Jean-François Bonnefon

jfbonnefon@gmail.com
http://jfbonnefon.github.io

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

Emploi actuel

2015—	**Directeur de recherche CNRS** Toulouse School of Economics + Toulouse School of Management

Responsabilités

2024–	Comité d'organisation, 2026 World Conference on Transport Research
2023–	Conseil de Laboratoire, Toulouse School of Management Research
2023–	Vice-Président du Comité d'éthique, Toulouse School of Economics
2022–	Advisory Board, Data Intelligence Institute of Paris
2021–	Leader de l'équipe Expanded Morality, Toulouse School of Economics
2019–	Chaire de Moral Artificial Intelligence, Institut 3IA ANITI, Toulouse
2018–	Leader du programme AI & Society, TSE Digital Center
2011–	Comité de pilotage, Institute for Advanced Study in Toulouse
2011–	Comité Éditorial, Thinking & Reasoning
2024	Éditeur Invité, Cognition
2021–2023	Comité scientifique, exposition IA Double Je, Paris, Toulouse
2022–2023	Conseiller, European Commission initiative on meaningful and ethical communication
2021–2023	Directeur Scientifique, Institute for Advanced Study in Toulouse
2021–2022	Comité de programme, Common Good Summit, Toulouse
2020–2021	Directeur du programme scientifique, 9th International Conference on Thinking
2019–2021	Comité de recrutement, Institute for Advanced Study in Toulouse
2013–2021	Comité National Français de Psychologie Scientifique
2019–2020	Président, European Commission expert group on the ethics of driverless mobility
2016–2020	Président du Comité d'éthique, Toulouse School of Economics
2011–2020	Conseil scientifique du Labex TULIP, Toulouse
2016–2018	Comité National de la Recherche Scientifique (Section 26, membre du bureau)
2016–2018	Comité de pilotage de l'Institut d'Étude des Systèmes Complexes de Toulouse
2016–2017	Éditeur Associé Senior, Cognition
2015–2018	Commission Scientifique de la Faculté des Sciences du Sport, Toulouse
2011–2018	Comités de visite AERES/HCERES (Président 2/4)
2016	Jury senior de l'Institut Universitaire de France
2015–2020	Comité Éditorial, Cognitive Science
2014	Comité d'évaluation ANR CES 30 Collaboratif/Jeunes Chercheurs
2013	Éditeur Invité, Argument & Computation
2013	Éditeur Invité, Thinking & Reasoning
2011–2015	Éditeur Associé, Cognition

Responsabilités, suite

2011–2013	Comité d'évaluation ANR SHS2 Blanc/Jeunes Chercheurs
2010–2017	Comité de Programme, Annual Conference of the Cognitive Science Society
2010–2014	Commission régionale d'action sociale du CNRS, Toulouse
2010–2014	Commission régionale d'interclassement ITA (BAP D, BAP F, BAP J)
2009–2014	Directeur d'unité, UMR5263 CLLE (150–200 agents)
2009–2014	Conseil de l'école doctorale CLESCO, Toulouse
2007–2015	Conseil de la Maison des SHS, Toulouse

Carrière et diplômes

2018—2019	**Chercheur invité** (12 mois) Massachusetts Institute of Technology, Cambridge, MA
2015—	**Directeur de recherche CNRS** Toulouse School of Economics + Toulouse School of Management
2013—2014	**Directeur de recherche CNRS** UMR5263, Toulouse (Cognition, Langues, Langage, Ergonomie)
2008	**Habilitation à diriger des recherches** Université Toulouse-2
2004—2012	**Chargé de recherche CNRS** UMR5263, Toulouse (Cognition, Langues, Langage, Ergonomie) UMR5505, Toulouse (Institut de Recherche en Informatique de Toulouse)
2003	**Doctorat en psychologie cognitive** Université Toulouse-2

Enseignement

ISAE	**Cursus ingénieur ENSICA**	
	La rationalité des décisions (2010-2014)	
TBS	**Master Commerce**	
	L'approche dual-process de la rationalité (2013-2016)	
ENS	**Cogmaster**	
	Raisonnement (2010–2016)	
Toulouse-2	**Doctorat Science du Comportement**	
	Initiation à LaTEX 2ε (2006–2017)	
	Le concours, la carrière et l'institution CNRS (2014–2018)	
	Écriture scientifique (2015–2018)	
	Master Psychologie	
	Théories Dual-Process	
	(2007–2015) Stratégie de	
	Publication (2012–2016) Enquête	
	Psychosociale (2002-2003)	
	Licence Psychologie	
	Psychométrie (2004-2010)	
	Psychologie Sociale (2000-2003)	
	Communication et Influence (2001)	
Toulouse-1	**Doctorat Économie**	
	Experimental methods (2020)	
	Master Management	
	Behavioral game theory	
	(2005–2012) Dual-process methods	
	(2017)	
	Reproducibility and Open Science (2018–2020)	
	Licence Économie	
	Introduction à la psychologie (2022–)	
	Nudges et psychologie de la décision	
	(2017–)	

Distinctions

2016–	**Outstanding Vitae Award**	TSE, renouvelé en 2022
2009–	**Prime d'Excellence Scientifique**	CNRS, renouvellée en 2013 et
2017 2009	**Lauriers de la Recherche**	Ville de Toulouse
2008	**Médaille de bronze du CNRS**	CID45

Expertises

Projets British Academy 2022 ERC, advanced grants 2016 Israel Science Foundation, personal grants 2019 Agence Nationale de la Recherche, jeunes chercheuses et jeunes chercheurs 2006/2009, blanc 2007/2010, retour post-doc 2012, recherche duale 2014/2017, générique 2019 · Biotechnologies and Biological Sciences Research Council (UK), basic grants 2009 · Canada Council for the Arts, Killam fellowships program 2009 · Leverhulme Trust (UK), grant program 2012 · Fond National Suisse pour la Science, programme Ambizione 2012, programme général 2014 · Villa Vigoni (Allemagne-France-Italie), ateliers trilatéraux 2014 · PSL*, chaires d'excellences 2013 · Région Aquitaine, appel à pro- jet recherche 2013 · Bourses Fernand Braudel 2014 · Netherlands Organisation for Scientific Research, innovational research incentives scheme 2016, Vidi program 2017 · Canadian Institute for Advanced Research, program review

2016 **Livres et Chapitres** MIT Press (USA) Oxford University Press (UK) Psychology Press (UK) College Publi- cations (UK) Springer (Germany) Palgrave Macmillan (USA) **Articles** *Acta Psychologica Argument & Computation Artificial Intelligence & Society Canadian Journal of Experimental Psychology Cognition Cognitive Psychology Cognitive Science Communications Psychology Current Directions in Psychological Science Current Psychology Letters Economics & Philosophy Ethics and Information Technology European Journal of Cognitive Psychology European Journal of Social Psychology Evolution and Human Behavior Experimental Psychology International Journal of Intelligent Systems Inter- national Review of Social Psychology Journal of Applied Non-Classical Logics Journal of Autonomous Agents and Multiagent Systems Journal of Behavioral Decision Making Journal of Business Ethics Journal of Economic Behavior & Organization Journal of Experimental Psychology : General Journal of Experimental Psychology : Learning, Memory, & Cognition Journal of Experimental Social Psychology Journal of Memory & Language Judgment and Decision Making Learning and Indivi- dual Differences Memory & Cognition Methodology Mind & Society Motivation & Emotion Nature Nature Human Behaviour Nature Machine Intelligence Personality and Individual Differences Philosophical Psychology Philosophical Transactions of the Royal Society B PLoS One PNAS Pragmatics & Cognition Proceedings of the IEEE Psychological Research Psychological Science Psychologie Française Pyschology & Marketing Psychonomic Bulletin and Review Public Opinion Quarterly Quarterly Journal of Experimental Psychology Review of Economic Studies Review of Philosophy and Psychology Risk Analysis Scientific Reports Sex Roles Social Psychological and Personality Science Spanish Journal of Psychology Synthese Thinking & Reasoning Trends in Cognitive Sciences* **Conférences** Annual Conference of the Cognitive Science Society (membre du comité de programme 2010–2017) Congrès Annuel de la Société Française de Psychologie (membre du comité de programme 2009) EuroAsianPacific Joint Conference on Cognitive Science European Conference on Artificial Intelligence (membre du comité de programme 2008) Information Processing and Management of Uncertainty Modèles Formels de l'Interaction (membre du comité de programme 2007–2013) Theoretical Aspects of Rationality and Knowledge International Joint Conference on Artificial Intelligence XPRAG Experimental Pragmatics Conference (membre du comité de programme 2019) NeurIPS (membre du comité de pro- gramme Workshop AI-psych-phil 2023)

Musées et expositions (11)

The Human Fair, Rotterdam New Institute (2016) · After Dark : Dangerous Ideas and Collisions, San Francisco Explo- ratorium (2017) · Permanent Collection, MIT (2018) · The Road Ahead : Reimagining Mobility, New York Smithso- nian 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) · Biennale Internationale de Design de Saint-Etienne (2022)

Jurys de thèse (12) et HDR (5)

Thèses Cardiff University, UK (Adam Harris, 2009) · Université de Genève, Suisse (Caroline Gauffroy, 2011) · Aix- Marseille Université (Clément Belletier, 2013) · Université de Montpellier (Julien Barthes, 2014 ; Jeanne Bovet, 2014)
· Université de Grenoble (Aaron Duke, 2015) · Université de Lyon (Emmanuel Trouche, 2016) · École Pratique des Hautes Études (Jean-Luc Stilgenbauer, 2016) · Université Pierre et Marie Curie (Lou Safra, 2017 ; Sébastien Lefort, 2017) · Université Paris-Descartes (Darren Frey, 2017) · Université de Toulouse (Florence Lespiau, 2017) **HDR** Uni- versité Paris-Descartes (Wim De Neys, 2013) · École Pratique des Hautes Études (Jean Baratgin, 2013) · Université de Lyon (Hugo Mercier, 2015) · Université de Toulouse (Gwenaël Kaminski, 2015 ; Astrid Hopfensitz, 2017)

Direction de thèses (7) et post-doctorats (5)

Zoë Purcell, 2020–2023 *post-doctorat* Moral Artificial Intelligence **Bence Bagó, 2020–2023** *post-doctorat* Moral Artificial Intelligence **Maxime Derex, 2017–2019** *post-doctorat* Experimental Programme Investigating Cumula- tive Culture (avec Alex Mesoudi) **Andrei Ivănescu, 2014–2017** Third-party expectations of nepotism and mating preferences from facial similarity (avec Astrid Hopfensitz et Gwenaël Kaminski) **Marina Miranda Lery Santos, 2014–2017** Décision de recherche d'aide en environnement numérique d'apprentissage du contrôle aérien (avec André Tricot) **Stefania De Vito, 2012–2013** *post-doctorat* Émotions stratégiques et interaction sociale **Marco Heimann, 2009–2013** *Prix de la Recherche PRI-FIR 2011, Prix Novela de la Ville de Toulouse 2012, Prix Garrigou de l'aca- démie de législation 2014, Prix de la Maison des Sciences de l'Homme et de la Société de Toulouse 2014, Prix de Thèse PRI-FIR 2014* Morals and Finance : Windfall gains, socially responsible investment, and compensation plans (avec Sébastien Pouget et Étienne Mullet) **Bastien Trémolière, 2009–2013** *Prix de Thèse ED CLESCO 2011, Prix de l'Académie des Sciences et Belles Lettres de Toulouse 2014* La rationalité des mortels : les pensées de mort perturbent la pensée analytique (avec Wim De Neys) **Stefania Pighin, 2009–2010** *post-doctorat* Influences contextuelles dans les activités ratio- nelles **Sylvie Leblois, 2008–2012** Donner et utiliser des conseils en situation de conflit d'intérêts **Manh-Hung Nguyen, 2007–2010** A logical framework for trust-related emotions : Formal and behavioral results (avec Domi- nique Longin) **Virginie Demeure, 2005–2008** Joindre l'utile au désagréable : Rôle de l'utilité et de la gestion de la face dans l'interprétation d'énoncés ambigus (avec Éric Raufaste)

Comités de thèses (4)

Eva Schmidt, 2023– Max Planck Institute, Berlin **Vanessa Houze-Cerfon, 2023–** Université de Toulouse **Clara Bersch, 2022–** Max Planck Institute, Berlin **Léo Fitouchi, 2021–2023** École Normale Supérieure, Paris

Séminaires sur invitation (62)

Amérique Berkeley University · Brown University · Harvard University · Massachusetts Institute of Technology · Princeton University · University of Oregon · Stanford University · Université du Québec à Montréal **Europe** Eu- ropean Political Strategy Centre, European Commission *Belgique* · Ministère de la Transition écologique et solidaire *France* · Kurt Lewin Instituut *Pays-Bas* Advanced Computation Laboratory, Cancer Research UK, Londres *Angleterre*

· Paris School of Economics · Université Blaise Pascal, Clermont-Ferrand *France* · Trinity College, Dublin *Irlande* · Université de Göttingen *Allemagne* · University of Durham *Angleterre* · Université de Poitiers *France* · Université de Provence *France* · Université Paris Sorbonne *France* · École de Guerre Économique *France* · Queen's University Belfast *Irlande du Nord* · Future of Humanity Institute, Université d'Oxford *Angleterre* · Université d'Oslo *Norvège* · Université Paris-8 *France* · Université de Leuven *Belgique* · Université de Grenoble *France* · Labex Cortex *France* · Kingston Business School *Angleterre* · École Normale Supérieure *France* · University of Essex *Angleterre* · University of Kent *Angleterre* · Université de Strasbourg *France* · Université d'Utrecht *Pays-Bas* · Université d'Amsterdam *Pays-Bas*

· Sorbonne Université *France* · Université de Montpellier *France* **Asie** Masdar Institute of Science and Technology *Émirats Arabes Unis* · Plaksha University *India* · Beida University *Chine* · Chinese Academy of Science *Chine* · East China Normal University *Chine* · Fudan University *Chine* · Shanghai Academy of Social Sciences *Chine* · Shanghai Normal University *Chine* · Shanghai University of Science and Technology *Chine* · University of Zhejiang *Chine*

Recherches financées

23. [Partenaire] Transformations numériques de la voix des salariés dans les entreprises de service *Programme de Recherche JCJC ANR (2022-2026).*

22. [Partenaire] In-group favoritism in intergroup conflicts *Programme de Recherche Collaborative ANR (2023-2027).*

21. [Partenaire] Uncovering and mitigating the detrimental effects of social learning on the discovery of new solutions *Programme ANR JCJC (2021-2025).*

20. **[Porteur de projet]** Moral Artificial Intelligence. *Chaire 3IA (2019–).*

19. **[Porteur de projet]** The morality of AI-based agents. *Microsoft Corporation (2018-2020).*

18. [Partenaire] Diagnostiquer les différences individuelles dans la détection du biais au cours du raisonnement. *Programme de Recherche Collaborative ANR (2016-2020).*

17. [Partenaire] Parenté, familiarité et cognition morale. *Bourse de thèse de la COMUE Toulouse (2014–2017).*

16. [Partenaire] K2MC : Kin, Kith, and Moral Cognition. *Programme ANR JCJC (2013–2016).*

15. [Partenaire] EmoTES : Les émotions dans l'interaction sociale. *Programme ANR Émotions (2011-2013).*

14. [Partenaire] STAC : Strategic Conversation. *ERC advanced grant obtenu par Nicholas Asher (2011–2016).*

13. **[Porteur de projet]** Uncertainties : Communication verbale de la source interne ou externe de l'incertitude. *PHC Aurora Norvège (2010–2011).*

12. **[Porteur de projet]** FaceTrust. *PHC Tournesol Belgique (2010–2011).*

11. [Partenaire] COMSOC : Computational social choice. *Programme blanc de l'ANR (2009–2012).*

10. **[Porteur de projet]** Experts Imprécis : Utilisation de recommandations expertes vagues en situation de conflit d'intérêt. *Programme ISCC (2009).*

9. **[Porteur de projet]** Restaurer la confiance des investisseurs : Une exploration expérimentale systématique de l'effet des politiques d'investissement éthique. *PRES Toulouse (2009–2012).*

8. **[Porteur de projet]** Rationality : Influences contextuelles et différences individuelles dans les activités ration- nelles. *Programme Jeunes Chercheuses et Jeunes Chercheurs de l'ANR (2007–2010).*

7. **[Porteur de projet]** La dynamique des émotions et de la confiance dans les systèmes d'intelligence ambiante. *Région Midi-Pyrénées (2007–2010).*

6. [Partenaire] Politeness expectations cancel the scalar implicature from Some. *PHC Alliance Angleterre (2007–2008).*

5. [Partenaire] MICRAC : Modèles informatiques et cognitifs du raisonnement causal. *Programme blanc de l'ANR (2005–2008).*

4. **[Porteur de projet]** Le rôle des justifications dans la pensée contrefactuelle et les regrets, au laboratoire et dans la vie. *Programme de coopération du consulat général de France à Shanghai (2004–2007).*

3. [Partenaire] Modélisation du raisonnement causal et de l'explication des actions. *Programme interdisciplinaire CNRS 'Traitement des connaissances, apprentissage et NTIC' (2003–2005).*

2. [Partenaire] Modélisation computationnelle de l'apprentissage de la pragmatique conversationnelle dans l'archi- tecture cognitive ACT-R. *Programme interdisciplinaire CNRS 'Traitement des connaissances, apprentissage et NTIC' (2003–2005).*

1. [Partenaire] Modélisation qualitative des préférences et de la prise de décision humaines sous incertitude. *Pro- gramme Cognitique du CNRS (2001–2003).*

Livres

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.

Rapports

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.

Articles

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

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 [Commentaire sur 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.*

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. [Commentaire sur 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. [Commentaire sur 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 prefe- rences 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 pro- social 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 deci- sions 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. *Com- munications 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 dimor- phism : 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 evi- dence 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 im- provement 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 provo- cative 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 aggre- gation 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 [Commentaire sur 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. [Éditorial]

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 trustworthi- ness 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 poli- cies 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 counterin- tuitive 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.
    — Reproduit dans 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 indivi- duals deal with benefits they did not deserve? *Behavioral and Brain Sciences* 36, 83. [Commentaire sur Baumard, André & Sperber.]

56. Bonnefon, J. F., Haigh, M., & Stewart, A. J. (2013). Utility templates for the interpretation of conditional state- ments. *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 acci- dentally 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. *Thin- king & Reasoning* 18, 379–393.

46. Bonnefon, J. F., Girotto, V., & Legrenzi, P. (2012). The psychology of reasoning about preferences and uncon- sequential 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 [Com- mentaire sur Elqayam & Evans, Substracting ought from is.]

43. Bonnefon, J. F. (2011). The doctrinal paradox, a new challenge for behavioral psychologists. *Advances in Psycho- logical 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 communica- tion. *Prenatal Diagnosis* 31, 809–813.
— Sélectionné dans *Obstretrical & Gynecological Survey* 2011, 66, 735–736.

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.
— Reproduit dans S. Cowley & F. Vallee-Tourangeau (Eds.), *Cognition beyond the brain : Computation, interactivity and human artifice,* Berlin : Springer-Verlag, 2013 (2016 pour la 2ᵉ édition).

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. [Commentaire sur Oaksford & Chater, Precis of 'Bayesian Rationality'.]

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 reaso- ning. *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 interpre- tation 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. [Commentaire sur Byrne, Precis of 'The Rational Imagination'.]

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 condi-tional 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. *Cog-nitive 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.

Chapitres et actes de colloques

42. Purcell, Z., Rahwan, I., & Bonnefon, J. F. (2022). The quantified moral self. *Proceedings of the 44th Annual Confe- rence of the Cognitive Science Society.*

41. Bonnefon, J. F. (2021) Utility conditionals. In Knauff, M. & Spohn, W. (Eds.), *The Handbook of Rationality.* Cam- bridge, MA : MIT Press.

40. Bonnefon, J. F., Shariff, A., & Rahwan, I. (2020). The moral psychology of AI and the ethical opt-out problem. In S. M. Liao (Ed.) *The Ethics of Artificial Intelligence.* Oxford University Press.

39. Juanchich, M., Sirota, M., & Bonnefon, J. F. (2019). Verbal uncertainty. In C. Cummins & N. Katsos (Eds.) *Oxford Handbook of Experimental Semantics and Pragmatics.* Oxford University Press.

38. Trémolière, B., De Neys, W., & Bonnefon, J. F. (2017). Reasoning and moral judgment : A common experimental toolbox. In L. J. Ball & V. A. Thompson (Eds.), *International Handbook of Thinking and Reasoning.* Hove, UK : Psychology Press.

37. Awad, E., Dsouza, S., Shariff, A., Bonnefon, J. F., & Rahwan, I. (2017). Moral psychology at scale : Lessons from 35 million decisions about autonomous vehicles. *Proceedings of the 2017 Conference on Digital Experimentation.* Boston, MA.

36. Holtgraves, T. M., & Bonnefon, J. F. (2017). Experimental approaches to linguistic (im)politeness. In J. Culpeper, M. Haugh, & D. Kádár (Eds.), *Palgrave handbook of linguistic (im)politeness.* Palgrave McMillan.

35. Bonnefon, J. F. (2016). Predicting behavior on the basis of arguments from consequences. In F. Paglieri, L. Bonelli, & S. Felletti (Eds.), *The psychology of argument : Cognitive approaches to argumentation and persuasion.* Studies in Logic and Argumentation Series, London : College Publications, pp. 1–17.

34. Bonnefon, J. F., & Billaut, E. (2016). Individual differences in reasoning, beyond ability and disposition. In L. Macchi, M. Bagassi, & R. Viale (Eds.) *Cognitive unconscious and human rationality.* Cambridge, MA : MIT Press.

33. Awad, E., Dsouza, S., Chang, P. J., Bonnefon, J. F., Shariff, A., & Rahwan, I. (2016). Moral Machine : Global survey of autonomous vehicle ethics using the trolley problem paradigm. *Proceedings of the 2016 Conference on Digital Experimentation,* Boston, MA.

32. Bonnefon, J. F., & Politzer, G. (2014). Intelligence artificielle et psychologie du raisonnement et de la décision. In P. Marquis, O. Papini, & H. Prade (Eds.) *L'intelligence artificielle : Frontières et applications.* Toulouse, France : Cépaduès.

31. Bonnefon, J. F. (2014). Politeness and reasoning : Face, connectives, and quantifiers. In T. M. Holtgraves (Ed.), *Oxford handbook of language and social psychology.* Oxford : Oxford University Press.

30. Trémolière, B., & Bonnefon, J. F. (2014). The rationality of mortals : Thoughts of death impair analytic proces- sing. In W. De Neys, & M. Osman (Eds.), *New approaches in reasoning research.* Hove, UK : Psychology Press.

29. Ben-Naim, J., Bonnefon, J.F., Herzig, A., Leblois, S., & Lorini, E. (2013). Computer-mediated trust in self-interested expert recommendations. In S. Cowley and F. Vallee-Tourangeau (Eds.), *Cognition beyond the brain : Computation, interactivity and human artifice.* Berlin : Springer-Verlag, pp. 52–70.

28. Bonnefon, J. F. (2012). Le raisonnement. In M. Denis (Ed.), *La psychologie cognitive.* Paris : Éditions de la Maison des sciences de l'homme.

27. Heimann, M., Pouget, S., Mullet, E., & Bonnefon, J. F. (2011). The experimental approach to trust in socially responsible funds. In W. Sun, C. Louche, & R. Pérez (Eds.), *Finance and Sustainability : Towards a new paradigm? A post-crisis agenda.* Bingley, UK : Emerald, pp. 169–183.

26. Bonnefon, J. F., De Neys, W., & Feeney, A. (2011). Processing scalar inferences in face-threatening contexts. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society.* Austin, TX : Cognitive Science Society.

25. Bonnefon, J.F., & Politzer, G. (2010). Pragmatic conditionals, conditional pragmatics, and the pragmatic com- ponent of conditional reasoning. In M. Oaksford & N. Chater (Eds.), *Cognition and conditionals : Probability and logic in human thinking.* Oxford : Oxford University Press, pp. 235–254.

24. Politzer, G., & Bonnefon, J.F. (2010). Two aspects of reasoning competence : A challenge for current accounts and a call for new conceptual tools. In M. Oaksford & N. Chater (Eds.), *Cognition and conditionals : Probability and logic in human thinking.* Oxford : Oxford University Press, pp. 357–374.

23. Rahwan, I., Bonnefon, J. F., Madakkatel, M. I., Awan, R. N, & Abdallah, S. (2010). Experiments for assessing floating reinstatement in argument-based reasoning. In R. Catrambone & S. Ohlsson (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society.* Portland, OR : Cognitive Science Society.

22. Bonnefon, J. F., & Vautier, S. (2009). Nouveaux problèmes de modélisation en psychologie du raisonnement. *Actes du Colloque Modèles et Apprentissage en Sciences Humaines et Sociales*.

21. Leblois, S., & Bonnefon, J. F. (2009). La bienveillance et les préférences comme paramètres de l'interprétation des recommandations vagues. *Actes des 5es Journées Francophones Modèles Formels de l'Interaction* (MFI'09).

20. Madakkatel, M. I., Rahwan, I., Bonnefon, J. F., Awan, R. N., & Abdallah, S. (2009). Formal argumentation and human reasoning : The case of reinstatement. In T. Bench-Capon, S. Parsons, & H. Prakken (Eds), *The Uses of computational argumentation : Papers from the AAAI Fall Symposium.* AAAI Press.

19. Nguyen, M. H., Longin, D. & Bonnefon, J. F. (2009). Relation of trust and social emotions : A logical approach. *Proceedings of the 2009 International Conference on Intelligent Agent Technology*. IEEE Computer Society Press.

18. Benferhat, S., Bonnefon, J. F., Chassy, P., Da Silva Neves, R. M., Dubois, D., Dupin de Saint-Cyr, F., Kayser, D., Nouioua, F., Nouioua-Boutouhami, S., Prade, H., & Smaoui, S. (2008). A comparative study of six formal models of causal ascription. In S. Greco, & T. Lukasiewicz (Eds.), *Scalable Uncertainty Management* (Lecture Notes in Artificial Intelligence 5291). Berlin : Springer, pp. 47–62.

17. Bonnefon, J. F., Dubois, D., & Fargier, H. (2008). An overview of bipolar qualitative decision rules. In G. Della Rica, D. Dubois, R. Kruse, & H. S. Lenz (Eds.), *Preferences and Similarities.* New York : SpringerWienNewYork, pp. 47–75.

16. Bonnefon, J. F., Dubois, D., & Prade, H. (2008). Transitive observation-based causation, saliency, and the Mar- kov condition. In S. Greco, & T. Lukasiewicz (Eds.), *Scalable Uncertainty Management* (Lecture Notes in Artificial Intelligence 5291). Berlin : Springer, pp. 78–91.

15. Amgoud, L., Bonnefon, J. F., & Prade, H. (2007). The logical handling of threats, rewards, tips, and warnings. In J. G. Carbonell, & J. Siekmann (Eds.), *Symbolic and Quantitative Approaches to Reasoning with Uncertainty* (Lecture Notes in Artificial Intelligence 4724). Berlin : Springer, pp. 235–246.

14. Demeure, V., Bonnefon, J. F., & Raufaste, É. (2007a). Rôle de la face et de l'utilité dans l'interprétation d'énoncés ambigus question/requête et incompréhension/désaccord. *Actes des 4es Journées Francophones Modèles Formels de l'Interaction* (MFI'07) Paris, 30 mai – 1er juin.

13. Demeure, V., Bonnefon, J. F., & Raufaste, É. (2007b). 'Is there any coffee left, boss ?' Face and utility concerns in the interpretation of ambiguous questions/requests. In S. Vosniadou, D. Kayser, & A. Protopapas (Eds), *Procee- dings of the European Cognitive Science Conference 2007,* Delphes, Grèce, 23–27 mai. Hove : Lawrence Erlbaum, pp 431–436.

12. Bonnefon, J. F., Da Silva Neves, R. M., Dubois, D., & Prade, H. (2006a). Background default knowledge and causality ascriptions. In G. Brewka, S. Coradeschi, A. Perini, & P. Traverso (Eds.), *Proceedings of the 17th European Conference on Artificial Intelligence* (ECAI2006), Riva di Garda, Italie, 28 août – 1 septembre. Zurich : IOS Press, pp 11–15.

11. Bonnefon, J. F., Da Silva Neves, R. M., Dubois, D., & Prade, H. (2006b). Model and experimental studies of causality ascriptions. In J. Dix & A. Hunter (Eds), *Proceedings of the 11th Workshop on Nonmonotonic Reasoning* (NMR06), Lake District, Angleterre, 30 mai – 1er juin. Clausthal, Allemagne : TU Clausthal, pp 193–199.

10. Bonnefon, J. F., & Fargier, H. (2006). Comparing sets of positive and negative arguments : Empirical assessment of seven qualitative rules. In G. Brewka, S. Coradeschi, A. Perini, & P. Traverso (Eds.), *Proceedings of the 17th European Conference on Artificial Intelligence* (ECAI2006), Riva di Garda, Italie, 28 août – 1 septembre. Zurich : IOS Press, pp 16–20.

9. Amgoud, L., Bonnefon, J. F., & Prade, H. (2005). An argumentation-based approach to multiple criteria decision. In L. Godo (Ed), *Proceedings of the 8th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty* (ECSQARU 2005), Barcelone, Espagne, 6–8 juillet. Berlin : Springer Verlag, pp 269–280.

8. Bonnefon, J. F., & Hilton, D. J. (2005). Getting the point of conditionals : An argumentative approach to the psychological interpretation of conditional premises. In G. Kern-Isberner, W. Rödder, & F. Kulmann (Eds), *Conditionals, Information, and Inference : International Workshop* (WCII 2002). Hagen, Allemagne, 12–15 mai. Ber- lin : Springer Verlag, pp 59–65.

7. Bonnefon, J. F., Longin, D., & Raufaste, É. (2005). Utilité et convention : Les deux faces d'une même pièce pour l'attribution d'intention. In A. Herzig, Y. Lespérance, & A.-I. Mouaddib (Eds.), *Actes des 3es Journées Francophones Modèles Formels de l'Interaction* (MFI'05). Caen, 25–27 mai. Toulouse : Cépaduès, pp 45–54.

6. Bonnefon, J. F., & Villejoubert, G. (2005). Communicating likelihood and managing face : Can we say it is probable when we know it to be certain ? In B. G. Bara, L. Barsalou, and M. Buciarelli (Eds.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society*, Stresa, Italie, 21–23 juillet. Mahwah, NJ : Erlbaum, pp 316–321.

5. Benferhat, S., Bonnefon, J. F., & Da Silva Neves, R. (2004a) An experimental analysis of possibilistic default rea- soning. In D. Dubois, C.A. Welty & M.-A. Williams (Eds.), *Principles of Knowledge Representation and Reasoning : Proceedings of the 9th International Conference* (KR2004), Whistler, Canada, 2–5 juin. AAAI Press, pp 130–140.

4. Benferhat, S., Bonnefon, J. F., & Da Silva Neves, R. (2004b). Une investigation expérimentale de l'approche possibiliste du raisonnement non monotone. *Actes du 14e Congrès Francophone de Reconnaissance des Formes et Intelligence Artificielle* (RFIA2004). Toulouse, 28–30 janvier.

3. Bonnefon, J. F., & Hilton, D. J. (2001). Formalizing human uncertain reasoning with default rules : A psycholo- gical conundrum and a pragmatic suggestion. In S. Benferhat & P. Bernard (Eds), *Proceedings of the 6th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty* (ECSQARU 2001). Toulouse, France, 19–21 Septembre. Berlin : Springer Verlag, pp 628–634.

2. Hilton, D. J., Bonnefon, J. F., & Kemmelmeier, M. (2001). Pragmatics at work : Formulation and interpretation of conditional instructions. In J.D. Moore & K. Stenning (Eds), *Proceedings of the 23rd Annual Conference of the Cognitive Science Society,* Edinburgh, Écosse, 1–4 Août. Hillsdale : Lawrence Erlbaum, pp 417–422.

1. Da Silva Neves, R., Bonnefon, J. F., & Raufaste, É. (2000). Rationality in human non-monotonic inference. *8th Intern. Workshop on Non-Monotonic Reasoning* (NMR2000). Breckenridge, Colorado, 9–11 Avril. Linköping Elec- tronic Articles in Computer and Information Science. http://www.ida.liu.se/ext/epa/cis/ufn-00/08/tcover.html

Conférences

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

71. **[Invitation]** Vers quelle société nous mènent l'IA et ses filtres ? (2023, septembre) *Colloque du Collège de France sur l'IA et ses Défis.*

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

69. **[Invitation]** AI for good. (2023, juin) *3rd Common Good Summit.*

68. **[Invitation]** Patient screening : Ethics and economics. (2023, juin) *1st CEPR Health Economics Conference.*

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

66. **[Invitation]** Uptake and social consequences of lie-detection algorithms. (2023, avril) *Workshop Autorité de contrôle prudentiel et de résolution.*

65. **[Invitation]** Will Autonomous Vehicles need road user education ? (2023, janvier) *Conference on road user edu- cation for goal zero.*

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

63. **[Invitation]** The psychology of human-machine interaction. (2021, avril) *Microsoft Data Science and Law Forum.*

62. **[Keynote]** Behavioral data for machine ethics. (2019, novembre) *Workshop on Morality, Social Choice, and Arti- ficial Intelligence.*

61. **[Invitation]** Transports innovants—le futur est déjà là. (2019, octobre). *Forum Innovation Intelligence Artificielle.*

60. **[Invitation]** The moral machine experiment. (2019, octobre) *Minds & Tech Conference.*

59. **[Invitation]** AI & ethics. (2019, septembre) *Tech Talks : The Politics of Cybersecurity.*

58. **[Invitation]** The massive crowdsourcing of machine ethics. (2019, mai) *Global Business Ethics Symposium.*

57. **[Invitation]** Public acceptance and adoption. (2019, avril)*European Union Conference on Connected and Automated Driving.*

56. **[Invitation]** The moral machine experiment. (2019, janvier) *Ethics and Policy Workshop : Ethics in Algorithms.*

55. **[Invitation]** The moral machine experiment. (2018, septembre) *Microsoft Data Science and Law Forum.*

54. **[Keynote]** The moral machine experiment. (2018, juillet) *27th International Joint Conference on Artificial Intelli- gence.*

53. **[Invitation]** Intelligence artificielle et cognition. (2018, juillet). *Conférence Olivier Legrain Sciences et Société.*

52. The moral machine experiment. (2017, décembre) *Workshop Social Ties and Social Intelligence.*

51. **[Invitation]** The moral machine experiment. (2017, décembre) *Conférence Nouvelles Pratiques du Journalisme.*

50. Split-second detection of cooperativeness from faces in the trust game. (2017, juillet) *39th Annual Meeting of the Cognitive Science Society.*

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

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

47. The social dilemma of autonomous vehicles. (2016, août) *International Conference on Thinking.*

46. Experimental ethics for autonomous vehicles. (2016, février) *2nd IAST-IEC Joint Workshop.* Toulouse, France.

45. **[Invitation]** Eye movements track the early stages of explaining anomalous decisions. (2015, novembre) *Leading Edge Workshop : The Process of Explanation.*

44. Trustworthiness detection. (2015, mars) *International Convention of the Association for Psychological Science.*

43. People see each other as selfish hedonic maximizers. (2015, mars) *International Convention of the Association for Psychological Science.*

42. **[Keynote]** Cognitive and moral reflection across social networks. (2015, février) *Conference on Reasoning, Ar- gumentation, and Critical Thinking Instruction.*

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

40. The double effect of social networks on utilitarian dilemmas. (2014, juin) *Shifting Attitudes : Radical Upheaval and its Legacies.*

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

38. People are more likely to lie when they are more likely to tell the truth. (2013, July) *6th London Reasoning Work- shop.*

37. **[Invitation]** Psychologie expérimentale de la mort. (2013, mai) *Archéologie des peuplements et des populations pré- et protohistorique.*

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

35. **[Invitation]** Reasoning in an imperfect world. (2012, juillet) *International Congress of Psychology, ICP2012.*

34. Utility templates for consequential arguments. (2012, juillet) *International Conference on Thinking, ICT2012.*

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

32. **[Invitation]** The causal structure of utility conditionals. (2012, mai) *Workshop 'How universal is causal cognition ?'*

31. Utility templates for consequential argumentation. (2012, mai) *International Workshop on Formal, Experimental and Informal Approaches to Argumentation.* Toulouse, France.

30. **[Invitation]** Nouveaux territoires pour les neurosciences de la rationalité. (2011, juin) *Colloque Annuel de l'ITMO Neurosciences.*

29. Inferences from preferences : A random sampling experiment. (2011, mai) *European Conference on Cognitive Science.*

28. From contract proposals to utility conditionals. (2010, octobre) *The intersection of logic, probability and pragmatics.*

27. Deductive reasoning from if-then personality signatures. (2010, juillet) *5th London Reasoning Workshop.* Londres.

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

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

24. Acceptable and unacceptable instances of one paradox of implication. (2010, mars) *Toulouse Meeting on Reasoning.*

23. Facework and uncertain reasoning in health communication. (2010, mars) *Toulouse Meeting on Reasoning.*

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

21. **[Invitation]** A comprehensive theory of reasoning from utility conditionals. (2009, septembre) *Workshop on Conditionals, Conditional Probability, and Conditionalization.*

20. The rational toll and emotional benefit of reasoning to a just world. (2009, juillet) *4th London Reasoning Workshop.*

19. Conditions for the transitivity of causal chains. (2009, juin) *International Multidisciplinary Workshop on Causality.*

18. Contexte d'énonciation, contexte du raisonnement. (2009, avril) *Symposium 'Quel sens pour la linguistique ?'* Tou- louse.

17. Expériences comportementales sur l'agrégation des jugements en situation de paradoxe doctrinal. (2008, no- vembre) *Workshop Multidisciplinaire sur l'Agrégation des Préférences et des Jugements.*

16. Utility matrices and folk principles of choice. (2008, août) *3rd London Reasoning Workshop.*

15. A theory of utility conditionals, or how decision processes leak into deductive inferences. (2008, août) *Interna- tional Conference on Thinking.*

14. Politeness effects in the interpretation of conditionals, probabilities, and quantifiers. (2008, août) *International Conference on Thinking.*

13. **[Invitation]** The causal analysis of traffic accident reports. (2008, mars) *Conference of the Eastern Psychological Association.*

12. **[Invitation]** Deductive reasoning about the consequences of moral and immoral behaviour. (2008, novembre) *The Economy of the Soul : Rational Choice and Moral Decision-Making.*

11. Pragmatic conditionals, conditional pragmatics, and the pragmatics of conditional reasoning. (2007, août) *2nd London Reasoning Workshop.*

10. A mixed Rasch model of dual-process conditional reasoning. (2006, juillet) *In two minds : Dual-process theories of rationality and reasoning.*

9. Individual differences in pragmatic modulation : A mixed Rasch model of reasoning with conditional arguments. (2006, mars) *International Meeting on Mental Models and Reasoning.*

8. Modus Tollens, Modus Shmollens : Conversational effects on contrapositive reasoning. (2006, septembre) *The London Reasoning Workshop* .

7. Reinstatement, floating conclusions, and the credulity of mental model reasoning. (2004, juillet) *5th International Conference on Thinking*

6. L'agrégation d'arguments dans la décision : Test expérimental de cinq modèles. (2003, décembre) *2ᵉ Journée thé- matique sur l'argumentation du Programme de Recherche en Sciences Cognitives de Toulouse* .

5. What's in a successful entrepreneur's mind? (2003, avril) *Toulouse Workshop on Judgement and Decision Making.*

4. Reasoning on conditionals with valued outcomes. (2003, mars) *2nd International Conference on Reasoning and Decision Making* ,

3. Consequential conditionals. (2003, mai) *Workshop on Causal and Conditional Reasoning.*

2. Compréhension, confiance et harmonie : En quoi la psychologie du raisonnement peut-elle bénéficier d'une approche argumentative? (2001, décembre) *1ʳᵉ Journée thématique sur l'argumentation, Programme de Recherche en Sciences Cognitives de Toulouse.*

1. Étude empirique d'un modèle formel pour l'étude du raisonnement conditionnel non-monotone : Logique Pos- sibiliste, Système P et Objets Conditionnels. (1999, janvier) *ACCION Interdisciplinarité en science de la cognition.*