

Moral Intuition

From the Human Mind to Artificial Agents

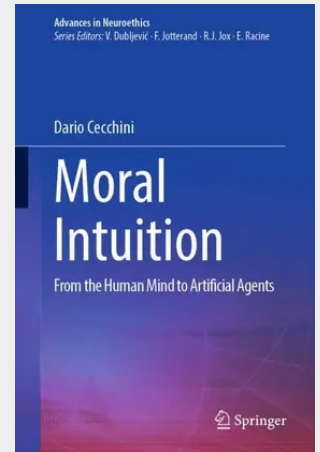
In the tradition of moral philosophy, long dominated by a rationalist paradigm, the idea of moral intuition has often been a source of embarrassment. How can the mind form a moral judgment within seconds, without any apparent reasoning?

In the spirit of neuroethics, this book demystifies moral intuition by examining the mental and neural processes that generate such automatic evaluations. Addressed to specialists in philosophy, psychology, and AI ethics, the book systematically investigates three questions: how moral intuitions work, how they can improve, and how they can be implemented in artificial agents.

Challenging the dominant default-interventionist view of moral reasoning, the first part argues that moral intuitions play a dual role: they detect harm and help in the environment, and they metacognitively regulate the deployment of cognitive resources, triggering reflection when intuitive outputs are uncertain or conflicting. Building on this foundation, the book offers a dyadic classification of the cognitive biases that shape moral intuitions and critically assesses strategies for mitigating them, including reasoning, expertise, and nudging. The final part extends this moral-psychological framework to artificial intelligence, arguing that the implementation of moral intuitions in artificial agents is both a feasible and a philosophically defensible goal, compatible with the functional capacities of contemporary AI systems.

In doing so, the book sets a new research agenda for understanding, improving, and implementing moral intuitions in both human and artificial agents.

In the tradition of moral philosophy, long dominated by a rationalist paradigm, the idea of moral intuition has often been a source of embarrassment. How can the mind form a moral judgment within seconds, without any apparent reasoning? In the spirit of neuroethics, this book demystifies moral intuition by examining the mental and neural processes that generate such automatic evaluations. Addressed to specialists in philosophy, psychology, and AI ethics, the book systematically investigates three questions: how moral intuitions work, how they can improve, and how they can be implemented in artificial agents. Challenging the dominant default-interventionist view of moral reasoning, the first part argues that moral intuitions play a dual role: they detect harm and help in the environment, and they metacognitively regulate the deployment of cognitive resources, triggering reflection when intuitive outputs are uncertain or conflicting. Building on this foundation, the book offers a dyadic classification of the cognitive biases that shape moral intuitions and critically assesses strategies for mitigating them, including reasoning, expertise, and nudging. The final part extends this moral-psychological framework to artificial intelligence, arguing that the implementation of moral intuitions in artificial agents is both a feasible and a philosophically defensible goal, compatible with the functional capacities of contemporary AI systems. In doing so, the book sets a new research agenda for understanding, improving, and implementing moral intuitions in both human and artificial agents.



149,79 €

139,99 € (zzgl. MwSt.)

Lieferfrist: bis zu 10 Tage

Artikelnummer: 9783032201164

Medium: Buch

ISBN: 978-3-032-20116-4

Verlag: Springer

Erscheinungstermin: 02.05.2026

Sprache(n): Englisch

Auflage: Erscheinungsjahr 2026

Serie: Advances in Neuroethics

Produktform: Gebunden

Gewicht: 578 g

Seiten: 226

Format (B x H): 160 x 241 mm

