

## Computational Modeling of Cognition and Behavior

Computational modelling is now ubiquitous in psychology, and researchers who are not modellers may find it increasingly difficult to follow the theoretical developments in their field. This book presents an integrated framework for the development and application of models in psychology and related disciplines. Researchers and students are given the knowledge and tools to interpret models published in their area, as well as to develop, fit, and test their own models.

Both the development of models and key features of any model are covered, as are the applications of models in a variety of domains across the behavioural sciences. A number of chapters are devoted to fitting models using maximum likelihood and Bayesian estimation, including fitting hierarchical and mixture models. Model comparison is described as a core philosophy of scientific inference, and the use of models to understand theories and advance scientific discourse is explained.

**Simon Farrell** is a professor in the School of Psychological Science at the University of Western Australia. He uses computational modelling and experiments to understand memory, judgement, choice, and the role of memory in decision-making. He is the co-author of *Computational Modeling in Cognition: Principles and Practice* (2011) and has published numerous papers on the application of models to psychological data. Simon was Associate Editor of the *Journal of Memory and Language* (2009–11) and the *Quarterly Journal of Experimental Psychology* (2011–16). In 2009 Farrell was awarded the Bertelson Award by the European Society for Cognitive Psychology for his outstanding early career contribution to European Cognitive Psychology.

**Stephan Lewandowsky** is a professor of cognitive science in the School of Experimental Psychology at the University of Bristol. He was awarded a Discovery Outstanding Researcher Award from the Australian Research Council in 2011 and received a Wolfson Research Fellowship from the Royal Society upon moving to Bristol in 2013. He was appointed a Fellow of the Academy of Social Sciences in 2017. His research examines people's memory and decision-making, with an emphasis on how people update information in memory. He has published over 200 scholarly articles and books, including numerous papers on how people respond to corrections of misinformation and what determines people's acceptance of scientific findings.

# Computational Modeling of Cognition and Behavior

SIMON FARRELL

University of Western Australia, Perth

STEPHAN LEWANDOWSKY

University of Bristol

Cambridge University Press  
978-1-107-10999-5 — Computational Modeling of Cognition and Behavior  
Simon Farrell, Stephan Lewandowsky  
Frontmatter  
[More Information](#)

## CAMBRIDGE UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, 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  
79 Anson Road, #06–04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

[www.cambridge.org](http://www.cambridge.org)  
Information on this title: [www.cambridge.org/9781107109995](http://www.cambridge.org/9781107109995)  
DOI: 10.1017/9781316272503

© Simon Farrell and Stephan Lewandowsky 2018

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2018

Printed in the United Kingdom by Clays, St Ives plc

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

*Library of Congress Cataloging-in-Publication Data*

Names: Farrell, Simon, 1976– author. | Lewandowsky, Stephan, 1958– author.

Title: Computational modeling of cognition and behavior / Simon Farrell,

University of Western Australia, Perth, Stephan Lewandowsky, University of Bristol.

Description: New York, NY : Cambridge University Press, 2018.

Identifiers: LCCN 2017025806 | ISBN 9781107109995 (Hardback) | ISBN 9781107525610 (paperback)

Subjects: LCSH: Cognition—Mathematical models. | Psychology—Mathematical models.

Classification: LCC BF311 .F358 2018 | DDC 153.01/5118—dc23

LC record available at <https://lcn.loc.gov/2017025806>

ISBN 978-1-107-10999-5 Hardback

ISBN 978-1-107-52561-0 Paperback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party Internet websites referred to in this publication and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

Cambridge University Press  
978-1-107-10999-5 — Computational Modeling of Cognition and Behavior  
Simon Farrell , Stephan Lewandowsky  
Frontmatter  
[More Information](#)

---

**To Jodi, Alec, and Sylvie, with love (S.F.)**

**To Annie and the tribe (Ben, Rachel, Thomas, Jess, and Zachary) with love (S.L.)**

## Contents

	<i>List of Illustrations</i>	page xiii
	<i>List of Tables</i>	xviii
	<i>List of Contributors</i>	xix
	<i>Preface</i>	xxi
<b>Part I</b>	<b>Introduction to Modeling</b>	<b>1</b>
<b>1</b>	<b>Introduction</b>	<b>3</b>
	1.1 Models and Theories in Science	3
	1.2 Quantitative Modeling in Cognition	6
	1.2.1 Models and Data	6
	1.2.2 Data Description	9
	1.2.3 Cognitive Process Models	13
	1.3 Potential Problems: Scope and Falsifiability	17
	1.4 Modeling as a “Cognitive Aid” for the Scientist	20
	1.5 <i>In Vivo</i>	22
<b>2</b>	<b>From Words to Models</b>	<b>24</b>
	2.1 Response Times in Speeded-Choice Tasks	24
	2.2 Building a Simulation	26
	2.2.1 Getting Started: R and RStudio	26
	2.2.2 The Random-Walk Model	27
	2.2.3 Intuition vs. Computation: Exploring the Predictions of a Random Walk	31
	2.2.4 Trial-to-Trial Variability in the Random-Walk Model	33
	2.2.5 A Family of Possible Sequential-Sampling Models	37
	2.3 The Basic Toolkit	38
	2.3.1 Parameters	38
	2.3.2 Connecting Model and Data	40
	2.4 <i>In Vivo</i>	40

<b>Part II</b>	<b>Parameter Estimation</b>	45
<b>3</b>	<b>Basic Parameter Estimation Techniques</b>	47
3.1	Discrepancy Function	47
3.1.1	Root Mean Squared Deviation (RMSD)	48
3.1.2	Chi-Squared ( $\chi^2$ )	49
3.2	Fitting Models to Data: Parameter Estimation Techniques	50
3.3	Least-Squares Estimation in a Familiar Context	50
3.3.1	Visualizing Modeling	51
3.3.2	Estimating Regression Parameters	53
3.4	Inside the Box: Parameter Estimation Techniques	57
3.4.1	Simplex	57
3.4.2	Simulated Annealing	61
3.4.3	Relative Merits of Parameter Estimation Techniques	64
3.5	Variability in Parameter Estimates	65
3.5.1	Bootstrapping	65
3.6	<i>In Vivo</i>	70
<b>4</b>	<b>Maximum Likelihood Parameter Estimation</b>	72
4.1	Basics of Probabilities	72
4.1.1	Defining Probability	72
4.1.2	Properties of Probabilities	73
4.1.3	Probability Functions	75
4.2	What Is a Likelihood?	80
4.3	Defining a Probability Distribution	85
4.3.1	Probability Functions Specified by the Psychological Model	86
4.3.2	Probability Functions via Data Models	86
4.3.3	Two Types of Probability Functions	91
4.3.4	Extending the Data Model	92
4.3.5	Extension to Multiple Data Points and Multiple Parameters	93
4.4	Finding the Maximum Likelihood	95
4.5	Properties of Maximum Likelihood Estimators	101
4.6	<i>In Vivo</i>	103
<b>5</b>	<b>Combining Information from Multiple Participants</b>	105
5.1	It Matters How You Combine Data from Multiple Units	105
5.2	Implications of Averaging	106
5.3	Fitting Aggregate Data	109
5.4	Fitting Individual Participants	111
5.5	Fitting Subgroups of Data and Individual Differences	113
5.5.1	Mixture Modeling	113
5.5.2	K-Means Clustering	118
5.5.3	Modeling Individual Differences	121
5.6	<i>In Vivo</i>	123

<b>6</b>	<b>Bayesian Parameter Estimation</b>	126
6.1	What Is Bayesian Inference?	126
6.1.1	From Conditional Probabilities to Bayes Theorem	126
6.1.2	Marginalizing Probabilities	129
6.2	Analytic Methods for Obtaining Posteriors	130
6.2.1	The Likelihood Function	130
6.2.2	The Prior Distribution	131
6.2.3	The Evidence or Marginal Likelihood	134
6.2.4	The Posterior Distribution	135
6.2.5	Estimating the Bias of a Coin	136
6.2.6	Summary	139
6.3	Determining the Prior Distributions of Parameters	139
6.3.1	Non-Informative Priors	139
6.3.2	Reference Priors	142
6.4	<i>In Vivo</i>	143
<b>7</b>	<b>Bayesian Parameter Estimation</b>	146
7.1	Markov Chain Monte Carlo Methods	146
7.1.1	The Metropolis-Hastings Algorithm for MCMC	147
7.1.2	Estimating Multiple Parameters	153
7.2	Problems Associated with MCMC Sampling	160
7.2.1	Convergence of MCMC Chains	161
7.2.2	Autocorrelation in MCMC Chains	162
7.2.3	Outlook	162
7.3	Approximate Bayesian Computation: A Likelihood-Free Method	163
7.3.1	Likelihoods That Cannot be Computed	163
7.3.2	From Simulations to Estimates of the Posterior	164
7.3.3	An Example: ABC in Action	166
7.4	<i>In Vivo</i>	170
<b>8</b>	<b>Bayesian Parameter Estimation</b>	172
8.1	Gibbs Sampling	172
8.1.1	A Bivariate Example of Gibbs Sampling	173
8.1.2	Gibbs vs. Metropolis-Hastings Sampling	176
8.1.3	Gibbs Sampling of Multivariate Spaces	176
8.2	JAGS: An Introduction	177
8.2.1	Installing JAGS	177
8.2.2	Scripting for JAGS	177
8.3	JAGS: Revisiting Some Known Models and Pushing Their Boundaries	182
8.3.1	Bayesian Modeling of Signal-Detection Theory	182
8.3.2	A Bayesian Approach to Multinomial Tree Models: The High-Threshold Model	186
8.3.3	A Bayesian Approach to Multinomial Tree Models	190
8.3.4	Summary	198
8.4	<i>In Vivo</i>	198

x	<b>Contents</b>	
<b>9</b>	<b>Multilevel or Hierarchical Modeling</b>	203
	9.1 Conceptualizing Hierarchical Modeling	203
	9.2 Bayesian Hierarchical Modeling	204
	9.2.1 Graphical Models	204
	9.2.2 Hierarchical Modeling of Signal-Detection Performance	207
	9.2.3 Hierarchical Modeling of Forgetting	211
	9.2.4 Hierarchical Modeling of Inter-Temporal Preferences	218
	9.2.5 Summary	226
	9.3 Hierarchical Maximum Likelihood Modeling	228
	9.3.1 Hierarchical Maximum Likelihood Model of Signal Detection	228
	9.4 Recommendations	233
	9.5 <i>In Vivo</i>	234
	<b>Part III Model Comparison</b>	239
<b>10</b>	<b>Model Comparison</b>	241
	10.1 Psychological Data and the Very Bad Good Fit	241
	10.1.1 Model Complexity and Over-Fitting	243
	10.2 Model Comparison	248
	10.3 The Likelihood Ratio Test	249
	10.4 Akaike's Information Criterion	256
	10.5 Other Methods for Calculating Complexity and Comparing Models	261
	10.5.1 Cross-Validation	262
	10.5.2 Minimum Description Length	262
	10.5.3 Normalized Maximum Likelihood	263
	10.6 Parameter Identifiability and Model Testability	264
	10.6.1 Identifiability	264
	10.6.2 Testability	269
	10.7 Conclusions	270
	10.8 <i>In Vivo</i>	271
<b>11</b>	<b>Bayesian Model Comparison Using Bayes Factors</b>	273
	11.1 Marginal Likelihoods and Bayes Factors	273
	11.2 Methods for Obtaining the Marginal Likelihood	277
	11.2.1 Numerical Integration	278
	11.2.2 Simple Monte Carlo Integration and Importance Sampling	280
	11.2.3 The Savage-Dickey Ratio	284
	11.2.4 Transdimensional Markov Chain Monte Carlo	287
	11.2.5 Laplace Approximation	294
	11.2.6 Bayesian Information Criterion	297
	11.3 Bayes Factors for Hierarchical Models	301
	11.4 The Importance of Priors	303
	11.5 Conclusions	306
	11.6 <i>In Vivo</i>	306



<b>Part IV Models in Psychology</b>	309
<b>12 Using Models in Psychology</b>	311
12.1 Broad Overview of the Steps in Modeling	311
12.2 Drawing Conclusions from Models	312
12.2.1 Model Exploration	312
12.2.2 Analyzing the Model	314
12.2.3 Learning from Parameter Estimates	315
12.2.4 Sufficiency of a Model	316
12.2.5 Model Necessity	318
12.2.6 Verisimilitude vs. Truth	323
12.3 Models as Tools for Communication and Shared Understanding	324
12.4 Good Practices to Enhance Understanding and Reproducibility	326
12.4.1 Use Plain Text Wherever Possible	326
12.4.2 Use Sensible Variable and Function Names	327
12.4.3 Use the Debugger	327
12.4.4 Commenting	328
12.4.5 Version Control	328
12.4.6 Sharing Code and Reproducibility	329
12.4.7 Notebooks and Other Tools	330
12.4.8 Enhancing Reproducibility and Runnability	331
12.5 Summary	332
12.6 <i>In Vivo</i>	332
<b>13 Neural Network Models</b>	334
13.1 Hebbian Models	334
13.1.1 The Hebbian Associator	334
13.1.2 Hebbian Models as Matrix Algebra	339
13.1.3 Describing Networks Using Matrix Algebra	348
13.1.4 The Auto-Associator	349
13.1.5 Limitations of Hebbian Models	356
13.2 Backpropagation	356
13.2.1 Learning and the Backpropagation of Error	360
13.2.2 Applications and Criticisms of Backpropagation in Psychology	364
13.3 Final Comments on Neural Networks	365
13.4 <i>In Vivo</i>	366
<b>14 Models of Choice Response Time</b>	369
14.1 Ratcliff's Diffusion Model	369
14.1.1 Fitting the Diffusion Model	371
14.1.2 Interpreting the Diffusion Model	383
14.1.3 Falsifiability of the Diffusion Model	385
14.2 Ballistic Accumulator Models	386
14.2.1 Linear Ballistic Accumulator	386
14.2.2 Fitting the LBA	388

xii	<b>Contents</b>	
	14.3 Summary	391
	14.4 Current Issues and Outlook	392
	14.5 <i>In Vivo</i>	393
<b>15</b>	<b>Models in Neuroscience</b>	395
	15.1 Methods for Relating Neural and Behavioral Data	397
	15.2 Reinforcement Learning Models	398
	15.2.1 Theories of Reinforcement Learning	398
	15.2.2 Neuroscience of Reinforcement Learning	404
	15.3 Neural Correlates of Decision-Making	410
	15.3.1 Rise-to-Threshold Models of Saccadic Decision-Making	410
	15.3.2 Relating Model Parameters to the BOLD Response	411
	15.3.3 Accounting for Response Time Variability	413
	15.3.4 Using Spike Trains as Model Input	415
	15.3.5 Jointly Fitting Behavioral and Neural Data	417
	15.4 Conclusions	420
	15.5 <i>In Vivo</i>	421
<b>Appendix A</b>	<b>Greek Symbols</b>	424
<b>Appendix B</b>	<b>Mathematical Terminology</b>	425
	<i>References</i>	427
	<i>Index</i>	455

## Illustrations

1.1	An example of data that defy easy description and explanation without a quantitative model.	4
1.2	The geocentric model of the solar system developed by Ptolemy.	5
1.3	Observed recognition scores as a function of observed classification confidence for the same stimuli (each number identifies a unique stimulus).	7
1.4	Observed and predicted classification (left panel) and recognition (right panel).	8
1.5	Sample power law learning function (solid line) and alternative exponential function (dashed line) fitted to the same data.	11
1.6	The representational assumptions underlying GCM.	14
1.7	The effects of distance on activation in the GCM.	15
1.8	Stimuli used in a classification experiment by Nosofsky (1991).	16
1.9	Four possible hypothetical relationships between theory and data involving two measures of behavior (A and B).	19
2.1	Graphical illustration of a simple random-walk model.	25
2.2	Predicted decision-time distributions from the simple random-walk model when the stimulus is non-informative.	31
2.3	Predicted decision-time distributions from the simple random-walk model with a positive drift rate (set to 0.03 for this example).	32
2.4	Predicted decision-time distributions from the modified random-walk model with a positive drift rate (set to 0.035 for this example) and trial-to-trial variability in the starting point (set to 0.8).	35
2.5	Predicted decision-time distributions from the modified random-walk model with a positive drift rate (set to 0.03 for this example) and trial-to-trial variability in the drift rate (set to 0.025).	36
2.6	Overview of the family of sequential-sampling models.	38
2.7	The basic idea: We seek to connect model predictions to the data from our experiment(s).	40
3.1	Data (plotting symbols) from Experiment 1 of Carpenter et al. (2008) (test/study condition) with the best-fitting predictions (solid line) of a power function.	48
3.2	An “error surface” for a linear regression model given by $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$ .	51
3.3	Two snapshots during parameter estimation of a simple regression line.	56
3.4	Two-dimensional projection of the error surface in Figure 3.2.	58

3.5	Probability with which a worse fit is accepted during simulated annealing as a function of the increase in discrepancy ( $\Delta f$ ) and the temperature parameter ( $T$ ).	63
3.6	The process of obtaining parameter estimates for bootstrap samples.	66
3.7	Histograms of parameter estimates obtained by the bootstrap procedure, where data are generated from the model and the model is fit to the generated bootstrap samples.	68
4.1	An example probability mass function: the probability of responding $A$ to exactly $N_A$ out of $N=10$ items in a categorization task, where the probability of an $A$ response to any particular item is $P_A = 0.7$ .	76
4.2	An example cumulative distribution function (CDF).	77
4.3	An example probability density function (PDF).	78
4.4	Reading off the probability of discrete data (top panel) or the probability density for continuous data (bottom panel).	81
4.5	Distinguishing between probabilities and likelihoods.	83
4.6	The probability of a data point under the binomial model, as a function of the model parameter $P_A$ and the data point $N_A$ , the number of $A$ responses in a categorization task.	84
4.7	Different ways of generating a predicted probability function, depending on the nature of the model and the dependent variable.	91
4.8	The joint likelihood function for the Wald parameters $m$ and $a$ given the data set $\mathbf{t} = [0.6 \ 0.7 \ 0.9]$ .	94
4.9	A likelihood function (left panel), and the corresponding log-likelihood function (middle) and deviance function ( $-2 \log$ likelihood; right panel).	97
4.10	A scatterplot between the individual data points (observed proportion $A$ responses for the 34 faces) and the predicted probabilities from GCM under the maximum likelihood parameter estimates.	101
5.1	Simulated consequences of averaging of learning curves.	107
5.2	A simulated saccadic response time distribution from the gap task.	114
5.3	Left panel: Accuracy serial position function for immediate free recall of a list of 12 words presented as four groups of three items. Right panel: Serial position functions for three clusters of individuals identified using K-means analysis.	118
5.4	The gap statistic for different values of $k$ .	120
5.5	A structural equation model for choice RT.	122
6.1	Two illustrative Beta distributions obtained by the R code in Listing 6.1.	133
6.2	Bayesian prior and posterior distributions obtained by a slight modification of the R code in Listing 6.1.	137
6.3	Jeffreys prior, Beta(0.5,0.5), for a Bernoulli process.	140
7.1	MCMC output obtained by Listing 7.1 for different parameter values.	150
7.2	MCMC output obtained by Listing 7.2 for different parameter values.	153
7.3	Experimental procedure for a visual working memory task in which participants have to remember the color of a varying number of squares.	154
7.4	Data (circles) from a single subject in the color estimation experiment of Zhang and Luck (2008) and fits of the mixture model (solid lines).	155

7.5	Posterior distributions of parameter estimates for $g$ and $\sigma_{vM}$ obtained when fitting the mixture model to the data in Figure 7.4.	160
7.6	Overview of a simple Approximate Bayesian Computation (ABC) rejection algorithm.	165
7.7	<b>a.</b> Data from an hypothetical recognition memory experiment in which people respond “old” or “new” to test items that are old or new. <b>b.</b> Signal-detection model of the data in panel <b>a.</b>	167
8.1	Illustration of a Gibbs sampler for a bivariate normal distribution.	174
8.2	Overview of how JAGS is being used from within R.	178
8.3	Output obtained from R using the <code>plot</code> command with an MCMC object returned by the function <code>coda.samples</code> .	181
8.4	<b>a.</b> Data from an hypothetical recognition memory experiment in which people respond “old” or “new” to test items that are old or new. <b>b.</b> Signal-detection model of the data in panel <b>a.</b>	182
8.5	Output from JAGS for the signal detection model illustrated in Figure 8.4.	185
8.6	Convergence diagnostics for the JAGS signal detection model reported in Figure 8.5.	186
8.7	The high-threshold (1HT) model of recognition memory expressed as a multinomial processing tree model.	187
8.8	Output from JAGS for the high-threshold (1HT) model illustrated in Figure 8.7.	190
8.9	<b>a.</b> Autocorrelation pattern for the output shown in Figure 8.8. <b>b.</b> The same autocorrelations after thinning. Only every fourth sample is considered during each MCMC chain.	191
8.10	The no-conflict MPT model proposed by Wagenaar and Boer (1987) to account for performance in the inconsistent-information condition in their experiment.	193
8.11	Output from a run of the no-conflict model for the data of Wagenaar and Boer (1987) using Listings 8.8 and 8.9.	197
8.12	Example of a 95% highest density interval (HDI).	199
8.13	Diagram of the normal model, in the style of the book, <i>Doing Bayesian Data Analysis</i> (Kruschke, 2015).	201
8.14	Diagram of the normal model, in the style of conventional graphical models.	202
9.1	Graphical model for the signal-detection example from Section 8.3.1.	205
9.2	Graphical model for a signal-detection model that is applied to a number of different conditions or participants.	206
9.3	Graphical model for a signal-detection model that is applied to a number of different conditions or participants.	207
9.4	Hierarchical estimates of individual hit rates (left panel) and false alarm rates (right) shown as a function of the corresponding individual frequentist estimates for the data in Table 9.2.	210
9.5	Graphical model for a hierarchical model of memory retention.	213
9.6	Results of a run of the hierarchical exponential forgetting model defined in Listings 9.3 and 9.4.	216
9.7	Posterior densities of the parameters $a$ , $b$ , and $\alpha$ of the hierarchical exponential forgetting model defined in Listings 9.3 and 9.4.	216

xvi	<b>List of Illustrations</b>	
9.8	Results of a run of the hierarchical power forgetting model defined in Listings 9.5 and 9.6.	218
9.9	Graphical model for a hierarchical model of intertemporal choice.	221
9.10	Data from 15 participants of an intertemporal choice experiment reported by Vincent (2016).	223
9.11	Snippet of the data file from the experiment by Vincent (2016) that is used by the R script in Listing 9.8.	224
9.12	Predictions of the hierarchical intertemporal choice model for the experimental conditions explored by Vincent (2016).	226
9.13	Posterior densities for the parameters of the hierarchical intertemporal choice model when it is applied to the experimental conditions explored by Vincent (2016).	227
10.1	Fits of the polynomial law of sensation to noisy data generated from a logarithmic function.	242
10.2	Predictions from a polynomial function of order 2 (left panel) and order 10 (right panel), with randomly sampled parameter values.	243
10.3	An illustration of the bias-variance trade-off.	246
10.4	The bias-variance trade-off. As model complexity (the order of the fitted polynomial) increases, bias decreases and variance increases.	247
10.5	Out-of-set prediction error.	248
10.6	The two functions underlying prospect theory.	251
10.7	K-L distance is a function of models and their parameters.	257
10.8	Prior probability (solid horizontal line) and posterior probabilities (lines labeled $\beta$ and $\epsilon$ ) for two parameters in a multinomial tree model that are “posterior-probabilistically-identified.”	267
11.1	Illustration of how the marginal likelihood can implement the principle of parsimony.	275
11.2	Illustration of the Savage-Dickey density ratio for the signal detection model, examining whether $b \neq 0$ .	286
11.3	Autocorrelations in samples of the model indicator $pM2$ using noninformative pseudo-priors (left panel) and pseudo-priors approximating the posterior (right panel).	294
11.4	Predicted hit and false alarm rates in a change-detection task derived using non-informative (left-hand quadrants) and informative (right-hand quadrants) prior distributions for two models of visual working memory.	305
12.1	A flowchart of modeling.	312
12.2	The effect of the response suppression parameter $\eta$ in Lewandowsky’s (1999) connectionist model of serial recall.	314
12.3	A schematic depiction of sufficiency and necessity.	319
13.1	Architecture of a Hebbian model of associative memory.	335
13.2	Different ways of representing information in a connectionist model.	336
13.3	Schematic depiction of the calculation of an outer product $\Delta \mathbf{W}$ between two vectors $\mathbf{o}$ and $\mathbf{c}$ .	342
13.4	Generalization in the Hebbian model.	346

13.5	Graceful degradation in a distributed model.	348
13.6	A set of 8 orthogonal (Walsh) vectors for an 8-element auto-associator to learn.	350
13.7	Classification performance of the Brain-State-in-a-Box model.	355
13.8	The logistic activation function.	357
13.9	The error between the network output and the target on each sweep.	362
13.10	Multidimensional scaling applied to hidden unit activations early (left) and late (right) in training.	363
14.1	Overview of the family of sequential-sampling models.	370
14.2	Overview of the diffusion model.	370
14.3	Histogram of a hypothetical RT distribution overlaid with quantiles 0.1, 0.3, 0.5, 0.7, and 0.9.	372
14.4	Quantile probability functions predicted by the diffusion model.	373
14.5	QPF for the synthetic data generated and plotted by Listing 14.5.	381
14.6	Graphical representation of a ballistic decision model for a lexical decision (word-nonword) task.	387
14.7	QPF for the synthetic data generated and plotted by Listing 14.8.	390
15.1	Learning of a basic reinforcement action model on the bandit task.	400
15.2	A simple maze. The squares are different states.	402
15.3	Sequencing of choice of action, delivery of reward, and move to a new state.	402
15.4	Learning for three different reinforcement learning models.	403
15.5	Activity in a single dopamine neuron consistent with reward prediction error.	405
15.6	Prediction error in a temporal difference model, at different stages of learning (see Listing 15.2).	408
15.7	The modeling framework for the modeling of FEF carried out by Purcell et al. (2010).	416
15.8	A schematic depiction of the model assumed in Turner et al. (2013) (top panel) and in van Ravenzwaaij et al. (2017) (bottom panel).	418

## Tables

5.1	Berkeley admission data broken down by department	106
6.1	Joint and marginal probabilities	129
7.1	Summary of all approaches to Bayesian parameter estimation that are discussed in this chapter.	147
8.1	Summary of the experiment by Wagenaar and Boer (1987).	192
8.2	Performance of subjects in the experiment by Wagenaar and Boer (1987) for all conditions and predictions of the no-conflict model presented in Listings 8.8 and 8.9	194
9.1	Notation for nodes used in graphical models	205
9.2	Observed and predicted hit and false alarm rates for one run of the hierarchical signal-detection model in Listing 9.2	210
10.1	Summary parameter estimates (means, with standard deviations in brackets) for fits of cumulative prospect theory to the data of Rieskamp (2008)	255
14.1	Comparison of parameter values used to generate synthetic data and the values recovered by fitting the diffusion model	379
14.2	Illustration of the speed-accuracy dilemma in a speeded choice task using data from three hypothetical participants and parameter estimates from fitting the diffusion model	384
A.1	Table of Greek Letters	424
B.1	Scalars, vectors, and functions	425
B.2	Summing, multiplying, and differentiation	425
B.3	Enumeration	425
B.4	Probability	426



## List of Contributors

**Nina R. Arnold**

University of Mannheim

**Amy H. Criss**

Syracuse University

**Chris Donkin**

University of New South Wales

**Birte U. Forstmann**

University of Amsterdam

**Robert M. French**

LEAD-CNRS, University of Burgundy-Franche Comté

**John K. Kruschke**

Indiana University

**Michael Lee**

University of California Irvine

**Jay Myung**

Ohio State University

**Klaus Oberauer**

University of Zurich

**Amy Perfors**

University of Adelaide

**Don van Ravenzwaaij**

University of Groningen

**Jennifer Trueblood**  
Vanderbilt University

**Brandon Turner**  
Ohio State University

**Joachim Vandekerckhove**  
University of California Irvine

**Eric-Jan Wagenmakers**  
University of Amsterdam

**Trisha van Zandt**  
Ohio State University

## Preface

This book presents an integrated approach to the application of computational and mathematical models in psychology. Computational models have been extensively applied to better understand many domains of human behavior, such as perception, memory, reasoning, decision-making, communicating, and deciding. Modeling is often applied in these areas to different purposes – measurement, prediction, and model testing. Our major goal here is to provide a unified view on the interface between theories, simulations, and data, with a view to answering the central question: how can we learn from models of behavior?

We cover several topics. Part I of the book explains what a computational model is and gives a general overview of models that have been applied to understanding human behavior. We also examine the process of converting theoretical statements into simulation code and give an overview of the various concepts required to understand modeling. Part II examines one use of models: parameter estimation. By fitting models to data, inferences can be made from the resulting parameter estimates, and statements made about the psychological mechanism(s) or representations that generated those data. We cover maximum likelihood estimation and Bayesian estimation, including estimation across multiple participants and hierarchical estimation. Part III explores how inferences can be made from models by using model comparison. We consider under what conditions statements of sufficiency and necessity can be made from data, and how model complexity can be conceptualized and quantified. Part III examines several approaches to accounting for complexity in model comparison, including information criteria and Bayes Factors. Part IV considers the role of computational modeling in advancing psychological theory. We explore use of models as adjuncts to human reasoning, and the interaction between human and artificial intelligence to guide theorizing and generation of conceptual insights. We also consider the use of models as tools to arrive at shared understanding between researchers (i.e. the use of models as common terms of reference), and practices for communicating and sharing models. We finish by giving an overview of the application of models in several popular areas: neural network models, models of choice response time, and the application of models to understand neural data.

To accomplish all this, we use a freely available computer language, called R, which was initially developed for statistical data analysis but has broad applicability and is now used by many modellers.

Some readers may know that we wrote a seemingly similar book some time ago (Lewandowsky and Farrell, 2011). The present book retains some of the features of the earlier book that seemed to be appreciated by readers – for example, we try to explain the important features in all our snippets of source code. Thus, while this is not a textbook in R programming, the book does point to the most important aspects of our programs that are relevant to the task at hand, namely how to understand the human mind by computational means. Beyond that, however, the present book is very different from our earlier volume. Whereas the earlier book was an introductory textbook, the present volume aspires to more lofty goals: we want to take the reader to the leading edge of current modeling practice, and we introduce several novel developments in the course of doing so.

As well as providing simulation code in the R language to complement the equations and descriptions in the text, each chapter ends with an *in vivo* section. For each *in vivo* example, we asked a researcher to share their experiences in working on that topic or method, some consideration of the philosophy of science in that area, or a counterpoint to our own views. We think these sections are insightful and illuminating (and amusing!), and we are very grateful to other members of the field for giving us the opportunity to share their thoughts with you.

As well as the authors of the *in vivo* sections throughout the book, we would like to thank the numerous friends and colleagues with whom we have discussed many issues in preparing this book. In particular, we thank Henrik Singmann and Benjamin Vincent for their comments on drafts of chapters in which their work was cited and used. We would also like to thank the instructors (Gordon Brown, Amy Criss, Adele Diederich, Chris Donkin, Bob French, Cas Ludwig, Klaus Oberauer, Jörg Rieskamp, Lael Schooler, Joachim Vandekerckhove, and Eric-Jan Wagenmakers) and students of the four European Summer Schools on Computational and Mathematical Modeling of Cognition that we have conducted over the past eight years, and that have attracted more than 120 students to date. Their feedback on drafts of this book have been invaluable and we thank students and instructors for many enthusiastic discussions. One thing that has been affirmed for us through these discussions is that models are used in many different ways in psychology. In presenting a unified and integrative theoretical framework for modeling, we have attempted to capture this variance, but recognize that there are many models and points of view that we could not explore here. We would also like to thank Janka Romero and her predecessor, Hetty Marx, at Cambridge University Press for their help and encouragement whilst proposing and writing the book, and Adam Hooper, Anup Kumar, Christina Taylor, and Sindhuja Ayyappan for their help during production.