

T&F PROOFS NOT FOR DISTRIBUTION

THE ROUTLEDGE HANDBOOK OF TRUST AND PHILOSOPHY

Edited by Judith Simon

T&F PROOFS NOT FOR DISTRIBUTION

First published 2020
by Routledge
52 Vanderbilt Avenue, New York, NY 10017

and by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

Routledge is an imprint of the Taylor & Francis Group, an informa business

© 2020 Taylor & Francis

The right of Judith Simon to be identified as the author of the editorial material, and of the authors for their individual chapters, has been asserted in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this book may be reprinted or reproduced or utilised in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

Trademark notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Library of Congress Cataloging-in-Publication Data
A catalog record for this title has been requested

ISBN: 978-1-138-68746-2 (hbk)
ISBN: 978-1-315-54229-4 (ebk)

Typeset in Times New Roman
by Taylor & Francis Books

T&F PROOFS NOT FOR DISTRIBUTION

16

TRUST: PERSPECTIVES IN PSYCHOLOGY

Fabrice Clément

Psychologists and philosophers, even when studying the same object, do not see things through the same epistemic glasses. Philosophy is traditionally invested in normative issues and, in the case of trust, the goal is to find some formal criteria to determine when it is justified to trust someone. Psychologists do not focus on how things *should* be, but rather on how information *is* processed in everyday life; this division of cognitive labor is something worth keeping in mind when involved in interdisciplinary work. But philosophers have obviously played an important role in the history of science: thanks to their conceptual work, they cleared the ground for empirical research by specifying the frontiers of the phenomenon to study. This is notably the case with trust, a “phenomenon we are so familiar with that we scarcely notice its presence and its variety” (Baier 1986:233). In order to determine the content of a concept, philosophers essentially question the way it is used in everyday speeches. The problem with the notion of trust is that different uses of the word can orient scientific research to different directions. Some favored cases, like contracts, where people decide if the amount of risk involved in cooperating with someone (usually unfamiliar), who will then reciprocate, is tolerable (Hardin 2002). These models are strongly inspired by dilemma games, where rational individuals make their decisions by counterbalancing personal and collective interests. In such contexts, the psychological processes involved are essentially conscious and, supposedly, rational – even if many empirical studies have shown that individuals do not often behave as egoistically as the economic model would suggest (Fehr and Schmidt 1999). At the other end of the spectrum, philosophers consider cases where trust seems to be of a different nature, like the trust infants have toward their parents. In such common cases, it would be odd to speak of a contract-based kind of trust; what is at stake is clearly not the result of a rational calculation, but a relationship characterized by a peaceful state of mind. As Baier (1986:234) put it, trust in such contexts can be seen as analogous to air: we inhabit a climate of trust and “notice it as we notice air, only when it becomes scarce or polluted.” Seen from that perspective, trust is therefore understood as a basic attitude (Faulkner 2015), most likely of an affective nature (Jones 1996). In a nutshell, when it defining the nature of trust, philosophers seem to oscillate between characterizing trust as a state of mind that could be compared either to a belief (for instance Gambetta 1988; see also Keren, this volume). or that could be compared to an emotion (see Lahno, this volume).

T&F PROOFS NOT FOR DISTRIBUTION

Fabrice Clément

An evolutionary perspective is often useful to disentangle the nature of complex human phenomena; why would something like trust (and distrust) have evolved in our species? The answer has probably to be sought in the complex forms of cooperation that characterize human social organizations. Humans are said to be highly cooperative (Tomasello 2009), in the sense that they can constitute coalitions that are not related to kinship. Without the guarantee of being reciprocated in the future, an act of altruism such as helping others can be costly (Trivers 1971). As evolutionary simulations have demonstrated (Axelrod and Hamilton 1981), a strategy can only be adaptive in repeated social interactions (like a prisoner's dilemma) when participants can eliminate those who do not reciprocate. In other words, it is advantageous to cooperate, even with a certain time delay, but only when the partner will reciprocate in a not too distant future. However, it is generally held that the benefit of cooperation (and of being part of a group where individuals can help each other when needed) is so great, that it is in the interest of the trustee to satisfy the expectations of the trustor so that he can benefit in the future when more goods are exchanged. But, notably when the trustee is not engaged in a long-term relationship, the temptation to defect can be powerful, for example, to get a greater reward, or to avoid some extra cost to fulfill one's obligation. Any exchange therefore involves a certain risk for the trustor. For some, this risk is even embedded in the etymology of the word: according to Hardin (2002), the word "trust" in English comes from the way villagers organized hunting during the Middle Ages. Most of them played the role of beaters to frighten smaller game like rabbits or birds out of the bushes and trees. On the other side of the wood, the hunters waited for the fleeing animals to come out to shoot them: they stand *tryst*. Trust was therefore a role whose success depended on the actions of other members of the group.

The lesson of this evolutionary story is that trusting others by favoring cooperation had such an adaptive advantage, that trust is hardly avoidable. At the same time, given the clear risk of being let down or even betrayed by those we trust unquestioningly, it seems unlikely that something like blind trust could have evolved. Indeed, several studies seem to indicate that fast and frugal processes are in place as soon as a potential cooperator enters the perceptual field. In other words, it is probable that some psychological processes had evolved to filter our potential cooperators. Such mechanisms work implicitly, in the background of our conscious mind, and trigger some affective alarm when detecting a potential risk of manipulation. Such automaticity has been illustrated by Todorov and his colleagues, who have shown in different experiments that individuals screen others' faces extremely quickly and evaluate whether someone is trustworthy or not within a few milliseconds. For instance, Willis and Todorov (2006) asked participants to judge faces on different personality traits (trustworthiness, competence, likeability, aggressiveness and attractiveness) after 100ms, 500ms or 1,000ms. Interestingly, the scores obtained after 100ms or 1,000ms were highly correlated, especially in the case of trustworthiness. In other words, detecting the amount of trustworthiness you can attribute to another person seems to be processed extremely fast by our brains. In another experiment, the effects of face trustworthiness were even detectable when the stimuli were presented for 33ms, i.e. below the threshold of objective awareness (Todorov and Duchaine 2008). Moreover, a study conducted with participants suffering from prosopagnosia, i.e. being unable to memorize faces or to perceive facial identity, showed that, in spite of their condition, the participants could make normal and fast judgments of trustworthiness (Todorov and Duchaine 2008),

T&F PROOFS NOT FOR DISTRIBUTION

Trust: Perspectives in Psychology

suggesting the existence of a specific mechanism for forming impressions about people. From a neurological perspective, this “fast and frugal” heuristic seems to principally involve the amygdala (Engell, Haxby and Todorov 2007). Thanks to 2D computer simulations based on multiple judgments of emotionally neutral faces, Oosterhof and Todorov (2008) managed to show that the cues used for trustworthiness heuristics are related to positive or negative valence, indicating that valence may be a good proxy of trustworthiness. In a nutshell, faces of people that are evaluated as trustworthy tend to look happier, even when they display an emotionally neutral face; on the contrary, people judged as less trustworthy, look angrier (see also Clément, Bernard, Grandjean and Sander 2013). Such attention to expressed emotions for the assessment of trustworthiness may be linked to the fact that different valences elicit approach versus avoidance behavior.

The automaticity of trustworthiness assessment is further supported by the discovery that even young children are sensitive to the valence displayed by people around them who could potentially be trusted. Cogsdill and her colleagues showed that 3- and 4-year-old children generated basic “mean”/ “nice” judgments to facial characteristics in a similar way to adults. This precocity could indicate that the ability to infer traits from faces might not result from “slow social-learning mechanisms that develop through the gradual detection and internalization of environmental regularities” (Cogsdill, Todorov, Spelke and Banaji 2014:7), but instead through cognitive mechanisms whose nature is still under-determined.

Besides the propensity to detect trustworthiness in others’ faces – a heuristic that is, incidentally, not always reliable (Bond, Berry and Omar 1994) – another adaptive solution could be to provide a mechanism for selectively attending to those who do not follow the rules of the game. This is precisely the hypothesis defended by the evolutionary psychologist Leda Cosmides (1989). Her starting point was the classic Wason selection task designed to test deductive reasoning abilities: a set of four cards was placed on a table, each of which had a number on one side and a colored patch on the other side. The visible faces of the cards showed 3, 8, red and brown. Participants were asked about the cards they should turn over in order to test the truth of the following proposition: “if a card shows an even number, then its opposite face shows a primary color.” The right deductive rule is to turn the 8 card (if it is not red, it violates the rule), and the brown card (if it is even, it violates the rule). However, most participants chose the 8 card and the *red* card, trying to confirm the rule. Cosmides introduced a variation of the game by asking participants to verify the following rule: “if a person is drinking beer, then he must be over 20 years old.” The four cards had information about four people sat a table: “drinking beer,” “drinking coke,” “25 years old,” and “16 years old.” This time, most participants responded correctly by choosing “drinking beer” and “16 years old.” Cosmides explains such differences based on the *content* of the problem instead of their *intrinsic difficulty*, by arguing that the human mind includes specific inferential procedures to detect cheating on social contracts (see also Sugiyama, Tooby and Cosmides 2002; for a critique, see Sperber and Girotto, 2002). Similarly, to avoid the risk of being manipulated, one can expect that cheaters are not only rapidly detected but also punished by the other players. This idea has been developed by behavioral economists under the “strong reciprocity” hypothesis. Indeed, there are many experiments that show that participants are actually willing to punish those who fail to cooperate, even if this “altruistic” punishment is costly for them (e.g. Fehr and Gächter 2002). Moreover, it has been demonstrated that punishing cheaters

T&F PROOFS NOT FOR DISTRIBUTION

Fabrice Clément

activates the dorsal striatum, an area of the brain that has been implicated in the processing of reward (de Quervain et al. 2004). In other words, our brains seem motivated to detect cheaters and to punish them. In such a world, trusting is a slightly less risky endeavor. Moreover, given the fact that humans are endowed with language, cheaters are at risk not only of being discovered, but also of being publicly denounced. By having their reputation tarnished, they run the risk of being excluded from future cooperation (Fehr 2004, see also Origgi, this volume). It has, indeed, been shown that many of our conversations are about others' behaviors and that the role of gossip can be seen as detecting others' deviations from what is socially expected (Dunbar 2004).

Detecting the potential cooperators is often associated with a positive bias for people belonging to the same group (in-group) and a negative bias for those belonging to another group (out-group). While being fallible and thus potentially harmful to both trustors and trustees (see Scheman as well as Medina, this volume), the evolutionary explanation for such preferences would be that people belonging to the same group are subject to the same norms and values and tend therefore to be more cooperative, notably because they fear the social repercussions of an unfair behavior (Gil-White 2005). Social psychologists have shown how people are inclined to judge in-group members to be more helpful than out-group members and participants choose preferentially in-group members (for instance, students from the same university) as cooperators in economic games (Foddy, Platow and Yamagishi 2009). Interestingly, in-group categorization is itself a very basic mechanism that can be triggered by apparently uninformative details, like shared taste for a given artist used by Tajfel in his minimal group paradigm (Tajfel 1970). This automaticity is also demonstrated by its very early presence in infants' mind: 10-month-old infants, for instance, preferentially selected a toy offered by a person who previously spoke in their native language over a person who spoke in a different language (Kinzler, Dupoux and Spelke 2007). Similarly, the simple observation of characters showing either the same or a different taste than their own choice triggers specific preferences by 9- and 14-month-olds: they prefer individuals who treat similar others well and treat dissimilar others poorly (Hamlin, Mahajan, Liberman and Wynn 2013). In-group and out-group categorizations are therefore fallible, but very powerful shortcuts when it comes to deciphering whom to trust.

Thanks to language, humans can benefit from another kind of cooperation: the exchange of information facilitated by verbal statements, and trust plays a crucial role here. As Quine and Ullian (1978:30) stated, language affords us "more eyes to see with." Others' testimonies can be compared as vicarious observation and increasing knowledge about our environment confer major benefits, notably in terms of reducing uncertainty (see Faulkner, this volume). However, this kind of "epistemic trust" is subject to the same potential issues as other forms of social cooperation: it is possible for an informer to omit a piece of relevant information, or even to intentionally transmit a piece of information that he knows to be erroneous. Such Machiavellian manipulation can be beneficial for the informant, who can "transfer" wrong information into others' minds (Byrne and Whiten 1988). If the others were to believe such "testimony," they could adopt behaviors that are favorable for the informant, but not for themselves. With Sperber and colleagues, we proposed that the risk of manipulation is so high in human communication that something like an "epistemic vigilance" had to evolve in our species (Sperber et al. 2010). According to this hypothesis, selective pressure favored cost-effective procedures to evaluate each piece of information that was

T&F PROOFS NOT FOR DISTRIBUTION

Trust: Perspectives in Psychology

provided by someone else. Given that the relevance of the testimony depends on the source's competence (since they could be wrong by ignorance) and honesty (since they could be lying), one could expect that these two dimensions depend upon a relatively automatic evaluation. A good way to test these initial checks is to look at child development: if humans are endowed with epistemic vigilance, one would expect children to exhibit it quite early (Clément 2010). Of course, honesty is hard to perceive but a good proxy is *benevolence*: when wishing somebody well, a source is, in principle, not trying to manipulate her. Indeed, an early sensitivity to benevolence has been shown by young children: participants as young as 3 years old prefer the testimony of a benevolent informant rather than a malevolent informant – a character who behaved aggressively towards the experimenter (Mascaro and Sperber 2009). Similarly, several studies have shown that children do not blindly trust what is said to them. For instance, even a source that has been reliable in the past loses her credibility when she says something that contradicts the child's own perception (Clément, Koenig and Harris 2004). Moreover, children use past accounts of a source's reliability when deciding whom to trust (Koenig, Clément and Harris 2004; Harris et al. 2012). It has even been shown that preschoolers use different social cues to evaluate others' statements. For instance, they are more ready to trust familiar people than unfamiliar ones (Harris and Koenig 2009) and prefer informants who also have their native accent (Kinzler, Corriveau and Harris 2011). Another line of research shows that preschoolers take the number of sources into account when listening to contradictory statements; without other specifications, the testimony of a majority is selected over a minority (Haun, van Leeuwen and Edelson 2013) and this propensity can even overcome reliability until 5 or 6 years of age (Bernard, Proust and Clément 2015): children at this age are able to “resist” the majority point of view when this latter had been wrong in the past. An evaluation of the content of the information transmitted also seems to take place quite early. At 16 months, for instance, infants are surprised when someone uses a familiar name inappropriately (Koenig and Echols 2003). Three-year-olds and 4-year-olds can distinguish good from bad reasons when they are evaluating testimony (Koenig 2012). Recently, it has been shown that preschoolers evaluate the “logical quality” of explanations, preferring, for instance, non-circular over circular statements (Corriveau and Kurkul 2014; Mercier, Bernard and Clément 2014).

In summary, research in psychology leads us to believe that our ability to evaluate others' testimony before deciding whether to trust them is already present, even very early in our ontogenesis. While there is no definitive proof that something like some vigilance is “built” into our cognitive system, existing evidence seems to indicate that it is “natural” for our species to evaluate different aspects of potential partners before deciding to trust them and, for instance, to follow their advice. Moreover, the fact that very young children are able to perform at least some of these evaluative operations indicates that different cues of our social environment are automatically and implicitly processed in situations where trust is at play. Based on the psychology literature, it is therefore difficult to limit trust to its strategic dimension. This option, often taken in economics, implicitly leads to a concept of trust based solely on the premise that it is the product of explicit thinking, as a calculation of the probability that a potential cooperator is reliable. Although such conscious deliberation may sometimes happen (for instance when deciding whom to contact to decorate your house), trusting someone is most often the result of an unconscious process that will determine a certain attitude. To conclude: trust manifests itself within a large continuum starting from the infant's trust in her mother to the businessman's trust in a potential partner.

T&F PROOFS NOT FOR DISTRIBUTION

Fabrice Clément

Two important questions remain unanswered. The first question is of a “rousseauist” nature: does everyone start with unquestioned faith towards others’ good will and then learn to be more selective? Or has trust to be won, overcoming mistrust? A line of research has recently brought interesting insight into this issue by studying the influence of a hormone that acts as an important neurotransmitter: *oxytocin*. This hormone is involved in key social contexts when the binding between two individuals is vital, like during sex, birth and breastfeeding. Its release is notably associated with a warm and positive state of mind that favors the ongoing social relationship. The influence of this hormone most likely exists since the beginning of life. This has been shown for instance by newborn macaques which, after inhaling oxytocin, were more prompt to initiate affiliative pro-social behavior to a human caregiver (Simpson et al. 2014). With human adults, experimental economists have shown that intranasal administration of oxytocin during a trust game increase the readiness to bear risks (Kosfeld, Heinrichs, Zak, Fischbacher and Fehr 2005). Recently, we have shown that even the endogenous level of oxytocin, checked before an “egg-hunting” game, predicted the level of cooperation, but *only as long as participants belonged to the same group* (McClung et al. 2018). This last specification is important because it highlights the fact that oxytocin cannot simply be considered as a “love hormone.” Indeed, other researchers have shown that oxytocin not only enhances trust and cooperation between in-groups but also promotes aggression toward potential out-groups (De Dreu et al. 2010; De Dreu and Kret 2016). The secretion of oxytocin plays an important role therefore in establishing a psychological state where the absence of anxiety in the presence of another person will facilitate collaboration. But such warm peacefulness, that could remain unnoticed until broken, is preceded by an evaluation of the nature of the ongoing relationship (Kaufmann and Clément 2014) and, in the case of potential competition, the same hormone can have radically different behavioral consequences.

The second remaining question is about the dynamic of trust: how do we move on the “scale of trust,” from an unselfconscious sense of security to a more explicit awareness of the risk, so often inherent to cooperation (Baier 1986:236)? To answer this question, we can imagine a scenario starting with the initial and implicit trust that a baby experiences, at least when the relationship develops harmoniously with her caregivers. It would be possible to imagine a widening circle to illustrate levees of trust. At the center of the circle one can expect trust to be largely unnoticed, like the air we breathe. However, things get more complex with the widening of the circle. Notably, a phenomenon well known to parents appears between 6 and 8 months of age: *Stranger Fear* (Sroufe 1977; Brooker et al. 2013). It is probably only when this anxiety can be – partially – overcome that the different implicit evaluation processes we have described can kick in. But, it is still not necessary at that point to presuppose explicit processes, as alternatives have been shown in recent research on metacognition. For a long time indeed, metacognition was thought of as a reflective process, a “cognition that reflects on, monitors, or regulates first-order cognition” (Kuhn 2000); in other words, metacognition would essentially be *metarepresentational*, involving representations that are about other representations (Sperber 2000). The problem with this conception is that it is cognitively demanding and it does not fit with young children’s executive abilities (Clément 2016). However, we have seen that preschoolers are able to “decide” whom to trust based on cues about various phenomena (reliability, benevolence, consensus, familiarity, etc.). To explain these precocious monitoring abilities, the philosopher Joelle Proust proposes that metacognition can indeed be differentiated from metarepresentations (Proust 2007). To monitor one’s cognition, we rely on different kinds of

T&F PROOFS NOT FOR DISTRIBUTION

Trust: Perspectives in Psychology

cues that are correlated with *feelings* (Proust 2013). The typical example, proposed by Asher Koriat, is the case of ‘the feeling of knowing’ (tip-of-the-tongue): even if access to a specific content is momentary impossible, we can feel that it is within “cognitive reach” (Koriat 1993). Within this perspective, trust can be understood as the intimate resonance of an evaluation process that is most often left opaque to us. With time and the development of a theory of mind, i.e. the ability to reflexively represent one’s mental states and those of others, more reflexive evaluations can take place. The economists’ version of trust can take its place at the other end of the spectrum: after a strategic weighting of the situation, you can consciously decide to trust a potential cooperator, with a reasonable doubt based on a rough estimation of the probability that s/he will reciprocate.

The psychological epistemic lenses adopted here highlight the unconscious processes that underlie social exchanges when the participants trust each other, i.e. accept the risk inherent of the cooperative situation with a degree of confidence. Many of these evaluations are not accessible to consciousness but the feelings they trigger can be re-evaluated and, eventually, put into question. As natural selection aims at efficient mechanisms, in terms of time and energy, the evaluation processes leading to trust are approximations which have worked sufficiently well in the past to ensure the success of our species. Therefore, it is possible that they lead us to trust – or distrust – someone, based on some unreliable cues. In some cases, these mechanisms can even have deleterious consequences, notably when decisions are based solely on group membership. The normative dimension of philosophy then becomes essential, specifying when it is justified to trust or to be skeptical, enabling us to consciously revise impressions that are too readily available. Such critical abilities are probably more desirable than ever.

References

- Axelrod, R. and Hamilton, W. (1981) “The Evolution of Cooperation,” *Science* 211: 1390–1396.
- Baier, A. (1986) “Trust and Antitrust,” *Ethics*, 96(2): 231–260.
- Bernard, S., Proust, J. and Clément, F. (2015) “Four- to Six-Year-Old Children’s Sensitivity to Reliability Versus Consensus in the Endorsement of Object Labels,” *Child Development* 86(4): 1112–1124.
- Bond, C., Berry, D.S. and Omar, A. (1994) “The Kernel of Truth in Judgments of Deceptiveness,” *Basic and Applied Social Psychology* 15(4): 523–534.
- Brooker, R.J., Buss, K.A., Lemery-Chalfant, K., Aksan, N., Davidson, R.J. and Goldsmith, H. H. (2013) “The Development of Stranger Fear in Infancy and Toddlerhood: Normative Development, Individual Differences, Antecedents, and Outcomes,” *Developmental Science* 16 (6): 864–878.
- Byrne, R.W. and Whiten, A. (1988) *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect*, Oxford: Oxford University Press.
- Clément, F. (2010) “To Trust or not to Trust? Children’s Social Epistemology,” *Review of Philosophy and Psychology* 1:531–549.
- Clément, F. (2016) “The Multiple Meanings of Social Cognition,” in J.A. Green, W.A. Sandoval and I. Bråten (Eds.), *Handbook of Epistemic Cognition*, Oxford: Routledge.
- Clément, F., Bernard, S., Grandjean, D. and Sander, D. (2013) “Emotional Expression and Vocabulary Learning in Adults and Children,” *Cognition and Emotion* 27: 539–548.
- Clément, F., Koenig, M. and Harris, P. (2004) “The Ontogenesis of Trust,” *Mind & Language* 19: 360–379.
- Cogsdill, E.J., Todorov, A.T., Spelke, E.S. and Banaji, M.R. (2014) “Inferring Character from Faces: A Developmental Study,” *Psychological Science* 25(5): 1132–1139.
- Corriveau, K.H. and Kurkul, K.E. (2014) “‘Why Does Rain Fall?’: Children Prefer to Learn from an Informant Who Uses Noncircular Explanations,” *Child Development* 85: 1827–1835.

T&F PROOFS NOT FOR DISTRIBUTION

Fabrice Clément

- Cosmides, L. (1989) "The Logic of Social Exchange: Has Natural Selection Shaped How Humans Reason? Studies with the Wason Selection Task," *Cognition* 31(3): 187–276.
- De Dreu, C.K.W. and Kret, M.E. (2016) "Oxytocin Conditions Intergroup Relations through Upregulated In-Group Empathy, Cooperation, Conformity, and Defense," *Biological Psychiatry* 79(3): 165–173.
- De Dreu, C.K.W., Greer, L.L., Handgraaf, M.J.J., Shalvi, S., Kleef, G.A.V., Baas, M. ... Feith, S.W. W. (2010) "The Neuropeptide Oxytocin Regulates Parochial Altruism in Intergroup Conflict among Humans," *Science* 328(5984): 1408–1411.
- de Quervain, D.J.-F., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A. and Fehr, E. (2004) "The Neural Basis of Altruistic Punishment," *Science* 305 (5688): 1254–1258.
- Dunbar, R.I.M. (2004) "Gossip in Evolutionary Perspective," *Review of General Psychology* 8: 100–110.
- Engell, A.D., Haxby, J.V. and Todorov, A. (2007) "Implicit Trustworthiness Decisions: Automatic Coding of Face Properties in the Human Amygdala," *Journal of Cognitive Neuroscience* 19: 1508–1519.
- Faulkner, P. (2015) "The Attitude of Trust is Basic," *Analysis* 75(3): 424–429.
- Fehr, E. (2004) "Don't Lose Your Reputation," *Nature* 432: 2–3.
- Fehr, E. and Gächter, S. (2002) "Altruistic Punishment in Humans," *Nature* 415: 137–140.
- Fehr, E. and Schmidt, K.M. (1999) "A Theory of Fairness, Competition, and Cooperation," *The Quarterly Journal of Economics* 114(3): 817–868.
- Foddy, M., Platow, M.J. and Yamagishi, T. (2009) "Group-Based Trust in Strangers: The Role of Stereotypes and Expectations," *Psychological Science* 20(4): 419–422.
- Gambetta, D. (ed.) (1988) *Trust: Making and Breaking Cooperative Relations*, Oxford: Blackwell.
- Gil-White, F. (2005) "How Conformism Creates Ethnicity Creates Conformism (and Why this Matters to Lots of Things)," *The Monist* 88: 189–237.
- Hamlin, J.K., Mahajan, N., Liberman, Z. and Wynn, K. (2013) "Not Like Me = Bad: Infants Prefer Those Who Harm Dissimilar Others," *Psychological Science* 24: 589–594.
- Hardin, R. (2002) *Trust and Trustworthiness*, New York: Russell Sage Foundation. www.russellsage.org/publications/trust-and-trustworthiness-1
- Harris, P.L. and Koenig, M. (2009) "Choosing Your Informants: Weighing Familiarity and Recent Accuracy," *Developmental Science* 12(3): 426–437
- Harris, P.L., Corriveau, K.H., Pasquini, E.S., Koenig, M., Fusaro, M. and Clément, F. (2012) "Credulity and the Development of Selective Trust in Early Childhood," in M.J. Beran, J. Brandl, J. Perner and J. Proust (Eds.), *Foundations of Metacognition*, Oxford: Oxford University Press.
- Haun, D.B., van Leeuwen, E.J. and Edelson, M.G. (2013) "Majority Influence in Children and Other Animals," *Developmental Cognitive Neuroscience* 3: 61–71.
- Jones, K. (1996) "Trust as an Affective Attitude," *Ethics* 107(1): 4–25.
- Kaufmann, L. and Clément, F. (2014) "Wired for Society: Cognizing Pathways to Society and Culture," *Topoi* 33(2): 20–45.
- Kinzler, K.D., Corriveau, K.H. and Harris, P.L. (2011) "Children's Selective Trust in Native-Accented Speakers," *Developmental Science* 14: 106–111.
- Kinzler, K.D., Dupoux, E. and Spelke, E.S. (2007) "The Native Language of Social Cognition," *Proc Natl Acad Sci USA* 104: 12577–12580.
- Koenig, M.A. (2012) "Beyond Semantic Accuracy: Preschoolers Evaluate a Speaker's Reasons," *Child Development* 83: 1051–1063.
- Koenig, M.A. and Echols, C.H. (2003) "Infants' Understanding of False Labeling Events: The Referential Roles of Words and the Speakers Who Use Them," *Cognition* 87: 179–208.
- Koenig, M.A., Clément, F. and Harris, P.L. (2004) "Trust in Testimony. Children's Use of True and False Statements," *Psychological Science* 15: 694–698.
- Koriat, A. (1993) "How Do We Know That We Know? The Accessibility Model of the Feeling of Knowing," *Psychological Review* 100(4): 609–637.
- Kosfeld, M., Heinrichs, M., Zak, P.J., Fischbacher, U. and Fehr, E. (2005) "Oxytocin Increases Trust in Humans," *Nature* 435: 673–676.
- Kuhn, D. (2000) "Metacognitive Development," *Current Directions in Psychological Science* 9: 178–181.
- McClung, J.S., Triki, Z., Clément, F., Bangerter, A. and Bshary, R. (2018) "Endogenous Oxytocin Predicts Helping and Conversation as a Function of Group Membership," *Proc. R. Soc. B* 285 (1882): 20180939.

T&F PROOFS NOT FOR DISTRIBUTION

Trust: Perspectives in Psychology

- Mascaro, O. and Sperber, D. (2009) "The Moral, Epistemic, and Mindreading Components of Children's Vigilance towards Deception," *Cognition* 112: 367–380.
- Mercier, H., Bernard, S. and Clément, F. (2014) "Early Sensitivity to Arguments: How Preschoolers Weight Circular Arguments," *Journal of Experimental Child Psychology* 125: 102–109.
- Oosterhof, N.N. and Todorov, A. (2008) "The Functional Basis of Face Evaluation," *Proceedings of the National Academy of Sciences* 105: 11087.
- Proust, J. (2007) "Metacognition and Metarepresentation: Is a Self-Directed Theory of Mind a Precondition for Metacognition?" *Synthese* 159: 271–295.
- Proust, J. (2013) *The Philosophy of Metacognition: Mental Agency and Self-Awareness*, Oxford: Oxford University Press.
- Quine, W.V.O. and Ullian, J.S. (1978) *The Web of Belief*, Vol. 2, New York: Random House.
- Simpson, E.A., Sclafani, V., Paukner, A., Hamel, A.F., Novak, M.A., Meyer, J.S. ... Ferrari, P.F. (2014) "Inhaled Oxytocin Increases Positive Social Behaviors in Newborn Macaques," *Proceedings of the National Academy of Sciences* 111(19): 6922–6927.
- Sperber, D. (2000). "Metarepresentations in an Evolutionary Perspective," in D. Sperber (ed.), *Metarepresentations: A Multidisciplinary Perspective*, Oxford: Oxford University Press.
- Sperber, D. and Girotto, V. (2002) "Use or Misuse of the Selection Task?" *Cognition* 85(3): 277–290.
- Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origg, G. and Wilson, D. (2010) "Epistemic Vigilance," *Mind and Language* 24: 359–393.
- Sroufe, L.A. (1977) "Wariness of Strangers and the Study of Infant Development," *Child Development* 48(3): 731–746.
- Sugiyama, L.S., Tooby, J. and Cosmides, L. (2002) "Cross-Cultural Evidence of Cognitive Adaptations for Social Exchange among the Shiwiar of Ecuadorian Amazonia," *Proc Natl Acad Sci USA* 99: 11537–11542.
- Tajfel, H. (1970) "Experiments in Intergroup Discrimination," *Scientific American* 223(5), 96–103.
- Todorov, A. and Duchaine, B. (2008) "Reading Trustworthiness in Faces without Recognizing Faces," *Cognitive Neuropsychology* 25: 395–410.
- Tomasello, M. (2009) *Why We Cooperate*, Cambridge, MA: MIT Press.
- Trivers, R. (1971) "The Evolution of Reciprocal Altruism," *The Quarterly Review of Biology* 46: 35–57.
- Willis, J. and Todorov, A. (2006) "First Impressions: Making up Your Mind after a 100-Ms Exposure to a Face," *Psychological Science* 17: 592–598.