Jean-François Bonnefon

jean-francois.bonnefon@iast.fr
http://jfbonnefon.github.io

Toulouse School of Economics
1, Esplanade de l'Université
31080 Toulouse Cedex 06 France

Current position

2015—	**Senior CNRS Research Scientist** (Toulouse, France) Toulouse School of Economics + Toulouse School of Management

Roles

2024–	Director, Institute for Advanced Study in Toulouse
2024–	Director, Department of Social and Behavioral Sciences, TSE
2024–	Scientific Board, Artificial and Natural Intelligence Toulouse Institute
2024–	Scientific Board, EXPLAIN initiative, DGRI
2024–	Program committee, 2026 World Conference on Transport Research
2024–	Co-Chair, Machine+Behavior Conference
2023–	Internal Board, Toulouse School of Management Research
2022–	Advisory Board, Data Intelligence Institute of Paris
2019–	Chair of Moral AI, Artificial and Natural Intelligence Toulouse Institute
2018–	Head of the AI & Society program, TSE Digital Center
2024–2025	Guest Editor, Cognition, special issue on morality and AI
2023–2024	Vice-President, TSE Ethical Review Board
2022–2023	Advisor, European Commission report on Trustworthy Public Communications
2021–2024	Leader of the Expanded Morality team, IAST
2021–2023	Scientific Board, 'IA Double Je' museum exhibit, Paris and Toulouse
2021–2023	Scientific Director, Institute for Advanced Study in Toulouse
2021	Program committee, Common Good Summit, Toulouse
2020–2021	Scientific Chair, 9th International Conference on Thinking
2013–2021	French National Committee for Scientific Psychology (CNFPS)
2019–2020	President, European Commission expert group on the ethics of driverless mobility
2016–2020	President, TSE Ethical Review Board
2011–2024	Steering committee, Institute for Advanced Study in Toulouse
2011–2020	Scientific board, TULIP Excellence Center, Toulouse
2016–2018	National Committee for Scientific Research (section 26)
2016–2018	Steering committee, Toulouse Institute for Complex Systems Studies
2016–2017	Senior Associate Editor, Cognition

[Roles continue next page]

Roles, continued

2016	Senior application panel, Institut Universitaire de France
2015–2020	Editorial Board, Cognitive Science
2015–2018	Scientific board, Faculty of Sports, Toulouse
2014	ANR grant panel on learning, education, health and work
2013	Guest Editor, Argument & Computation
2013	Guest Editor, Thinking & Reasoning
2011–2015	Associate Editor, Cognition
2011–2013	ANR grant panel on human development, cognition, language and communication
2010–2017	Program Committee, Conference of the Cognitive Science Society
2010–2014	CNRS committee for social action, Toulouse
2009–2014	Director of the CLLE research institute in Toulouse (staff of 150–200)
2009–2014	Board of post-graduate cognitive studies, University of Toulouse
2007–2015	Board of the MSHS-T federative research institute, Toulouse

Career History

2018—2019	**Visiting Scientist (12 months)** (Cambridge, MA) Massachusetts Institute of Technology
2015—	**Senior CNRS Research Scientist** (Toulouse, France) Toulouse School of Economics + Toulouse School of Management
2013—2015	**Senior CNRS Research Scientist** (Toulouse, France) Research Unit 5263, Cognition, Language & Human Factors
2008	**Habilitation** French post-doctoral degree required for senior scientific positions
2004—2012	**Junior CNRS Research Scientist** (Toulouse, France) Research Unit 5263, Cognition, Language & Human Factors Research Unit 5505, Institute for Research in Computer Science
2003	**Ph.D., cognitive psychology** University of Toulouse

Research Grants

Continuous funding (24 grants) as PI or co-investigator since year 2000. Funding bodies include Microsoft, the European Union, the French Ministère de la Recherche, the French Ministère des Affaires Étrangères, the Agence Nationale de la Recherche, the Région Midi-Pyrénées, the Federal University of Toulouse, and the Centre National de la Recherche Scientifique.

Teaching

UT1	**Doctoral Program in Economics**
	Experimental Methods (2020)
	Master Program in Data Science
	Social and behavioral science for the ethics of AI (2025–)
	Master Program in Management
	Behavioral game theory (2005–2012)
	Dual-process methods for management scientists (2017)
	Reproducibility and Open Science (2018–2020)
	Undergraduate Program in Economics
	Introduction to psychology (2022–)
	Nudges (2017–2024)
ISAE	**Master Program in Aerospace Engineering**
	Rational decisions (2010–2014)
TBS	**Master Program in Business Administration**
	Dual-process rationality (2013-2016)
ENS	**Master Program in Cognitive Science**
	Reasoning (2010–2016)
UT2	**Doctoral Program in Behavioral Science**
	Introduction to LaTeX 2_ε (2006–2017)
	Introduction to French research agencies (2014–2018)
	Scientific writing (2015–2018)
	Master Program in Psychology
	Dual-process theories of cognition (2007–2015)
	Writing and publishing an experimental article (2012–2016)
	Running a survey in social psychology (2002–2003)
	Undergraduate Program in Psychology
	Psychometrics (2004-2010)
	Experimental Social Psychology (2000-2003)
	Communication and Persuasion (2001)

Doctoral and Post-Doctoral Supervision (13)

Malaurie Fauré, 2024–2025 [post-doc] Digital transformation of employees' voices in the service industry **Zoë Purcell, 2020–2023** [post-doc] Moral Artificial Intelligence **Bence Bagó, 2020–2023** [post-doc] Moral Artificial Intelligence **Maxime Derex, 2017–2019** [post-doc] Experimental programme investigating cumulative culture **Andrei Ivănescu, 2014–2017** Third-party expectations of nepotism and mating preferences from facial similarity **Marina Miranda Lery Santos, 2014–2017** Help seeking in a digital learning environment for traffic control **Stefania De Vito, 2012–2013** [post-doc] Strategic emotions in social interactions **Marco Heimann, 2009–2013** *Research Award of the Responsible Investment Forum 2011, Garrigou Prize of the Academy of Legislation 2014, Best Thesis Award of the Maison des Sciences de l'Homme et de la Société de Toulouse 2014, Thesis Award of the Responsible Investment Forum 2014* Morals and finance: Windfall gains, socially responsible investment, and compensation plans **Bastien Trémolière, 2009–2013** *Thesis Award of the Post-Graduate Program 2011, Dissertation Award of the Toulouse Academy of Sciences and Humanities 2014* The rationality of mortals: Thoughts of death impair analytic processing **Stefania Pighin, 2009–2010** [post-doc] Contextual influences in rational activities **Sylvie Leblois, 2008–2012**

Giving and using advice when interests are in conflict **Manh-Hung Nguyen, 2007–2010** A logical framework for trust-related emotions: Formal and behavioral results **Virginie Demeure, 2005–2008** Utility and facework in the interpretation of ambiguous statements.

Museums and Exhibits (15)

The Human Fair, Rotterdam New Institute (2016) · After Dark: Dangerous Ideas and Collisions, San Francisco Exploratorium (2017) · Permanent Collection, MIT (2018) · The Road Ahead: Reimagining Mobility, New York Smithsonian Design Museum (2018) · The Brain, Lisbon Calouste Gulbenkian Foundation (2019) · Paderborn Heinz Nixdorf MuseumsForum (2020) · British Science Museum (2020) · Linz Ars Electronica Center (2020) · Leuven Docville Film Festival (2021) · Musée des Arts et Métiers de Paris (2022) · Saint-Etienne Biennale Internationale de Design (2022) · We the Curious, Bristol (2022) · Quai des Savoirs de Toulouse (2024) · Freedom, MS Wissenschaft (2024) · Palais de la Découverte de Paris (2025)

Invited Seminars (66)

European Commission *Belgium* · Ministère de la Transition écologique et solidaire *France* · Harvard University *USA* · Massachusetts Institute of Technology *USA* · Princeton University *USA* · Berkeley University *USA* · Brown University *USA* · Stanford University *USA* · Future of Humanity Institute *England* · Paris School of Economics · Kurt Lewin Instituut *NL* · University of Oregon *USA* · Université du Québec à Montréal *Canada* · University of Göttingen *Germany* · Advanced Computation Laboratory, Cancer Research UK London *England* · Université Blaise Pascal, Clermont-Ferrand *France* · Trinity College, Dublin *Ireland* · University of Durham *England* · Université de Poitiers *France* · Université de Provence *France* · Université Paris Sorbonne *France* · École de Guerre Économique *France* · Queen's University Belfast *Northern Ireland* · University of Oslo *Norway* · Université Paris-8 *France* · University of Leuven *Belgium* · Université de Grenoble *France* · Labex Cortex *France* · Kingston Business School *England* · École Normale Supérieure *France* · University of Essex *England* · University of Kent *England* · Kedge Business School *France* · Université de Strasbourg *France* · Utrecht University *NL* · Amsterdam University *NL* · Sorbonne Université *France* · Université de Montpellier *France* · Beida University *China* · Chinese Academy of Science *China* · East China Normal University *China* · Fudan University *China* · Shanghai Academy of Social Sciences *China* · Shanghai Normal University *China* · Shanghai University of Science and Technology *China* · University of Zhejiang *China* · Masdar Institute of Science and Technology *United Arab Emirates* · Plaksha University *India* · Chinese University in Hong Kong

Keynote Lectures and Invited Conference Talks

44. **[Keynote]** Delegating to AI can increase dishonest behavior. (2025, June) *Conference on AI assertion.*
43. L'IA générative, un remède à la désinformation? (2024, December) *Colloque du Collège de France sur l'IA et ses Défis.*
42. The ethical dilemmas of autonomous vehicles. (2024, October) *International AI Governance Roundtable.*
41. **[Keynote]** The moral psychology of AI. (2024, June) *10th International Conference on Thinking.*
40. **[Keynote]** Public expectations for ethical AI. (2024, April) *12th biennial Postal Economics Conference.*

39. **[Keynote]** The moral psychology of AI. (2023, November) *2nd Workshop on Ethical Public Robots and Artificial Intelligence.*

38. Vers quelle société nous mènent l'IA et ses filtres? (2023, September) *Colloque du Collège de France sur l'IA et ses Défis.*

37. **[Keynote]** The moral psychology of AI: a flash review. (2023, June) *Conference on the Moral Psychology of AI.*

36. AI for good. (2023, June) *3rd Common Good Summit.*

35. Patient screening: Ethics and economics. (2023, June) *1st CEPR Health Economics Conference.*

34. **[Keynote]** From machine behavior to machine culture. (2023, May) *5th Data Science and Law Forum.*

33. Uptake and social consequences of lie-detection algorithms. (2023, April) *Workshop Autorité de contrôle prudentiel et de résolution.*

32. Will Autonomous Vehicles need road user education? (2023, January) *Conference on road user education for goal zero.*

31. **[Keynote]** The moral machine experiment. (2022, June) *International Conference on Cognitive Aircraft Systems.*

30. The psychology of human-machine interaction. (2021, April) *Microsoft Data Science and Law Forum.*

29. **[Keynote]** Behavioral data for machine ethics. (2019, November) *Workshop on Morality, Social Choice, and Artificial Intelligence.*

28. Transports innovants—le futur est déjà là. (2019, October). *Forum Innovation Intelligence Artificielle.*

27. The moral machine experiment. (2019, October) *Minds & Tech Conference.*

26. AI & ethics. (2019, September) *Tech Talks: The Politics of Cybersecurity.*

25. The massive crowdsourcing of machine ethics. (2019, May) *Global Business Ethics Symposium.*

24. Public acceptance and adoption. (2019, April)*European Union Conference on Connected and Automated Driving.*

23. The moral machine experiment. (2019, January) *Ethics and Policy Workshop : Ethics in Algorithms.*

22. The moral machine experiment. (2018, September) *Microsoft Data Science and Law Forum.*

21. **[Keynote]** The moral machine experiment. (2018, July) *27th International Joint Conference on Artificial Intelligence.*

20. Intelligence artificielle et cognition. (2018, July). *Conférence Olivier Legrain Sciences et Société.*

19. The moral machine experiment. (2017, December) *Conférence Nouvelles Pratiques du Journalisme.*

18. **[Keynote]** Ethical challenges for self-driving cars. (2017, May) *4th Conference of the Italian Society for Neuroethics.* Padova, Italie.

17. **[Keynote]** Le dilemme éthique des véhicules autonomes. (2017, January) *Décrypta-Géo 2017.*

16. Eye movements track the early stages of explaining anomalous decisions. (2015, November) *Leading Edge Workshop: The Process of Explanation.*

15. **[Keynote]** Cognitive and moral reflection across social networks. (2015, February) *Conference on Reasoning, Argumentation, and Critical Thinking Instruction.*

14. **[Keynote]** The perils of politeness. (2014, July) *28th International Congress of Applied Psychology.*

13. **[Keynote]** Reasoning about decisions. (2014, March) *Third Annual Meeting of Priority Programme SPP 1516 New Frameworks of Rationality.*

12. Psychologie expérimentale de la mort. (2013, May) *Archéologie des peuplements et des populations pré- et protohistorique.*

11. **[Keynote]** Experiments on politeness and reasoning. (2012, August) *Experimental and Empirical Approaches to Politeness and Impoliteness.*

10. Reasoning in an imperfect world. (2012, July) *International Congress of Psychology, ICP2012.*

9. **[Keynote]** Inference from preferences. (2012, July) *International Conference on Thinking, ICT2012.*

8. The causal structure of utility conditionals. (2012, May) *Workshop 'How universal is causal cognition?'*

7. Nouveaux territoires pour les neurosciences de la rationalité. (2011, June) *Colloque Annuel de l'ITMO Neurosciences.*

6. How do we aggregate collective judgments that simultaneously point to inconsistent conclusions? (2010, July) *Rethinking Problem Solving: A Symposium on Distributed Cognition.*

5. **[Keynote]** Cognitive inferences from/to preferences. (2010, July) *9th Conference on Logic and the Foundations of Game and Decision Theory.*

4. **[Keynote]** Intelligence Artificielle et psychologie. (2009, October) *Journées d'Intelligence Artificielle Fondamentale.*

3. A comprehensive theory of reasoning from utility conditionals. (2009, September) *Workshop on Conditionals, Conditional Probability, and Conditionalization.*

2. The causal analysis of traffic accident reports. (2008, March) *Conference of the Eastern Psychological Association.*

1. Deductive reasoning about the consequences of moral and immoral behaviour. (2008, November) *The Economy of the Soul: Rational Choice and Moral Decision-Making.*

Books

5. Bonnefon, J. F. (2019). *La voiture qui en savait trop: l'intelligence artificielle a-t-elle une morale?* Humensciences.

   — *The car that knew too much: Can a machine be moral?* MIT Press, 2021.

4. Bonnefon, J. F. (2017). *Reasoning unbound: Thinking about morality, delusion, and democracy.* Palgrave Macmillan.

3. Bonnefon, J. F., & Trémolière, B. (Eds). (2017). *Moral Inferences.* Hove: Psychology Press.

2. Elqayam, S., Bonnefon, J. F., & Over, D. E. (Eds). (2016). *New paradigm psychology of reasoning: Basic and Applied perspectives.* London: Routledge.

1. Bonnefon, J. F. (2011). *Le raisonneur et ses modèles.* Grenoble, France : Presses Universitaires de Grenoble.

Reports

1. Horizon 2020 Commission Expert Group to advise on specific ethical issues raised by driverless mobility (2020, chair: J.F. Bonnefon). *Ethics of Connected and Automated Vehicles: recommendations on road safety, privacy, fairness, explainability and responsibility.* Publication Office of the European Union.

Journal Articles

131. Bago, B., Muller, P., & Bonnefon, J. F. (2025). Using generative AI to increase sceptics' engagement with climate science. *Nature Climate Change* 15, 1176–1182.

130. Schmidt, E. M., Bersch, C., Köbis, N., Bonnefon, J. F., Rahwan, I., & Dong, M. (2025). First interactions with generative chatbots shape local but not global sentiments about AI. *Computers in Human Behavior: Artificial Humans* 6, 100223.

129. Köbis, N., Rahwan, Z., Rilla, R., Supriyatno, B. I., Bersch, C., Ajaj, T., Bonnefon, J. F., & Rahwan, I. (2025). Delegation to artificial intelligence can increase unethical behaviour. *Nature* 646, 126–134.

128. Rahwan, I., Shariff, A., & Bonnefon, J. F. (2025). The science fiction science method. *Nature* 644, 51–58.

127. Dong, M., Bonnefon, J. F., & Rahwan, I (2025). Heterogeneous preferences and asymmetric insights for AI use among welfare claimants and non-claimants. *Nature Communications* 16, 7973.

126. Bonnefon, J. F. (2025). Editorial to the special issue on morality and AI. *Cognition* 265, 106229.

125. Makovi, K., Bonnefon, J. F., Oudah, M., Sargsyan, A., & Rahwan, T. (2025). Rewards and punishments help humans overcome biases against cooperation partners assumed to be machines. *iScience* 28, 112833.

124. Derex, M., Bonnefon, J. F., Boyd, R., McElreath, R., & Mesoudi, A. (2025). Social learning preserves both useful and useless theories by canalizing learners' exploration. *Proceedings of the Royal Society B* 292, 20242499.

123. Dong, M., Conway, J. R., Bonnefon, J. F., Shariff, A., & Rahwan, I. (2025). Fears about Artificial Intelligence across 20 countries and six domains of application. *American Psychologist.*

122. Bonnefon, J.F., Landier, A., Sastry, P., & Thesmar, D. (2025) The moral preferences of investors: experimental evidence. *Journal of Financial Economics* 163, 103955.

121. Bago, B., & Bonnefon, J. F. (2024) Generative AI as a tool for truth. *Science* 385, 1164-1165.

120. Capraro, V., Lentsch, A., Acemoglu, D., Akgun, S., Akhmedova, A., Bilancini, E., Bonnefon, J. F., Brañas-Garza, P., Butera, L., Douglas, K. M., Everett, J. A. C., Gigerenzer, G., Greenhow, C., Hashimoto, D. A., Holt-Lunstad, J., Jetten, J., Johnson, S., Longoni, C., Lunn, P., Natale, S., Rahwan, I., Selwyn, N., Singh, V., Suri, S., Sutcliffe, J., Tomlinson, J., van der Linden, S., Van Lange, P. A. M., Wall, F., Van Bavel, J. J., & Viale, R. (2024). The impact of generative artificial intelligence on socioeconomic inequalities and policy making. *PNAS Nexus* 3, pgae058.

119. Bonnefon, J. F., Rahwan, I., & Shariff, A. (2024). The moral psychology of Artificial Intelligence. *Annual Review of Psychology* 75, 653-675.

118. von Schenk, A., Klockmann, V., Bonnefon, J. F., Rahwan, I., & Köbis, N. (2024). Lie detection algorithms disrupt the social dynamics of accusation behavior. *iScience* 27, 110201.

117. Dong, M., Bonnefon, J. F., & Rahwan, I. (2024). Toward human-centered AI management: Methodological challenges and future directions. *Technovation* 131, 102953

116. Katiyar, T., Bonnefon, J. F., Mehr, S. A., & Singh, M. (2024). Discovering the unknown unknowns of research cartography with high-throughput natural description. *Behavioral and Brain Sciences* 47, e50 [Commentary on Almaatouq et al.]

115. Brinkmann, L., Baumann, F., Bonnefon, J. F., Derex, M., Müller, T., Nussberger, A. M., Czaplicka, A., Acerbi, A., Griffiths, T., Henrich, J., Leibo, J. Z., McElreath, R., Oudeyer, P. Y., Stray, J., & Rahwan, I. (2023). Machine culture. *Nature Human Behaviour* 7, 1855–1868.

114. Makovi, K., Sargsyan, A., Li, W., Bonnefon, J. F., & Rahwan, T. (2023). Trust within human-machine collectives depends on the perceived consensus about cooperative norms. *Nature Communications* 14, 3108.

113. Bonnefon, J. F. (2023). Moral artificial intelligence and machine puritanism. *Behavioral and Brain Sciences* 46, E297. [Commentary on Fitouchi et al.]

112. Purcell, Z. A., & Bonnefon, J. F. (2023). Research on Artificial Intelligence is reshaping our definition of morality. *Psychological Inquiry* 34, 100-101 [Commentary on Dahl]

111. Purcell, Z. A., & Bonnefon, J. F. (2023). Humans feel too special for machines to score their morals. *PNAS Nexus* 2, pgad179.

110. Awad, E., Bago, B., Bonnefon, J. F., Christakis, N. A., Rahwan, I., & Shariff, A. (2022). Polarized citizen preferences for the ethical allocation of scarce medical resources in twenty countries. *Medical Decision Making Policy & Practice* 7, 23814683221113573.

109. Duch, R., Roope, L. S. J., Violato, M., Becerra, M. F., Robinson, T., Bonnefon, J. F., Friedman, J., Loewen, P., Mamidi, P., Melegaro, A., Blanco, M., Vargas, J., Seither, J., Candio, P., Cruz, A. G., Hua, X., Barnett, A., & Clarke, P. M. (2021). Citizens from 13 countries share similar preferences for COVID-19 vaccine allocation priorities. *PNAS* 38, e2026382118.

108. Köbis, N., Bonnefon, J. F., & Rahwan, I. (2021). Bad machines corrupt good morals. *Nature Human Behaviour* 5, 679–685.

107. Clarke, P. M., Roope, L. S. J., Loewen, P. J., Bonnefon, J. F., Melegaro, A., Friedman, J., Violato, M., Barnett, A., & Duch, R. (2021). Public opinion on global rollout of COVID-19 vaccines. *Nature Medicine* 27, 935–936.

106. Bago, B., Bonnefon, J. F., & De Neys, W. (2021). Intuition rather than deliberation determines selfish and prosocial choices. *Journal of Experimental Psychology: General* 150, 1081–1094.

105. Shariff, A., Bonnefon, J. F., & Rahwan, I. (2021). How safe is safe enough? Psychological mechanisms underlying extreme safety demands for self-driving cars. *Transportation Research Part C: Emerging Technologies* 126, 103069.

104. Awad, E., Dsouza, S., Shariff, A., Rahwan, I., & Bonnefon, J. F. (2020). Reply to Claessens et al.: Maybe the footbridge sacrifice is indeed the only one that sends a negative social signal. *PNAS* 117, 13205–13206.

103. Bonnefon, J. F., & Rahwan, I. (2020). Machine thinking, fast and slow. *Trends in Cognitive Sciences* 24, 1019–1027.

102. Rahwan, I., Crandall, J., & Bonnefon, J. F. (2020). Intelligent machines as social catalysts. *PNAS* 117, 7555–7557.

101. Awad, E., Dsouza, S., Shariff, A., Rahwan, I., & Bonnefon, J. F. (2020). Universals and variations in moral decisions made in 42 countries by 70,000 participants. *PNAS* 117, 2332–2337.

100. Awad, E., Levine, S., Kleiman-Weiner, M., Dsouza, S., Tenenbaum, J. B., Shariff, A., Bonnefon, J. F., & Rahwan, I. (2020). Drivers are blamed more than their automated cars when both make mistakes. *Nature Human Behaviour* 4, 134–143.

99. Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J. F., & Rahwan, I. (2020). Reply to: Life and death decisions of autonomous vehicles. *Nature* 579, E3–E5.

98. Awad, E., Dsouza, S., Bonnefon, J. F., Shariff, A., & Rahwan, I. (2020). Crowdsourcing moral machines. *Communications of the ACM* 63, 48–55.

97. Juanchich, M., Sirota, M., & Bonnefon, J. F. (2020. Anxiety-induced miscalculations, more than differential inhibition of intuition, explain the gender gap in cognitive reflection. *Journal of Behavioral Decision Making* 33, 427–443.

96. Borau, S., & Bonnefon, J. F. (2020). Gendered products act as the extended phenotype of human sexual dimorphism: They increase physical attractiveness and desirability. *Journal of Business Research* 120, 498–508.

95. Salvia, E., Mevel, K., Borst, G., Poirel, N., Simon, G., Orliac, F., Etard, O., Hopfensitz, A., Houdé, O., Bonnefon, J. F., & De Neys, W. (2020). Age-related neural correlates of facial trustworthiness detection during economic interaction. *Journal of Neuroscience, Psychology, and Economics* 13, 19–33.

94. Miranda Lery Santos, M., Tricot, A., & Bonnefon, J. F. (2020). Do learners declining to seek help conform to rational principles? *Thinking & Reasoning* 26, 87–117.

93. Ishowo-Oloko, F., Bonnefon, J. F., Soroye, Z., Crandall, J., Rahwan, I, & Rahwan, T. (2019). Behavioural evidence for a transparency-efficiency tradeoff in human-machine cooperation. *Nature Machine Intelligence* 1, 517–521.

92. Derex, M., Bonnefon, J. F., Boyd, R., & Mesoudi, A. (2019). Causal understanding is not necessary for the improvement of culturally evolving technology. *Nature Human Behaviour* 3, 446-452.

91. Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J. F., Breazeal, C., Crandall, J. W., Christakis, N. A., Couzin, I. D., Jackson, M. O., Jennings, N. R., Kamar, E., Kloumann, I., Larochelle, H., Lazer, D., McElreath, R., Mislove, A., Parkes, D. C., Pentland, A., Roberts, M. E., Shariff, A., Tenenbaum, J. B., & Wellman, M. (2019). Machine behaviour. *Nature* 568, 477–486.

90. Bonnefon, J. F., Shariff, A., & Rahwan, I. (2019). The trolley, the bull bar, and why engineers should care about the ethics of autonomous cars. *Proceedings of the IEEE* 107, 502–504.

89. Borau, S., & Bonnefon, J. F. (2019). The imaginary intrasexual competition: Advertisements featuring provocative female models trigger women to engage in indirect aggression. *Journal of Business Ethics* 157, 45–63.

88. Juanchich, M., Sirota, M., & Bonnefon, J. F. (2019). The polite wiggle-room effect in charity donation decisions. *Journal of Behavioral Decision Making* 32, 179–193.

87. Sirota, M., Juanchich, M., & Bonnefon, J F. (2018). '1-in-X' bias: '1-in-X' format causes overestimation of health-related risks. *Journal of Experimental Psychology: Applied* 24, 431–439.

86. Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.F., & Rahwan, I. (2018). The moral machine experiment. *Nature* 563, 59–64.

85. Bonnefon, J. F. (2018). The pros and cons of identifying critical thinking with System 2 processing. *Topoi* 37, 113–119.

84. Crandall, J. W., Oudah, M., Tennom, Ishowo-Oloko, F., Abdallah, S., Bonnefon, J. F., Cebrian, M., Shariff, A., Goodrich, M. A., & Rahwan, I. (2018). Cooperating with machines. *Nature Communications, 9,* 233.

83. Shariff, A., Bonnefon, J. F., & Rahwan, I. (2017). Psychological roadblocks to the adoption of self-driving vehicles. *Nature Human Behaviour* 1, 694–696.

82. De Neys, W., Hopfensitz, A., & Bonnefon, J. F. (2017). Split-second trustworthiness detection from faces in an economic game. *Experimental Psychology* 64, 231–239.

81. Bonnefon, J. F., Hopfensitz, A., & De Neys, W. (2017). Can we detect cooperators by looking at their face? *Current Directions in Psychological Science* 26, 276–281.

80. Awad, E., Bonnefon, J. F., Caminada, M., Malone, T., & Rahwan, I. (2017). Experimental assessment of aggregation principles in argumentation-enabled collective intelligence. *ACM Transactions on Internet Technology* 17, #29.

79. Bonnefon, J. F., Hopfensitz, A., & De Neys, W. (2017). Trustworthiness perception at zero acquaintance: consensus, accuracy, and prejudice. *Behavioral and Brain Sciences* 40, 24–25 [Commentary on Jussim].

78. Bonnefon, J. F., Heimann, M., & Lobre-Lebraty, K. (2017). Value similarity and overall performance: Trust in responsible investment. *Society and Business Review* 12, 200–215.

77. Borau, S., & Bonnefon, J. F. (2017). The advertising performance of non-ideal female models as a function of viewers' body mass index: a moderated mediation analysis of two competing affective pathways. *International Journal of Advertising* 36, 457–476.

76. Bonnefon, J. F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science* 352, 1573–1576.

75. Sloman, S. A, Kim, A., Bonnefon, J. F., Wagemans, J., Frank, M. C., Arnold, J. E., Murphy, G., Tsakiris, M., Feldman, J., Lourenco, S. F., & Wynn, K. (2016). Introducing a fund for open-access fees. *Cognition* 154, iii-iv. [Editorial]

74. Bonnefon, J. F., Dahl, E., & Holtgraves, T. M. (2015). Some but not all dispreferred turn markers help to interpret scalar terms in polite contexts. *Thinking and Reasoning* 21, 230–249.

73. Bonnefon, J. F., Hopfensitz, A. & De Neys, W. (2015). Face-ism and kernels of truth in facial inferences. *Trends in Cognitive Sciences* 19, 421–422.

72. Bouvet, R., & Bonnefon, J.F. (2015). Non-reflective thinkers are predisposed to attribute supernatural causation to uncanny experiences. *Personality and Social Psychology Bulletin* 41, 955–961.

71. De Neys, W., Hopfensitz, A., & Bonnefon, J. F. (2015). Adolescents gradually improve at detecting trustworthiness from the facial features of unknown adults. *Journal of Economic Psychology* 47, 17–22.

70. De Vito, S., Buonocore, A., Bonnefon, J. F., & Della Sala, S. (2015). Eye movements disrupt episodic future thinking. *Memory* 23, 796–805.

69. Haigh, M., & Bonnefon, J.F. (2015a). Conditional sentences create a blind spot in theory of mind during narrative comprehension. *Acta Psychologica* 160, 194–201.

68. Haigh, M., & Bonnefon, J.F. (2015b). Eye movements reveal how readers infer intentions from the beliefs and desires of others. *Experimental Psychology* 62, 206–213.

67. Heimann, M., Bonnefon, J. F., & Mullet, E. (2015). People's views about the acceptability of remuneration policies and executive bonuses. *Journal of Business Ethics* 127, 661–671.

66. Trémolière, B., Kaminski, G., & Bonnefon, J. F. (2015). Intrasexual competition shapes men's anti-utilitarian moral decisions. *Evolutionary Psychological Science* 1, 18–22.

65. De Vito, S., & Bonnefon, J. F. (2014). People believe each other to be selfish hedonic maximizers. *Psychonomic Bulletin & Review* 21, 1331–1338.

64. De Vito, S., Buonocore, A., Bonnefon, J. F., & Della Sala, S. (2014). Eye movements disrupt spatial but not visual mental imagery. *Cognitive Processing* 15, 543–549.

63. Heimann, M., Mullet, E., & Bonnefon, J.F. (2014). Legitimacy of executive compensation plans: A preliminary study of French laypersons' acceptability. *Psicologica* 35, 543–558.

62. Rahwan, I., Krasnoshtan, D., Shariff, A., & Bonnefon, J. F. (2014). Analytical reasoning task reveals limits of social learning in networks. *Journal of the Royal Society Interface* 20131211.

61. Trémolière, B., & Bonnefon, J. F. (2014). Efficient kill-save ratios ease up the cognitive demands on counterintuitive moral utilitarianism. *Personality and Social Psychology Bulletin* 40, 923–930.

60. Trémolière, B., De Neys, W., & Bonnefon, J. F. (2014). The grim reasoner: Analytic reasoning under mortality salience. *Thinking and Reasoning* 20, 333–351.

59. Bonnefon, J. F. (2013). Formal models of reasoning in cognitive psychology. *Argument and Computation* 4, 1–3.

58. Bonnefon, J. F. (2013). New ambitions for a new paradigm: Putting the psychology of reasoning at the service of humanity. *Thinking and Reasoning* 19, 381–398.

    - Reproduced in S. Elqayam, J. F. Bonnefon, & D. E. Over (Eds.), *New paradigm psychology of reasoning*, London: Routledge, 2016.

57. Bonnefon, J.F., Girotto, V., Heimann, M., & Legrenzi, P. (2013). Can mutualistic morality predict how individuals deal with benefits they did not deserve? *Behavioral and Brain Sciences* 36, 83. [Commentary on Baumard, André & Sperber.]

56. Bonnefon, J. F., Haigh, M., & Stewart, A. J. (2013). Utility templates for the interpretation of conditional statements. *Journal of Memory and Language* 68, 350–361.

55. Bonnefon, J. F., Hopfensitz, A., & De Neys, W. (2013). The modular nature of trustworthiness detection. *Journal of Experimental Psychology: General* 142, 143–150.

54. Bonnefon, J. F., & Sloman, S. A. (2013). The causal structure of utility conditionals. *Cognitive Science* 37, 193–209.

53. De Neys, W., & Bonnefon, J. F. (2013). The whys and whens of individual differences in thinking biases. *Trends in Cognitive Sciences* 17, 172–178.

52. De Neys, W., Hopfensitz, A., & Bonnefon, J. F. (2013). Low second-to-fourth digit ratio predicts indiscriminate social suspicion, not improved trustworthiness detection. *Biology Letters* 9, 20130037.

51. Feeney, A., & Bonnefon, J. F. (2013). Politeness and honesty contribute additively to the interpretation of scalar expressions. *Journal of Language and Social Psychology* 32, 181-190.

50. Heimann, M., Girotto, V., Legrenzi, P., & Bonnefon, J. F. (2013). Decision makers use norms, not cost-benefit analysis, when choosing to conceal or reveal unfair rewards. *PLoS One* 8, e73223.

49. Leblois, S. & Bonnefon, J. F. (2013). People are more likely to be insincere when they are more likely to accidentally tell the truth. *Quarterly Journal of Experimental Psychology, 66* 1486–1492.

48. Bonnefon, J. F., Da Silva Neves, R. M., Dubois, D., & Prade, H. (2012). Qualitative and quantitative conditions for the transitivity of perceived causation: Theoretical and experimental results. *Annals of mathematics and Artificial Intelligence* 64, 311–333.

47. Bonnefon, J. F. (2012). Utility conditionals as consequential arguments: A random sampling experiment. *Thinking & Reasoning* 18, 379–393.

46. Bonnefon, J. F., Girotto, V., & Legrenzi, P. (2012). The psychology of reasoning about preferences and unconsequential decisions. *Synthese* 187, 27–41.

45. Trémolière, B., De Neys, W., & Bonnefon, J. F. (2012). Mortality salience and morality: Thinking about death makes people less utilitarian. *Cognition* 124, 379–384.

44. Bonnefon, J. F. (2011). Norms for reasoning about decisions. *Behavioral and Brain Sciences* 34, 249–250 [Commentary on Elqayam & Evans]

43. Bonnefon, J. F. (2011). The doctrinal paradox, a new challenge for behavioral psychologists. *Advances in Psychological Science* 19, 617–623.

42. Bonnefon, J. F., Feeney, A., & De Neys, W. (2011). The risk of polite misunderstandings. *Current Directions in Psychological Science* 20, 321–324.

41. Bonnefon, J. F., & Politzer, G. (2011). Pragmatics, mental models, and one paradox of the material conditional. *Mind & Language* 26, 141–155.

40. Pighin, S., & Bonnefon, J. F. (2011). Facework and uncertain reasoning in health communication. *Patient Education and Counseling* 85, 169–172.

39. Pighin, S., Bonnefon, J. F., & Savadori, L. (2011). Overcoming number numbness in prenatal risk communication. *Prenatal Diagnosis* 31, 809–813.

38. Pighin, S., Savadori, L., Barilli, E., Cremonesi, L., Ferrari, M., & Bonnefon, J. F. (2011). The '1 in X' effect on the subjective assessment of medical probabilities. *Medical Decision Making* 31, 721–729.

37. Zhang, J, Bonnefon, J. F., & Deng, C. (2011). Zhi zai zhong guo ren de jia she si wei zhong de jue se. [On the role of zhi in Chinese counterfactual thinking.] *Acta Psychologica Sinica* 43, 1–10.

36. Bonnefon, J. F. (2010). Deduction from if-then personality signatures. *Thinking and Reasoning* 16, 157–171.

35. Bonnefon, J. F. (2010). Behavioral evidence for framing effects in the resolution of the doctrinal paradox. *Social Choice and Welfare* 34, 631–641.

34. Bonnefon, J. F., & Vautier, S. (2010). Modern psychometrics for the experimental psychology of reasoning. *Acta Psychologica Sinica* 42, 99-110.

33. Ben-Naim, J., Bonnefon, J. F., Herzig, A., Leblois, S., & Lorini, E. (2010). Computer-mediated trust in self-interested expert recommendations. *AI & Society* 25, 413–422.

    - Reproduced in S. Cowley & F. Vallee-Tourangeau (Eds.), *Cognition beyond the brain: Computation, interactivity and human artifice,* Berlin: Springer-Verlag, 2013.

32. Rahwan, I., Madakkatel, M. I., Bonnefon, J. F., Awan, R. N., & Abdallah, S. (2010). Behavioural experiments for assessing the abstract argumentation semantics of reinstatement. *Cognitive Science* 34, 1483–1502.

31. Bonnefon, J. F. (2009). A theory of utility conditionals: Paralogical reasoning from decision-theoretic leakage. *Psychological Review* 116, 888–907.

30. Bonnefon, J. F., Feeney, A., & Villejoubert, G. (2009). When some is actually all: Scalar inferences in face-threatening contexts. *Cognition* 112, 249–258.

29. Bonnefon, J. F., Longin, D., & Nguyen, M. H. (2009). A logical framework for trust-related emotions. *Electronic Communications of the EASST* 22.

28. Demeure, V., Bonnefon, J. F., & Raufaste, É. (2009). Politeness and conditional reasoning: Interpersonal cues to the indirect suppression of deductive inferences. *Journal of Experimental Psychology: Learning, Memory, & Cognition* 35, 260–266.

27. Martinez, F., Bonnefon, J. F., & Hoskens, J. (2009). Active involvement, not illusory control, increases risk taking in a gambling game. *Quarterly Journal of Experimental Psychology* 62, 1063–1071.

26. Politzer, G., & Bonnefon, J. F. (2009). Let us not put the probabilistic cart before the uncertainty bull. *Behavioral and Brain Sciences* 32, 100–101. [Commentary on Oaksford & Chater]

25. Bonnefon, J. F. (2008). Two routes for bipolar information processing, and a blind spot in between. *International Journal of Intelligent Systems* 23, 1–7.

24. Bonnefon, J. F., Da Silva Neves, R. M., Dubois, D., & Prade, H. (2008). Predicting causality ascriptions from background knowledge: Model and experimental validation. *International Journal of Approximate Reasoning* 48, 752–765.

23. Bonnefon, J. F., Dubois, D., Fargier, H., & Leblois, S. (2008). Qualitative heuristics for balancing the pros and cons. *Theory & Decision* 65, 71–95.

22. Bonnefon, J. F., Eid, M., Vautier, S., & Jmel, S. (2008). A mixed Rasch model of dual-process conditional reasoning. *Quarterly Journal of Experimental Psychology* 61, 809–824.

21. Bonnefon, J. F., & Vautier, S. (2008). Defective truth tables and falsifying cards: Two measurement models yield no evidence of an underlying fleshing-out propensity *Thinking & Reasoning* 14, 231–243.

20. Bonnefon, J. F., & Zhang, J. (2008). The intensity of recent and distant life regrets: An integrated model and a large scale survey. *Applied Cognitive Psychology* 22, 653-662.

19. Demeure, V., Bonnefon, J. F., & Raufaste, É. (2008). Utilitarian relevance and face-management in the interpretation of ambiguous question/request statements. *Memory & Cognition* 36, 873–881.

18. Dubois, D., Fargier, H., & Bonnefon, J. F. (2008). On the qualitative comparison of decisions having positive and negative features. *Journal of Artificial Intelligence Research* 32, 385–417.

17. Vautier, S., & Bonnefon, J. F. (2008). Is the above-average effect measurable at all? The validity of the self-reported happiness minus other's perceived happiness construct. *Applied Psychological Measurement* 32, 575–584.

16. Bonnefon, J. F. (2007). How do individuals solve the doctrinal paradox in collective decisions? An empirical investigation. *Psychological Science* 18, 753–755.

15. Bonnefon, J. F. (2007). Reasons to act and the mental representation of consequentialist aberrations. *Behavioral and Brain Sciences* 30, 454–455. [Commentary on Byrne]

14. Bonnefon, J. F., Vautier, S., & Eid, M. (2007). Modeling individual differences in contrapositive reasoning with continuous latent state and trait variables. *Personality and Individual Differences* 42, 1403–1413.

13. Bonnefon, J. F., & Villejoubert, G. (2007). Modus Tollens, Modus Shmollens: Contrapositive reasoning and the pragmatics of negation. *Thinking & Reasoning* 13, 207–222.

12. Bonnefon, J. F., Zhang, J., & Deng, C. (2007). L'effet des justifications sur le regret est-il direct ou indirect? *Revue Internationale de Psychologie Sociale* 20, 131–145.

11. Bonnefon, J. F., & Villejoubert, G. (2006). Tactful or doubtful? Expectations of politeness explain the severity bias in the interpretation of probability phrases. *Psychological Science* 17, 747–751.

10. Politzer, G., & Bonnefon, J. F. (2006). Two varieties of conditionals and two kinds of defeaters help reveal two fundamental types of reasoning. *Mind & Language* 21, 484–503.

9. Raufaste, É., Longin, D., & Bonnefon, J. F. (2006). Utilitarisme pragmatique et reconnaissance d'intention dans les actes de langage indirects. *Psychologie de l'Interaction* 21/22, 189–209.

8. Benferhat, S., Bonnefon, J. F., & Da Silva Neves, R. (2005). An overview of possibilistic handling of default reasoning, with experimental studies. *Synthese* 146, 53–70.

7. Hilton, D. J., Kemmelmeier, M., & Bonnefon, J. F. (2005). Putting ifs to work: Goal-based relevance in conditional directives. *Journal of Experimental Psychology: General* 135, 388–405.

6. Hilton, D. J., Villejoubert, G., & Bonnefon, J. F. (2005). How to do things with logical expressions: Creating collective value through co-ordination. *Interaction Studies* 6, 103–117.

5. Zhang, J., Walsh, C., & Bonnefon, J. F. (2005). Between-subject or within-subject measures of regret: dilemma and solution. *Journal of Experimental Social Psychology* 41, 559–566.

4. Bonnefon, J. F. (2004). Reinstatement, floating conclusions, and the credulity of mental model reasoning. *Cognitive Science* 28, 621–631.

3. Bonnefon, J. F., & Hilton, D. J. (2004). Consequential conditionals: Invited and suppressed inferences from valued outcomes. *Journal of Experimental Psychology: Learning, Memory, & Cognition* 30, 28–37.

2. Bonnefon, J. F., & Hilton, D. J. (2002). The suppression of Modus Ponens as a case of pragmatic preconditional reasoning. *Thinking & Reasoning* 8, 21–40.

1. Da Silva Neves, R., Bonnefon, J. F., & Raufaste, É. (2002). An empirical test for patterns of nonmonotonic inference. *Annals of Mathematics and Artificial Intelligence* 34, 107–130.