The Measure of All Minds

Are psychometric tests valid for a new reality of artificial intelligence systems, technology-enhanced humans, and hybrids yet to come? Are the Turing Test, the ubiquitous CAPTCHAs, and the various animal cognition tests the best alternatives? In this fascinating and provocative book, José Hernández-Orallo formulates major scientific questions, integrates the most significant research developments, and offers a vision of the universal evaluation of cognition.

By replacing the dominant anthropocentric stance with a universal perspective where living organisms are considered as a special case, long-standing questions in the evaluation of behavior can be addressed in a wider landscape. Can we derive task difficulty intrinsically? Is a universal *g* factor – a common general component for all abilities – theoretically possible? Using algorithmic information theory as a foundation, the book elaborates on the evaluation of perceptual, developmental, social, verbal and collective features and critically analyzes what the future of intelligence might look like.

JOSÉ HERNÁNDEZ-ORALLO is Professor of Information Systems and Computation at the Technical University of Valencia. He has published four books and more than a hundred articles and papers in artificial intelligence, machine learning, data mining, cognitive science, and information systems. His work in the area of machine intelligence evaluation has been covered by both scientific and popular outlets including The Economist and New Scientist. He pioneered the application of algorithmic information theory to the development of artificial intelligence tests.

The Measure of All Minds

Evaluating Natural and Artificial Intelligence

José Hernández-Orallo

CAMBRIDGE
UNIVERSITY PRESS

Shaftesbury Road, Cambridge CB2 8EA, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India

103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of Cambridge University Press & Assessment, a department of the University of Cambridge.

We share the University's mission to contribute to society through the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781107153011

First published 2017
First paperback edition 2024

A catalogue record for this publication is available from the British Library

Library of Congress Cataloging-in-Publication data
Names: Hernández-Orallo, José, author.
Title: The measure of all minds : evaluating natural and artificial intelligence / José Hernández-Orallo.
Description: Cambridge, United Kingdom ; New York, NY : Cambridge University Press, 2017. | Includes bibliographical references and index.
Identifiers: LCCN 2016028921 | ISBN 9781107153011 (hardback : alk. paper)
Subjects: | MESH: Mental Processes – physiology | Artificial Intelligence – trends | Man-Machine Systems | Social Skills | Evaluation Studies as Topic
Classification: LCC R855.3 | NLM WL 26.5 | DDC 616.8900285–dc23
LC record available at https://lccn.loc.gov/2016028921

ISBN 978-1-107-15301-1 Hardback

Contents

Colour plates are to be found between pages 176 and 177

Panels

Preface

The quintessence of intelligence is one of the big questions still beyond our understanding. In the past, science has unravelled many other previously puzzling questions through measurement, a fundamental tool for the identification, comparison and classification of natural phenomena. Not surprisingly, a very significant portion of our still scant knowledge about what intelligence is – and what it is not – comes from this measurement effort. For more than a century, psychometrics, comparative psychology and other disciplines have developed a rich collection of measurement instruments for quantifying various behavioural properties in the animal kingdom, prominently placing humans as a yardstick.

Beyond the enormous landscape of behaviours in the animal kingdom, there is yet another gigantic space to be explored: the machine kingdom. A plethora of new types of 'creatures' is emerging: robots, animats, chatbots, digital assistants, social bots, automated avatars and artificial life forms, to name a few, including hybrids and collectives, such as machine-enhanced humans, cyborgs, artificial swarms, human computation systems and crowd computing platforms. These systems display behaviours and capabilities as peculiar as their developers and constituents can contrive. Universal psychometrics presents itself as a new area dealing with the measurement of behavioural features in the machine kingdom, which comprises any interactive system, biological, artificial or hybrid, individual or collective.

The focus on an enlarged set of subjects generates plenty of new questions and opportunities. Are IQ tests valid for arbitrary machines? Can we devise universal cognitive tests? Can we have a formal definition of intelligence solely based on computational principles? Can the structure of cognitive abilities and empirical latent factors, including the dominant *g* factor, be extrapolated beyond biological creatures? Can this be studied theoretically? How should artificial personalities be measured? Do we need intelligence to evaluate intelligence universally? The classical paradigms used to evaluate natural and

artificial systems have not been able to answer (or even formulate) these questions precisely. Also, customary evaluation tools are gamed by these new kinds of systems.

Recently, however, there has been a significant progress in a principled approach to the evaluation of behaviour based on information theory and computation. The anthropocentric stance is replaced by a universal perspective where life forms are considered as particular cases. Classical tools in human psychometrics, comparative psychology and animal cognition are not jettisoned but rethought for a wider landscape and substantiated on algorithmic grounds.

This book provides a comprehensive account of the concepts, terminology, theory and tools that should compose a unified framework for the universal evaluation of behavioural features. The exposition does not avoid some notions that are less consolidated, such as the arrangement of the space of abilities, the evaluation of personality or the process of ability development. The ideas that do not work are openly criticised, to aid the understanding of the many scattered scientific contributions that have recently appeared in different areas. In fact, some of these theories only make real sense – or no sense at all – when they are put together.

Many of the current conundrums in the evaluation of natural intelligence derive from the empirical evaluation of 'populations' (human groups, age ranges, species, etc.). The consideration of any conceivable behaviour (natural or artificial) and any imaginable 'machine population' provides a falsifiability criterion for any general claim, theory or test about behavioural features. The machine kingdom also brings a myriad of subjects to evaluate, with fewer experimentation constraints than those posed by humans and other animals. The theoretical underpinning on computation and information theory leads to several key formalisations, such as the concepts of task difficulty and policy-general intelligence. These new grounds illuminate blatant questions such as what human intelligence tests really measure.

Artificial intelligence can also benefit from the distinction between task-oriented evaluation and feature-oriented evaluation, jointly with a less anthropocentric methodology for the development and assessment of general-purpose agents. If properly overhauled, many tools from psychometrics can enter the scene of artificial intelligence evaluation, such as item response theory and adaptive testing. Similarly, the experience in the design of interfaces from animal evaluation can be crucial beyond natural intelligence.

Psychometrics, comparative psychology and artificial intelligence evaluation usually speak different languages. A great effort has been made to render

this book accessible and valuable for researchers and students in all these areas and, extensively, to any interested reader outside these disciplines. As a result of the integration of different areas, some paradigms will be challenged and some hypotheses will be refuted. The outcome for the future is an integration of well-founded principles for the evaluation of behaviour in humans, non-human animals and all other machines.

BOOK STRUCTURE

The book is organised in five parts.

Part I presents and frames the goals. Chapter 1 describes the diversity of behaviours resulting from a surge in the types of computers, robots, enhanced humans, hybrid systems and collectives thereof, with various types of communication. How can these systems be analysed and, ultimately, measured? This chapter specifies the conceptual characterisation of the so-called machine kingdom and the space of behavioural features, defines the scientific inquiry as a universal generalisation of psychometrics and enumerates the questions that are addressed during the rest of the book. Chapter 2 delineates the methodological principles to answer these questions, some fundamental concepts of measurement theory, the motivation for using the theoretical tools from computation and algorithmic information theory and the strengthened refutation power of those theoretical and empirical results over an enlarged set of subjects.

Part II provides the necessary background from the three areas universal psychometrics is built upon: human psychometrics, comparative (animal) cognition and artificial intelligence (AI); their existing links; and the barriers against a unified approach. The purpose of these chapters is not to give a comprehensive review (for which many specialised textbooks are available) but to focus on the concepts and tools that may be required or questioned during the book. Chapter 3 gives an account of psychometrics, IQ tests, the *g* factor, item response theory and adaptive testing in general and out-of-the-norm populations. Chapter 4 portrays a very particular view of the evaluation of non-human biological behaviour, ranging from apes, in many ways comparable to humans, to the detection of the so-called minimal cognition in bacteria, plants and extraterrestrial life. Chapter 5 analyses the chaotic state of AI evaluation, with disparate approaches ranging from Turing's imitation game to robotic competitions and the unsuccessful attempts so far towards a feature-oriented evaluation. Chapter 6 confronts the three previous chapters. What is common and distinctive in

the different approaches to the evaluation of intelligence and other behavioural features?

Part III presents the foundations for universal psychometrics based on computation and algorithmic information theory (AIT). Chapter 7 introduces AIT and shows how it pervades cognition. AIT can be used to generate test items that look very much the same as those that appear in some IQ tests, unveiling what these tests really are. Chapter 8 defines cognitive tasks in a universal way. Many different notions of difficulty are described, and a general difficulty function is formalised and derived from Levin's universal search. Chapter 9 elaborates agent characteristic curves from this new concept of difficulty. Tests must be constructed through an effective sampling over a range of difficulties, analysing the role of discriminating power in non-adaptive and adaptive tests. Chapter 10 tackles a controversial issue: how analytical notions of task similarity can be used to define what abilities are and to arrange the space of abilities from specific to general. Chapter 11 interprets general intelligence as the result of considering all tasks and, alternatively, in terms of whether a universal *g* factor exists.

Part IV delves into the significance of intelligence and other behavioural features in environments that harbour other systems, competing, co-operating or enhancing the subject's abilities. Chapter 12 investigates how cognitive development can be evaluated, from early perception to more conceptual abstraction. In the context of universal machines, such as humans and computers, potential features must be carefully understood in a probabilistic way. Chapter 13 deals with social skills, covering both competition and co-operation of humans, non-human animals and multi-agent systems in artificial intelligence. The Darwin-Wallace distribution is introduced as a way of characterising agent-populated environments. Chapter 14 is devoted to communication, which is key in knowledge exchange and development and in co-ordinating social organisations and collectives. Chapter 15 analyses the evaluation of groups and hybrids. How do the abilities of collective or symbiotic systems depend on the abilities of their members and their organisation? A recurrent question emerges all throughout this part: how crucial and distinctive are developmental, social, verbal and collective skills and drives?

Finally, Part V discusses what lies ahead. Chapter 16 considers what a universal cognitive test might look like. Test adaptation and interface customisation are key to evaluating a subject for which we lack any previous information. Chapter 17 has a more speculative character, arguing that measurement must play a crucial role in appraising the cognitive systems that the future may bring. Chapter 18 closes the book with the implications and the lessons learnt from universal psychometrics, and the way in which it can have a significant impact.

Shaded panels are spread throughout the book to introduce stand-alone concepts and questions, and keynote boxes spotlight the most important ideas. The highlights at the end of each chapter capture its take-away messages and the essential bits for subsequent chapters. This is meant as a checklist for those readers from diverse backgrounds who defer or skim through a chapter and wonder whether they are nevertheless ready to undertake the next one, especially in more technical parts of the book.

ACKNOWLEDGEMENTS

From the range of behavioural features that characterise us, we all experience a decisive struggle between our shortcomings and the way to overcome them. The people I relied on – my extended mind – tipped the scales for this book.

First and foremost, paraphrasing Stanisław Lem's *The Futurological Congress*, *I never would have* written this book *if it had not been for Professor* David L. Dowe, *who gave me clearly to understand that this was expected of me*. Most of the motivation, essence and even style of this book are the consequence of more than a decade of joint intellectual ventures. He always asked the ingenious questions. I just played Sancho.

The anYnt project turned from fancy into milestone for many of the key concepts and test models that pervade the book. I am most grateful to all the members of the project: David L. Dowe, again, Sergio España, Javier Insa and, most especially, Mariví Hernández-Lloreda.

Many people have contributed with valuable comments and corrections to the book drafts, insightful conversations, providing help with graphical material or answering some technical questions. Thanks to Sam S. Adams, Stuart Armstrong, Frank Bergmann, Harold Boley, Miles Brundage, Angelo Cangelosi, Nader Chmait, Douglas K. Detterman, David L. Dowe, Cèsar Ferri, Peter Flach, Arthur Franz, David J. Gunkel, Thomas Hendrey, M. Victoria Hernández-Lloreda, Enrique Hernández-Orallo, Bill Hibbard, Katja Hofmann, Frank Jäkel, Wendy Johnson, Michal Kosinski, Meelis Kull, Jan Leike, Leonid A. Levin, Miquel Llorente, Fernando Martínez-Plumed, Alexey Melkikh, Elena Messina, Carlos Monserrat, Shane Mueller, Stephen Muggleton, Adolfo Plasencia, Huw Price, Ricardo B. C. Prudêncio, M. José Ramírez-Quintana, John Rust, Ute Schmid, Aaron Sloman, Albert Soler, Robert Sparrow, David Stillwell, Claes Strannegård, Jared P. Taglialatela, Jan Arne Telle, Andrés Terrasa, Kristinn R. Thórisson and Susana Urbina.

I owe special gratitude to Lauren Cowles, Adam Kratoska, the anonymous reviewers and the rest of the team at Cambridge University Press for their

guidance, invaluable feedback and professionalism throughout the entire book process.

Above all, my warmest thanks go to Neus, Enric and Jaume, who responded so generously to the new glutton in the family. I hope they – and everyone else – appreciate that, regardless of the wrongs of this book, writing it was the right thing to do.