Mechanisms in Classical Conditioning

A Computational Approach

What mechanisms are involved in enabling us to generate predictions of what will happen in the near future? Although we use associative mechanisms as the basis to predict future events, such as using cues from our surrounding environment, timing, attentional and configural mechanisms are also needed to improve this function. Timing mechanisms allow us to determine when those events will take place. Attentional mechanisms ensure that we keep track of cues that are present when unexpected events occur and that we disregard cues present when everything happens according to our expectations. Configural mechanisms make it possible to combine separate cues into one signal that predicts an event different from that predicted individually by separate cues. Written for graduates and researchers in neuroscience, computer science, biomedical engineering and psychology, the author presents neural network models that incorporate these mechanisms and shows, through computer simulations, how they explain the multiple properties of associative learning.

DR. SCHMAJUK has been an Associate Professor of Biomedical Engineering in Buenos Aires (Argentina), an Assistant Professor of Psychology at Northwestern University, and is presently a Professor of Psychology and Neuroscience at Duke University. Here he has developed several neural network models of classical conditioning, operant conditioning, animal communication, creativity, spatial learning, cognitive mapping and prepulse inhibition. Previous books by this author: *Animal Learning and Cognition. A Neural Network Approach* (1997), *Occasion Setting. Associative Learning and Cognition in Animals* (1998), *Latent Inhibition and its Neural Substrates* (2002).

Mechanisms in Classical Conditioning

A Computational Approach

Nestor Schmajuk

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore,
São Paulo, Delhi, Dubai, Tokyo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521887809

First published 2010

Printed in the United Kingdom at the University Press, Cambridge

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

Library of Congress Cataloguing in Publication data
Schmajuk, Nestor A.
Mechanisms in classical conditioning : a computational approach / Nestor Schmajuk.
p. cm.
ISBN 978-0-521-88780-9 Hardback
1. Conditioned response–Computer simulation. 2. Neural networks
(Computer science) I. Title.
BF319.S357 2010
153.1′526–dc22 2009036649

ISBN 978-0-521-88780-9 Hardback

To my wife, Mabel, and my children Mariana, Gabriela, and Hans.

Contents

Preface *ix*
Acknowledgments *xi*
Abbreviations *xv*

Part I Introduction 1

1 Classical conditioning: data and theories 3

Part II Attentional and associative mechanisms 19

2 An attentional–associative model of conditioning 21
3 Simple and compound conditioning 34
4 The neurobiology of fear conditioning 57
5 Latent inhibition 64
6 The neurobiology of latent inhibition 93
7 Creativity 119
8 Overshadowing and blocking 137
9 Extinction 165
10 The neurobiology of extinction 234

Part III Configural mechanisms 243

11 A configural model of conditioning 245
12 Occasion setting 257
13 The neurobiology of occasion setting 316

Part IV Attentional, associative, configural and timing mechanisms 357

14 Configuration and timing: timing and occasion setting 359

15 Attention and configuration: extinction cues 392

16 Attention, association and configuration: causal learning and inferential reasoning 403

Part V Conclusion: mechanisms of classical conditioning 415

17 Conclusion: mechanisms of classical conditioning 417

References 425
Author Index 460
Subject Index 468

Preface

This book extends the application of the neural network models described in my previous book on “Animal learning and cognition: A neural network approach” to a whole range of important classical conditioning paradigms, including recovery from overshadowing, recovery from blocking, backward blocking and recovery from backward blocking; extinction, and occasion setting, as well as the neurophysiology of some of those phenomena.

In the last decades, models of conditioning have shown increasing completeness and precision. This book describes a number of computational mechanisms (associations, attention, configuration, and timing) that first seemed necessary to explain a small number of conditioning results and then proved able to account for a large part of the extensive body of conditioning data. These computational mechanisms are implemented by artificial neural networks, which can be mapped onto different brain structures. Therefore, the approach permits to establish clear brain-behavior relationships.

The book is organized as follows. Part I presents major classical conditioning data and describes several theories proposed to explain them. Part II presents a neural network theory, which includes attentional and associative mechanisms, and applies it to the description of conditioning, latent inhibition, overshadowing and blocking, extinction, and creative processes. In addition, it examines the neurobiological bases of latent inhibition and extinction. Part III describes another neural network, which includes configural mechanisms, and applies it to the description of occasion setting. In addition, it examines the neurobiological bases of occasion setting. Finally, Part IV shows how the combination of attentional, associative, configural and timing mechanisms applies to timing in occasion setting, extinction cues, causal learning and inferential reasoning.

In addition to summarizing recent progress in conditioning theory, the material in the book suggests directions for future efforts. The ability of the models both to replicate and explain behavior opens the door to the design of novel experiments, as well as to the conception of more effective therapeutic methods to treat some psychological disorders and the development of adaptive robots.

Acknowledgments

This book reflects the work during the years I have spent so far at Duke University. I want to thank my collaborators during this time, including Ying Wan Lam, Jeffrey Gray, Catalin Buhusi, Jef Lamoureux, Peter Holland, Landon Cox, Beth Christiansen, Jose Larrauri, Kevin LaBar, Gunes Kutlu, Diana Aziz and Margaret Bates. Many of them were my colleagues at the time of our collaborations; others were students who later became colleagues.

I thank the many people who gave me feedback at different stages of this project. The list is long and includes Andy Barto, Mark Bouton, Cody Brooks, Gonzalo de la Casa, Jennifer Higa, John Moore, Santiago Pellegrini, Oskar Pineño, Juan Manuel Rosas, John Staddon, Mabel Tyberg, Lise Wallach and Michael Wallach. I am also grateful to my editor, Martin Griffiths, for his help during the preparation of the book.

This book includes material from the following papers:

1. Schmajuk, N.A., Lam, Y.W. & Gray, J.A. (1996). Latent inhibition: a neural network approach. *Journal of Experimental Psychology: Animal Behavior Processes*, **22**, 321–349. (With kind permission from the American Psychological Association)
2. Buhusi, C. & Schmajuk, N.A. (1996). Attention, configuration, and hippocampal function. *Hippocampus*, **6**, 621–642. (With kind permission from John Wiley & Sons, Inc.)
3. Schmajuk, N.A. & Buhusi, C. (1997). Occasion setting, stimulus configuration, and the hippocampus. *Behavioral Neuroscience*, **111**, 235–258. (With kind permission from the American Psychological Association)
4. Schmajuk, N.A., Lamoureux, J. & Holland, P.C. (1998). Occasion setting and stimulus configuration: a neural network approach. *Psychological Review*, **105**, 3–32. (With kind permission from the American Psychological Association)

5. Schmajuk, N.A. & Buhusi, C.V. (1997). Spatial and temporal cognitive mapping: a neural network approach. *Trends in Cognitive Sciences*, **1**, 109–114. (With kind permission from Elsevier)
6. Buhusi, C.V., Gray, J.A. & Schmajuk, N.A. (1998). The perplexing effects of hippocampal lesions on latent inhibition: a neural network solution. *Behavioral Neuroscience*, **112**, 316–351. (With kind permission from the American Psychological Association)
7. Schmajuk, N.A., Buhusi, C.V. & Gray, J.A. (1998). The pharmacology of latent inhibition: a neural network approach. *Behavioural Pharmacology*, **9**, 711–730. (With kind permission from Lippincott, Williams & Wilkins)
8. Buhusi, C.V. & Schmajuk, N.A. (1999). Timing in simple conditioning and occasion setting: a neural network approach. *Behavioral Processes*, **45**, 33–57. (With kind permission from Elsevier)
9. Schmajuk, N.A., Cox, L. & Gray, J.A. (2001). Nucleus accumbens, entorhinal cortex and latent inhibition: a neural network approach. *Behavioral Brain Research*, **118**, 123–141. (With kind permission from Elsevier)
10. Schmajuk, N.A., Christiansen, B.A. & Cox, L. (2000). Haloperidol reinstates latent inhibition impaired by hippocampal lesions: data and theory. *Behavioral Neuroscience*, **114**, 659–670. (With kind permission from the American Psychological Association)
11. Schmajuk, N.A. (2001). Hippocampal dysfunction in schizophrenia. *Hippocampus*, **11**, 599–613. (With kind permission from John Wiley & Sons, Inc.)
12. Schmajuk, N.A., Gray, J.A. & Larrauri, J.A. (2005). A pre-clinical study showing how dopaminergic drugs administered during pre-exposure can impair or facilitate latent inhibition. *Psychopharmacology*, **177**, 272–279. (With kind permission of Springer Science and Business Media)
13. Schmajuk, N.A. (2005). Brain-behaviour relationships in latent inhibition: a computational model. *Neuroscience and Biobehavioral Reviews*, **29**, 1001–1020. (With kind permission from Elsevier)
14. Schmajuk, N.A. & Larrauri, J.A. (2006). Experimental challenges to theories of classical conditioning: application of a computational model of storage and retrieval. *Journal of Experimental Psychology: Animal Behavior Processes*, **32**, 1–20. (With kind permission from the American Psychological Association)
15. Schmajuk, N.A., Larrauri, J.A. & LaBar, K. (2006). Reinstatement of conditioned fear: an attentional-associative model. *Behavioral Brain Research*, **177**, 242–253. (With kind permission from Elsevier)
16. Larrauri, J.A. & Schmajuk, N.A. (2008). Attentional, associative, and configural mechanisms in extinction. *Psychological Review*, **115**, 640–676. (With kind permission from the American Psychological Association)

17. Schmajuk N. A. & Larrauri J. A. (2008). Associative models can describe both causal learning and conditioning. *Behavioural Processes*, **77**, 443–445. (With kind permission from Elsevier)
18. Schmajuk, N. A., Aziz, D. R. & Bates, M. J. B. (2009). Attentional-associative interactions in creativity. *Creativity Research Journal*, **21**, 92–103. (With kind permission from Taylor & Francis Group)
19. Schmajuk, N. (2008). Attentional and error-correcting associative mechanisms in classical conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, in press. (With kind permission from the American Psychological Association)
20. Schmajuk, N. (2010). From latent inhibition to retrospective revaluation: an attentional–associative model. In *Latent Inhibition*, eds. I. Weiner & B. Lubow. New York: Cambridge University Press. (With kind permission from Cambridge University Press)
21. Schmajuk, N. A. & Kutlu, G. M. (2009). The computational nature of associative learning. *Behavioral Brain Science*, **32**, 223–224. (With kind permission from Cambridge University Press)
22. Dunsmoor, J., and Schmajuk, N. A. (2009). Interpreting patterns of brain activation in human fear conditioning with an attentional–associative learning model. *Behavioral Neuroscience*, **123**, 851–855.

Abbreviations

ACQ	acquisition
Amyg	amygdala
an_j	output of the hidden units
as_i	output of the input units
BB	backward blocking
B_{CS}	aggregate prediction of CS
B_{CX}	aggregate prediction of CX
$\overline{B}_k$	average aggregate prediction of event k
BS	between-subject
BT	buffer threshold
B_{US}	aggregate prediction of the US
BW	between-subject with interspersed water presentations
CA	cornu ammonis
CAT	conditioned attention theory
CAQ	creative achievement questionnaire
CC	classical conditioning
CER	conditioned emotional response
CL	cortical lesion
CN	central nucleus *or* configural stimulus
CPS	creative personality scale
CR	conditioned response (strength)
CS	conditioned stimulus (intensity)
CSg	intensity of common elements (generalized) in different contexts
CXc	intensity of contextual cues in the training cage
CXg	intensity of contextual cues shared by all contexts
CXh	intensity of contextual cues in the home cage
DA	dopaminergic
dlPFC	dorsolateral prefrontal cortex

EC	entorhinal cortex *or* extinction cue
EH	error signal for hidden units
EO	error signal for output units
fMRI	functional magnetic resonance imaging
FN	feature negative
FP	feature positive
FTI	feature–target interval
GABA	γ-aminobutyric acid
HAL	haloperidol
HFL	hippocampus formation lesion
HP	hippocampus proper
HPL	hippocampus proper lesion
ISI	interstimulus interval
ITI	intertrial interval
LI	latent inhibition
$\bar{\lambda}_k$	average observed value of event k
LN	lateral nucleus
LTM	long-term memory
MGB	medial geniculate body
NAC	nucleus accumbens
NCX	neocortex
NM	nictitating membrane
NMDA	*N*-methyl D-aspartate
NMR	nictitating membrane response
Novelty⊠	total novelty normalized between 0 and 1
NP	negative patterning
OR	orienting response (strength)
PFC	prefrontal cortex
PIN	posterior intralaminar nucleus
PP	positive patterning
PRE	preresponse
PREE	partial reinforcement extinction effect
PUP	paired–unpaired–paired
RAT	remote associates test
RF_E	reinforcement excitatory
RF_I	reinforcement inhibitory
RRP	rest–rest–paired
RW	model Rescorla–Wagner model
SAL	saline

SCR	skin conductance response
SD model	Schmajuk–DiCarlo model
SL	sham lesion
SLG model	Schmajuk–Lam–Gray model
SLH model	Schmajuk–Lamoureux–Holland model
SOP model	Wagner's Sometimes Opponent Process Standard Operating Procedures model
STM	short-term memory
SUB	subiculum
τ_{CS}	trace of the CS
τ_{CX}	trace of the CX
THAL	thalamus
TTX	tetrodotoxin
t.u.	time unit
US	unconditioned stimulus
UUP	unpaired–unpaired–paired
$V_{CS,CX}$	association between CS and CX
$V_{CS,US}$	association between CS and US
$V_{CS4,CS2}$	association between CS_1 and CS_2
VEH	vehicle
VH	CS–CN association
VN	CN–US association
VP	ventral pallidum
VS	CS–US association
VTA	ventral tegmental area
WS	within-subject
X_{CS}	representation of the CS
X_{CX}	representation of the CX
z_{CS}	attention to the CS
z_{CX}	attention to the CX