing to change: Anecdotes aren't data. *The Harvard Gazette*.  
<https://news.harvard.edu/gazette/story/2019/06/focal-point-harvard-professor-steven-pinker-says-the-truth-lies-in-the-data/>
- Pinker, S., (2018). *Enlightenment Now. The Case for Reason, Science, Humanism, and Progress*. Allen Lane.
- Pinker, S. (2011). *The better angels of our nature: Why violence has declined*. Penguin.
- Pinker, S., (2002) *The Blank Slate: The Modern Denial of Human Nature*. New York, NY: Penguin
- Pogorelskiy, K., & Shum, M. (2019). News we like to share: How news sharing on social networks influences voting outcomes. [CAGE Online Working Paper Series 427](#), Competitive Advantage in the Global Economy (CAGE).  
<https://ideas.repec.org/p/cge/wacage/427.html>
- Pronin, E. (2007). Perception and misperception of bias in human judgment. *Trends in Cognitive Sciences*, 11, 37-43. [10.1016/j.tics.2006.11.001](https://doi.org/10.1016/j.tics.2006.11.001)
- Quine, W. V., (1948). On What There Is. *The Review of Metaphysics*, 2(1): 21–38.

- Rogowski, J. C., & Sutherland, J. L. (2015). How Ideology Fuels Affective Polarization. *Political Behavior*, 38(2), 485–508. doi:[10.1007/s11109-015-9323-7](https://doi.org/10.1007/s11109-015-9323-7)
- Roser, M. (2016a). Economic growth. Our World in Data.  
<https://ourworldindata.org/economic-growth/>
- Roser, M. (2016b). Hunger and undernourishment. Our World in Data.  
<https://ourworldindata.org/hunger-and-overnourishment/>
- Rosling, H, Rosling, O., Rosling, Rönnlund, A. (2018). *Factfulness: Ten Reasons We're Wrong About the World – and Why Things Are Better Than You Think*. Flatiron Books.
- Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: an illusion of explanatory depth. *Cognitive Science*, 26(5), 521–562.  
doi:[10.1207/s15516709cog2605\\_1](https://doi.org/10.1207/s15516709cog2605_1)
- Rubin, P. H. (2003) Folk economics. *Southern Economic Journal* 70(1):157–71. doi: 10.2307/1061637.
- Rubin, P. H. (2014). Emporiophobia (Fear of Markets): Cooperation or Competition? *Southern Economic Journal*, 80(4), 875–889. doi:[10.4284/0038-4038-2013.287](https://doi.org/10.4284/0038-4038-2013.287)
- Schlosser, G. & Wagner, G. P. (Eds.) (2004). *Modularity in development and evolution*. Chicago, IL: University of Chicago Press.
- Shin, J., & Thorson, K. (2017). Partisan selective sharing: The biased diffusion of fact-checking messages on social media. *Journal of Communication*, 67, 233-255. <https://doi.org/10.1111/jcom.12284>
- Sidgwick, H. (1884). *The methods of ethics* (3rd ed.). London, England: Macmillan.
- Singh, M. & Hoffman, M. (2020). Commitment and impersonation A reputation-based theory of principled behavior. PsyArXiv : [10.31234/osf.io/ua57r](https://doi.org/10.31234/osf.io/ua57r)
- Singer, P. (2009). *The Life You Can Save*. Random House.

- Siqueira, D. S., de Almeida Meystre, J., Hilário, M. Q., Rocha, D. H. D., Menon, G. J., & da Silva, R. J. (2019). Current perspectives on nuclear energy as a global climate change mitigation option. *Mitigation and Adaptation Strategies for Global Change*, 24(5), 749–777. <https://doi.org/10.1007/s11027-018-9829-5>
- Skitka, L. J. (2010). The Psychology of Moral Conviction. *Social and Personality Psychology Compass*, 4(4), 267–281. doi:[10.1111/j.1751-9004.2010.00254.x](https://doi.org/10.1111/j.1751-9004.2010.00254.x)
- Skitka, L. J., & Mullen, E. (2002). The dark side of moral conviction. *Analyses of Social Issues and Public Policy (ASAP)*, 2(1), 35–41. <https://doi.org/10.1111/j.1530-2415.2002.00024.x>
- Skitka, L. J., Bauman, C. W., & Sargis, E. G. (2005). Moral conviction: Another contributor to attitude strength or something more? *Journal of Personality and Social Psychology*, 88, 895–917. [10.1037/0022-3514.88.6.895](https://doi.org/10.1037/0022-3514.88.6.895)
- Sloman, S. & Fernbach, P. (2017) *The Knowledge Illusion. Why We Never Think Alone.* MacMillan
- Slote, M. (1995). Agent-based virtue ethics. *Midwest Studies in Philosophy*, 20, 83–101. <https://doi.org/10.1080/05568640409485132>
- Sperber, D. (2009). Culturally transmitted misbeliefs. *Behavioral and Brain Sciences*, 32(6), 534–535. <https://doi.org/10.1017/s0140525x09991348>
- Sperber, D., & Baumard, N. (2012). Moral Reputation: An Evolutionary and Cognitive Perspective. *Mind & Language*, 27(5), 495–518. <https://doi.org/10.1111/mila.12000>
- Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & Language*, 25(4), 359–393. <https://doi.org/10.1111/j.1468-0017.2010.01394.x>
- Sperber, D. & Hirschfeld, L.A. (2004). The cognitive foundations of cultural stability and diversity. *Trends in Cognitive Science*, 8(1): 40-46. <https://doi.org/10.1016/j.tics.2003.11.002>



- Sperber, D., & Wilson, D. (1987). Précis of Relevance: Communication and Cognition. *Behavioral and Brain Sciences*, 10(04), 697.  
<https://doi.org/10.1017/s0140525x00055345>
- Spranca, M., Minsk, E., & Baron, J. (1991). Omission and commission in judgment and choice. *Journal of Experimental Social Psychology*, 27, 76–105.  
[https://doi.org/10.1016/0022-1031\(91\)90011-T](https://doi.org/10.1016/0022-1031(91)90011-T)
- Ståhl T., Zaal M.P., Skitka L.J. (2016). Moralized Rationality: Relying on Logic and Evidence in the Formation and Evaluation of Belief Can Be Seen as a Moral Issue. *PLoS ONE* 11(11): e0166332. doi:10.1371/journal.pone.0166332
- Stanford, P. K. (2018). The difference between ice cream and Nazis: Moral externalization and the evolution of human cooperation. *Behavioral and Brain Sciences*, 41, e95.  
<https://doi.org/10.1017/S0140525X17001911>
- Stanovich, K. E. & West, R. F. (2007) Natural myside bias is independent of cognitive ability. *Thinking & Reasoning* 13(3):225–47. <https://doi.org/10.1080/13546780600780796>
- Stone, D., Patton, B., & Heen, S. (2010). *Difficult conversations: How to discuss what matters most*. New York, NY: Penguin.
- Stroud, N. J. (2008). Media use and political predispositions: Revisiting the concept of selective exposure. *Political Behavior*, 30, 341–366. doi:10.1007/s11109-007-9050-9
- Tappin, B. M., Pennycook, G. and Rand, D. G. (2020). Thinking clearly about causal inferences of politically motivated reasoning: why paradigmatic study designs often undermine causal inference, *Current Opinion in Behavioral Sciences*, 34:81–87.  
<https://doi.org/10.1016/j.cobeha.2020.01.003>
- Tetlock, P. E. (2003). Thinking the unthinkable: sacred values and taboo cognitions. *Trends in Cognitive Sciences*, 7(7), 320–324. [https://doi.org/10.1016/S1364-6613\(03\)00135-9](https://doi.org/10.1016/S1364-6613(03)00135-9)
- Tetlock, P. E., Kristel, O. V, Elson, S. B., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical

counterfactuals. *Journal of Personality and Social Psychology*, 78(5), 853–870.

<https://doi.org/10.1037//0022-3514.78.5.853>

Tetlock, P. E., Kristel, O. V., Elson, B., Green, M., & Lerner, J. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology*, 78, 853-870.

Tooby, J. (2017). Coalitional instincts. Edge. Retrieved from <https://www.edge.org/response-detail/27168>

Tooby, J. & Cosmides, L. (2010) Groups in mind: The coalitional roots of war and morality. In: *Human morality and sociality: Evolutionary and comparative perspectives*, ed. H. Høgh-Olesen, pp. 191–234. Palgrave MacMillan.

Tooby, J., & Cosmides, L. (1990). The past explains the present – Emotional adaptations and the structure of ancestral environments. *Ethology and Sociobiology*, 11, 375-424.

[https://doi.org/10.1016/0162-3095\(90\)90017-Z](https://doi.org/10.1016/0162-3095(90)90017-Z)

Trivers, R. L. (1971) The evolution of reciprocal altruism. *Quarterly Review of Biology* 46:35–57.

Tybur, J. M., Lieberman, D., Kurzban, R., & DeScioli, P. (2013). Disgust: Evolved function and structure. *Psychological Review*, 120(1), 65–84.

<https://doi.org/10.1037/a0030778>

Uhlmann, E.L., Pizarro, D. A. Diermeier, D. (2015) A person-centered approach to moral judgment. *Perspectives on Psychological Science*. 10(1) 72 –81.

DOI: [10.1177/1745691614556679](https://doi.org/10.1177/1745691614556679)

Van Bavel, J. J., & Pereira, A. (2018). The Partisan Brain: An Identity-Based Model of Political Belief. *Trends in Cognitive Sciences*, 22(3), 213–224.

doi:[10.1016/j.tics.2018.01.004](https://doi.org/10.1016/j.tics.2018.01.004)

- van Prooijen, J.-W., & van Vugt, M. (2018). Conspiracy Theories: Evolved Functions and Psychological Mechanisms. *Perspectives on Psychological Science*, 13(6), 770–788.  
<https://doi.org/10.1177/1745691618774270>
- Von Neumann J, Morgenstern O. (1944). *The Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.
- von Hippel, W., & Trivers, R. (2011). The evolution and psychology of self-deception. *Behavioral and Brain Sciences*, 34(1), 1–16.  
<https://doi.org/10.1017/s0140525x10001354>
- Wainryb, C., Shaw, L., Langley, M., Cottam, K., & Lewis, R. (2004). Children’s thinking about diversity of belief in the early school years: Judgments of relativism, tolerance, and disagreeing persons. *Child Development*, 75:3, 687-703.
- Walmsley, J., & O’Madagain, C. (2020). The Worst-Motive Fallacy: A Negativity Bias in Motive Attribution. *Psychological Science*, 31(11), 1430–1438.  
<https://doi.org/10.1177/0956797620954492>
- Weeks, B. E., Ardèvol-Abreu, A., & Gil de Zúñiga, H. (2017). Online influence? Social media use, opinion leadership, and political persuasion. *International Journal of Public Opinion Research*, 29, 214-239. <https://doi.org/10.1093/ijpor/edv050>
- Wesseler, J., & Zilberman, D. (2014). The economic power of the Golden Rice opposition. *Environment and Development Economics*, 19(06), 724–742.  
doi:[10.1017/S1355770X1300065X](https://doi.org/10.1017/S1355770X1300065X)
- Whitehouse, H. (2019). Dying for the group: Towards a general theory of extreme self-sacrifice. *Behavioral and Brain Sciences*, 1–64.  
<https://doi.org/10.1017/S0140525X18000249>
- Yamagishi T, Terai S, Kiyonari T, Mifune N, Kanazawa S (2007). The social exchange heuristic: Managing errors in social exchange. *Rationality and Society*, 19(3), 259–291. <https://doi.org/10.1177/1043463107080449>

Young, L. & Tsoi, L. (2013). When mental states matter, when they don't, and what that means for morality. *Social and Personality Psychology Compass* 7/8 (2013): 585–604, doi: 10.1111/spc3.12044