THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e. g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

INSECT NEUROETHOLOGY OF REINFORCEMENT LEARNING

EVRIPIDIS GKANIAS

Doctor of Philosophy
School of Informatics
University of Edinburgh

2022

Evripidis Gkanias:
Insect neuroethology of reinforcement learning
Doctor of Philosophy, 2022

SUPERVISORS:
Professor Barbara Webb, Ph.D.
Professor Subramanian Ramamoorthy, Ph.D.

EXAMINERS:
Professor Thomas Nowotny, Ph.D.
Professor J. Douglas Armstrong, Ph.D.

ABSTRACT

Historically, reinforcement learning is a branch of machine learning founded on observations of how animals learn. This involved collaboration between the fields of biology and artificial intelligence that was beneficial to both fields, creating smarter artificial agents and improving the understanding of how biological systems function. The evolution of reinforcement learning during the past few years was rapid but substantially diverged from providing insights into how biological systems work, opening a gap between reinforcement learning and biology. In an attempt to close this gap, this thesis studied the insect neuroethology of reinforcement learning, that is, the neural circuits that underlie reinforcement-learning-related behaviours in insects. The goal was to extract a biologically plausible plasticity function from insect-neuronal data, use this to explain biological findings and compare it to more standard reinforcement learning models. Consequently, a novel dopaminergic plasticity rule was developed to approximate the function of dopamine as the plasticity mechanism between neurons in the insect brain. This allowed a range of observed learning phenomena to happen in parallel, like memory depression, potentiation, recovery, and saturation. In addition, by using anatomical data of connections between neurons in the mushroom body neuropils of the insect brain, the neural incentive circuit of dopaminergic and output neurons was also explored. This, together with the dopaminergic plasticity rule, allowed for dynamic collaboration amongst parallel memory functions, such as acquisition, transfer, and forgetting. When tested on olfactory conditioning paradigms, the model reproduced the observed changes in the activity of the identified neurons in fruit flies. It also replicated the observed behaviour of the animals and it allowed for flexible behavioural control. Inspired by the visual navigation system of desert ants, the model was further challenged in the visual place recognition task. Although a relatively simple encoding of the olfactory information was sufficient to explain odour learning, a more sophisticated encoding of the visual input was required to increase the separability among the visual inputs and enable visual place recognition. Signal whitening and sparse combinatorial encoding were sufficient to boost the performance of the system in this task. The incentive circuit enabled the encoding of increasing familiarity along a known route, which dropped proportionally to the distance of the animal from that route. Finally, the proposed model was challenged in delayed reinforcement tasks, suggesting that it might take the role of an adaptive critic in the context of reinforcement learning.

LAY SUMMARY

This thesis tried to bridge the gap between the fields of reinforcement learning and animal behaviour, by exploring how the neural circuit of the fruit flies can implement reinforcement learning principles. A common assumption for a specific brain area of fruit flies is that it transforms sensory input into the behaviour of the animal. Therefore, a novel rule was proposed in order to describe how motivational neurons alter their response to a variety of sensory stimuli, like odours and views, which suggested potential mechanisms for the acquisition, forgetting, and consolidation of memories in the brain. Along with a novel circuit of neurons, which was accurately extracted by the neuroanatomy of the fly brain, this rule allowed for the formation of short- and long-term memories, as well as transfer between them. This circuit was further implemented in computer simulations, explaining a big volume of behavioural data and neural recordings from the fruit-fly brain, and providing insights into the underlying mechanisms driving them. Inspired by the visual navigation capabilities of desert ants, the same circuit was further challenged in recognising a previously experienced route by learning its visual surroundings. The results suggested that the circuit encoded a relative familiarity value for the experienced views, which was increasing along the familiar route. This increase was proportional to the proximity of the animal to that route, which could allow the animal to calculate its navigating direction in order to successfully follow the route. The circuit was also challenged by more complicated tasks, like driving a taxi or avoiding a cliff, where a strategy was required in order to find a solution. The results showed that the proposed circuit alone cannot solve these tasks and suggested that a collaboration between the modelled and other brain regions could allow the animals to develop a strategy.

ACKNOWLEDGEMENTS

This is one of the few opportunities in life to formally acknowledge the people that have had a positive impact on you and your career and I will surely exploit it fully.

First and foremost, I would like to thank my supervisor and mentor Barbara Webb for introducing me to the field of biorobotics and computational neuroethology, and for creating an intellectual and diverse environment that allowed me to dream big and different. Barbara's mechanistic perspective on animals and her passion for the neural circuit and behaviour of insects inspired me to pursue my study in insect neuroethology from a computational perspective. So, I thank her for that, for trusting me in pursuing numerous projects before and during my doctorate studies, and for providing opportunities for me to study the behaviour of insects in their natural habitat, which I found very inspiring and informative for my research. I would also like to thank her for her enormous support during my studies, putting her personal needs aside, and providing extensive feedback and exceptionally long meetings when I needed them the most. Not everyone would do that, and I feel truly blessed that Barbara was part of my research journey.

I would also like to thank other people that were directly involved in this study. I am grateful to my thesis committee, Subramanian Ramamoorthy and Matthias Hennig, who provided insightful discussions and feedback during my annual review meetings. Subramanian was particularly helpful in putting my computational models in the context of reinforcement learning ensuring a robotics perspective to my thesis, while Matthias was equally helpful regarding neuro-computational modelling methods. Thanks to my thesis examiners, Thomas Nowotny and Douglas Armstrong, for their valuable feedback that improved this thesis. I also thank Li Yan McCurdy and Michael Nitabach from Yale School of Medicine for providing the data used in Chapter 3, without which this study would have been incomplete. Also thanks for our lengthy discussions around their data and potential circuits that can explain them. I am also grateful for the help of Vanessa Ruta, who kindly provided the data that supported the dopaminergic plasticity rule in Chapter 2, and of James Bennett, who commented on the data used in Section 4.2.2. Thanks to Edinburgh Centre for Robotics for providing the funding and support throughout this study. I specifically thank Anne Murphy for being such a wonderful administrator, Sethu Vijayakumar and David Lane for securing the funding for all the students of the Centre for Doctoral Training in Robotics and Autonomous Systems, and the UKRI for providing the funding.

I would also like to thank the old and new members of the Insect Robotics group that had a positive impact on me and my thesis. More specifically, I would like to thank Thomas Stone, Benjamin Risse, Kostas Lagogiannis, Jane Loveless, Tianqi Wei, James Garforth, Jan Stankiewicz, Theodoros Stouraitis, Ioannis Pisokas, Le Zhu, Roman Goulard, Emily Rolley-Parnell, Robert Mitchell, Anna Hadjitofi, Florent le Moël, Rana El Khouri Maroun, Aruna Raman, Mohamed Sorour, and Yihe Lu for the useful discussions and debates both in personal and academic aspects. A big thanks to Robert and Theodoros, who proofread big parts of my thesis; also thanks to Yihe, Anna, Aruna, and Rana, who proofread smaller parts. Thanks to my bachelor's and master's students, Jiewen Deng, Xuechun Qiao, Yijie Chen, Komal Afzal, and Alina Scaria, for their courage and trust in my supervising of their research projects.

I am also grateful to the mentors I worked with in the past because they were part of my research journey and they contributed to what I am today. I thank Grigorios Tsoumakas for introducing me to multi-label data and machine learning through neural networks during my bachelor thesis, Petros Daras for trusting me to work on European projects for computer vision, graphics, and machine learning, although I was naïve in the research community, and Michael Mangan for teaching me how to conduct experiments with desert ants in the field, how to design and realise experimental set-ups, and how to collect data. I also thank Antoine Wystrach, Sebastian Schwarz, Xim Cerda, Cody Freas, Cornelia Buehlmann, Leo Clement, and Florent le Moël for their support during numerous field seasons with desert ants in Seville, as well as my volunteers from the University of Seville, Carlos Merino, Gabriel Rivas Mena, and Manuel Serrano Jimenez for helping me realise my ambitious experiments.

I am grateful for the unconditional support of my family and friends, who always reminded me of what is important in life and that everything is a matter of perspective. Thanks to my parents, Achilleas and Georgia, who made me who I am today, and ensured a stable and happy childhood for me. Also, for being patient and supportive while I was away from home for so long. Thanks to my sisters, Maria and Nafsika, for always being enthusiastic about my studies and for boosting my confidence in explaining my work. Thanks to my friends, Stergios, Theodoros, Jiayi, Giannis, Kostas, Eleftheria, Ioanna, Sotiria, Apostolos, Lena, Stavros, Antigone, and Virginia who were in touch with me and reminded me of my achievements and mistakes when I needed it the most. A special thanks to Stergios, Theodoros, and Jiayi for being my Scottish family and keeping me social during my studies and the pandemic.

Finally, I would like to thank my partner Paschalina for being there when I needed it the most. Thanks for her kindness and understanding, for moving to Scotland to keep our relationship alive, and for feeding me goods that kept me going during the final stages of my thesis.

PUBLICATIONS

Parts of the research leading to this thesis have previously appeared in the following peer-reviewed publications. Some passages have been quoted verbatim from the respective sources.

JOURNAL ARTICLES

- E. Gkanias, B. Risse, M. Mangan, and B. Webb (2019). "From skylight input to behavioural output: a computational model of the insect polarised light compass". *PLoS Computational Biology*, 15(7), e1007123. (Appendix D)
- E. Gkanias, L. Y. McCurdy, M. N. Nitabach, and B. Webb (2022). "An incentive circuit for memory dynamics in the mushroom body of *Drosophila melanogaster*". *Elife* 11, e75611. (Chapter 2, Chapter 3, and Chapter 4)

CONFERENCE PRESENTATIONS

- E. Gkanias, A. Scaria, N. A. Vladis, B. Risse, M. Mangan, and B. Webb (2019). "Robustness of a model of the insects' celestial compass in realistic conditions". *International Conference on Invertebrate Vision*. (Appendix D)
- E. Gkanias, B. Risse, M. Mangan, and B. Webb (2019). "From skylight input to behavioural output: a computational model of the insect polarised light compass". *International Navigation Conference*. (Appendix D)
- E. Gkanias, L. Y. McCurdy, M. N. Nitabach, and B. Webb (2021). "How flies acquire, forget and assimilate memories: a computational perspective". *Mushroom Body Meeting*. (Chapter 3)
- E. Gkanias, L. Y. McCurdy, M. N. Nitabach, and B. Webb (2021). "The incentive circuit of the fruit fly brain: a computational perspective". *Bernstein Conference*. (Chapter 2, Chapter 3, and Section 4.2.3)
- E. Gkanias, L. Y. McCurdy, M. N. Nitabach, and B. Webb (2022). "An anatomically accurate circuit for short- and long-term motivational learning in fruit flies". *Cosyne Conference*. (Chapter 2, Chapter 3, and Chapter 4)
- E. Gkanias, and B. Webb (2022). "How could the mushroom body and central complex combine for visual homing in insects?". *International Congress of Neuroethology*. (Chapter 5)

DECLARATION

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

Edinburgh, 2022

Evripidis Gkanias,
December 20, 2022

"Begin at the beginning," the King said gravely,
"and go on till you come to the end: then stop."
— Lewis Carroll, *Alice in Wonderland*

CONTENTS

1 INTRODUCTION 1

1.1 Introduction to computational neuroethology of insects . . . 3

1.2 Scope . . . 4

1.3 Contributions . . . 5

2 SYNAPTIC PLASTICITY 7

2.1 Related work . . . 8

2.1.1 Reward prediction error rule . . . 9

2.1.2 Basic rule . . . 10

2.1.3 Template rule . . . 10

2.1.4 Hebbian rule . . . 11

2.1.5 Neural-modulation rule . . . 12

2.1.6 Stochastic Hebbian rule . . . 12

2.1.7 Anti-Hebbian rule . . . 13

2.2 Results . . . 13

2.2.1 The dopaminergic plasticity rule . . . 14

2.2.2 Derivation of the dopaminergic plasticity rule . . . 16

2.3 Discussion . . . 18

2.4 Methods . . . 20

2.4.1 Backward conditioning experiments . . . 20

2.4.2 Normalised mean change . . . 22

3 MEMORY DYNAMICS 23

3.1 The insect brain . . . 24

3.1.1 The mid-brain . . . 24

3.1.2 The mushroom body . . . 26

3.1.3 Convergence zones . . . 36

3.2 Related work . . . 36

3.2.1 Firing-rate models . . . 37

3.2.2 Spiking neural network models . . . 41

3.2.3 Neuromorphic models . . . 44

3.3 Results . . . 45

3.3.1 The incentive circuit . . . 45

3.3.2 Susceptible & restrained memories . . . 47

3.3.3 Reciprocal short-term memories . . . 50

3.3.4 Long-term memory . . . 52

3.3.5 Reciprocal long-term memories . . . 54
3.3.6 Memory assimilation mechanism . . . 55
3.4 Discussion . . . 56
3.4.1 Predictions of the computational model . . . 57
3.4.2 Additional mushroom body connections . . . 58
3.4.3 Beyond attraction and aversion . . . 60
3.5 Methods . . . 62
3.5.1 Parameters of the model . . . 62
3.5.2 Forward propagation . . . 66
3.5.3 Modelling the neural responses . . . 67
3.5.4 Validation of the dopaminergic factor . . . 68
3.5.5 Data collection . . . 69
3.5.6 Calculating off- and on-shock values . . . 70
4 OLFACTION AS A CASE STUDY 71
4.1 Background . . . 72
4.1.1 Conditioning types . . . 72
4.1.2 Olfactory conditioning experimental paradigms . . . 73
4.1.3 Neural activity interventions . . . 80
4.2 Results . . . 81
4.2.1 T-maze experiments . . . 82
4.2.2 Neural activity intervention experiments . . . 86
4.2.3 Arena experiments . . . 88
4.3 Discussion . . . 91
4.3.1 Advantages of the dopaminergic plasticity rule . . . 91
4.3.2 The effects of olfactory conditioning . . . 92
4.4 Methods . . . 94
4.4.1 T-maze experiments . . . 95
4.4.2 Neural activity intervention experiments . . . 96
4.4.3 Arena experiment . . . 97
5 VISUAL PLACE RECOGNITION 101
5.1 Background . . . 102
5.1.1 Insect vision . . . 102
5.1.2 Visual projection neurons . . . 104
5.1.3 Sparse coding in the Kenyon cells . . . 105
5.1.4 The behavioural readout . . . 106
5.2 Related work . . . 106
5.2.1 Rendering the visual surroundings . . . 106
5.2.2 Visual place recognition . . . 109

5.3 Results . . . 111
5.3.1 Decorrelation of the visual input . . . 113
5.3.2 Efficient sparse coding and combinatorial optimisation . . . 114
5.3.3 Predicting the familiarity . . . 115
5.3.4 Route following by differential familiarity . . . 117
5.4 Discussion . . . 120
5.4.1 Visual projection neurons . . . 120
5.4.2 Sparse coding by the Kenyon cells . . . 121
5.4.3 The effect of consistent familiarity . . . 123
5.4.4 Integration with the central complex . . . 125
5.5 Methods . . . 128
5.5.1 The simulated environment . . . 128
5.5.2 Rendering the environment . . . 128
5.5.3 Principle component projection neurons . . . 131
5.5.4 Kenyon cells combinatorial coding . . . 132
5.5.5 Running the experiments . . . 135
5.5.6 Steepness analysis . . . 138
6 DELAYED REINFORCEMENTS 141
6.1 Background . . . 142
6.1.1 Associative reinforcement learning . . . 144
6.1.2 Sequential reinforcement learning . . . 145
6.1.3 Standard benchmarks . . . 148
6.2 Results . . . 150
6.2.1 Spatially correlated states . . . 151
6.2.2 Temporally correlated states . . . 152
6.3 Discussion . . . 153
6.3.1 The representation of Q-values . . . 154
6.3.2 Temporally correlated representations . . . 155
6.3.3 Reinforcement learning architecture of the insect brain . . . 158
6.4 Methods . . . 160
6.4.1 The OpenAI gym simulations . . . 160
6.4.2 Representations of the state . . . 160
6.4.3 Eligibility trace . . . 161
6.4.4 The incentive circuit for multiple actions . . . 162
6.4.5 Sarsa and Q-learning . . . 162
6.4.6 Calculating the average reward . . . 163
7 CONCLUSION 165
7.1 Overview . . . 166

7.1.1 The dopaminergic plasticity rule . . . 166
7.1.2 The incentive circuit . . . 167
7.1.3 Memory dynamics for olfactory conditioning . . . 168
7.1.4 The visual place recognition task . . . 169
7.1.5 Delayed reinforcements . . . 171
7.2 Future perspectives . . . 172
7.3 Epilogue . . . 173
A THE INCENTIVE WHEEL 177
B LATERAL INHIBITION AMONG OMMATIDIA 181
B.1 Results and discussion . . . 181
B.2 Methods . . . 182
C ZERNIKE MOMENTS FROM OMMATIDIA VALUES 185
C.1 Results and discussion . . . 185
C.2 Methods . . . 186
D THE POLARISED-LIGHT COMPASS 189
D.1 Methods . . . 192
D.1.1 Overview . . . 192
D.1.2 Skylight . . . 193
D.1.3 The insect eye . . . 194
D.1.4 The compass . . . 197
D.1.5 Central complex . . . 204
D.1.6 Evaluation . . . 204
D.2 Results . . . 206
D.2.1 Compass accuracy without tilt . . . 206
D.2.2 Effects of head tilt . . . 207
D.2.3 Exploration of the structural parameters . . . 209
D.2.4 Path integration . . . 209
D.2.5 Experimental paradigm . . . 212
D.3 Discussion . . . 213
D.3.1 Obtaining solar azimuth from polarisation information . . . 214
D.3.2 Neurobiological plausibility . . . 215
D.3.3 The sensor array . . . 217
D.3.4 POL-compass design for robotics . . . 217
D.3.5 Conclusion . . . 219
E SUPPLEMENTARY MATERIAL 221
BIBLIOGRAPHY 255

LIST OF FIGURES

Figure 2.1 The effects of the dopaminergic plasticity rule 15
Figure 2.2 The effect of the ER−Ca^{2+} and cAMP based on the order of the conditioned and unconditioned stimuli 18
Figure 2.3 Parameter exploration for τ_{short} and τ_{long} 21
Figure 3.1 The mid-brain of fruit flies 25
Figure 3.2 Overview of the mushroom body circuit 27
Figure 3.3 The involvement of mushroom body output and dopaminergic neurons in aversive and appetitive reinforcement 29
Figure 3.4 The MBON-network layers and the anterior paired lateral neuron 35
Figure 3.5 Schematic representation of the olfactory processing layers of insects related to the mushroom bodies 39
Figure 3.6 The incentive circuit 46
Figure 3.7 Description of the experimental setup and the aversive olfactory conditioning paradigms 47
Figure 3.8 The susceptible and restrained microcircuits of the mushroom body 48
Figure 3.9 The reciprocal short-term memories microcircuit of the mushroom body 51
Figure 3.10 The long-term memory microcircuits of the mushroom body 53
Figure 3.11 The reciprocal long-term memories microcircuit of the mushroom body 54
Figure 3.12 The memory assimilation mechanism microcircuit of the mushroom body 56
Figure 3.13 The mammalian limbic system and the suggested parallels in the proposed incentive circuit 60
Figure 3.14 The synaptic weights and connections among the neurons of the incentive circuit 65
Figure 3.15 Description of the simulation process from the experiments . 68
Figure 4.1 T-maze elemental olfactory conditioning paradigm 75
Figure 4.2 The activity of the six mushroom body output neurons was translated into forces that drive a simulated fly towards or away from odour sources 82

Figure 4.3 Testing the performance of the incentive circuit on the T-maze 85
Figure 4.4 Testing the performance of the incentive circuit in neural activity intervention experiments . . . 87
Figure 4.5 The behaviour of the animals controlled by their neurons during the simulation where they can freely move . . . 89
Figure 4.6 The preference index of the (simulated) flies during the classic unpaired conditioning paradigm . . . 93
Figure 5.1 Basic structure of the compound eye . . . 103
Figure 5.2 Comparison of the visual rendering by panoramic camera-like processes and by the compound eye . . . 107
Figure 5.3 Overview of the visual place recognition task . . . 112
Figure 5.4 Pearson correlation between 50 pairs of views randomly selected from each of the 16 routes tested . . . 114
Figure 5.5 Summary of the visual place recognition contributions of this work . . . 115
Figure 5.6 Comparison of the familiarity predictions among the models and for the different tasks . . . 116
Figure 5.7 Comparison of the estimated steepness across the models . . . 118
Figure 5.8 Hypothetical model of how the familiarity predicted by the incentive circuit can control three visual place recognition behaviours for route following . . . 124
Figure 5.9 Integration of the central complex with the differential familiarity model and the visual place recognition behaviours . . . 126
Figure 6.1 Reinforcement learning architectures . . . 143
Figure 6.2 Discrete action space benchmarks based on grid space representation . . . 149
Figure 6.3 The average reward of the last episode for the different models and state representations . . . 152
Figure 6.4 The average reward of the last episode for the different models and state representations with added eligibility traces . . . 153
Figure 6.5 Factor graph representation of value-update by the long-term memory microcircuit and temporal difference . . . 155
Figure 6.6 Factor graph representation of value-update by the long-term memory microcircuit with tags from the eligibility traces . . . 156
Figure 6.7 Reinforcement learning architectures of the mushroom body 159
Figure A.1 The incentives wheel model . . . 178
Figure B.1 Comparison of the familiarity predictions among the models with additional lateral inhibition . . . 182

Figure C.1 Comparison of the familiarity predictions among the models with Zernike moments . . . 186
Figure D.1 Overview of the modelling pipeline . . . 191
Figure D.2 Sample output from the skylight dome model . . . 193
Figure D.3 Processing stages of light in the biological and artificial dorsal rim area . . . 195
Figure D.4 Overview of the compass model . . . 198
Figure D.5 The gating function that compensates for tilt . . . 200
Figure D.6 Using confidence of the estimate to compensate for time . . . 202
Figure D.7 Step-by-step processing of the compass model . . . 203
Figure D.8 The objective function and the accuracy of the compass . . . 205
Figure D.9 Dealing with time and light disturbance . . . 207
Figure D.10 Dealing with tilt for a variety of gating parameters . . . 208
Figure D.11 Optimal compass structural parameters . . . 210
Figure D.12 Behavioural simulation for the path integration task . . . 211
Figure D.13 Real and simulated response of compass neurons for artificial and natural polarised light . . . 214
Figure E.1 All the chemical levels and neural activities calculated based on the order of the conditioned and unconditioned stimuli . . . 224
Figure E.2 Anatomy of olfactory pathways in the fly brain . . . 225
Figure E.3 The responses from all the recorded neurons in the *Drosophila melanogaster* mushroom body during the aversive olfactory conditioning paradigm . . . 226
Figure E.4 The susceptible and restrained memory microcircuits of the mushroom body . . . 227
Figure E.5 The responses of the neurons of the incentive circuit using only the connections of the susceptible and restrained memory microcircuits . . . 228
Figure E.6 The synaptic weights of the neurons of the incentive circuit using only the connections of the susceptible and restrained memory microcircuits . . . 229
Figure E.7 The responses of the neurons of the incentive circuit using only the connections of the susceptible, restrained, and reciprocal short-term memories microcircuits . . . 230
Figure E.8 The synaptic weights of the neurons of the incentive circuit using only the connections of the susceptible, restrained, and reciprocal short-term memories microcircuits . . . 231

Figure E.9 The responses of the neurons of the incentive circuit using only the connections of the susceptible, restrained, reciprocal short-term, and long-term memories microcircuits 232
Figure E.10 The synaptic weights of the neurons of the incentive circuit using only the connections of the susceptible, restrained, reciprocal short-term, and long-term memories microcircuits . . 233
Figure E.11 The responses of the neurons of the incentive circuit using all the connections except of the memory assimilation mechanism microcircuit 234
Figure E.12 The synaptic weights of the neurons of the incentive circuit using all the connections except of the memory assimilation mechanism microcircuit 235
Figure E.13 The synaptic weights of the neurons of the incentive circuit using all its connections 236
Figure E.14 The reconstructed responses of the neurons of the incentive circuit using the reward prediction error plasticity rule 237
Figure E.15 The synaptic weights of the neurons of the incentive circuit using the reward prediction error plasticity rule 238
Figure E.16 The responses of the dopaminergic and mushroom body output neurons of the circuit when altering the pre-synaptic strengths of output neuron during the reversal condition 239
Figure E.17 The responses of the dopaminergic and mushroom body output neurons of the circuit when altering the dopaminergic modulation strengths during the reversal condition 240
Figure E.18 The responses of the dopaminergic and mushroom body output neurons of the circuit when altering the dopaminergic and output biases during the reversal condition 241
Figure E.19 The mean synaptic weights over the simulated flies that visited both odours 242
Figure E.20 Behavioural summary of simulated flies grouped by the areas that they visited 243
Figure E.21 Behavioural summary of simulated flies when controlled by the different output neuron types separately 244
Figure E.22 Paths of the flies when using the dopaminergic plasticity rule and during all 10 repeats of the experiment 245
Figure E.23 Behavioural summary of a subset of simulated flies that visited both odours at some time using the reward prediction error plasticity rule 246

Figure E.24 Paths of simulated flies when using the reward prediction error plasticity rule . . . 247
Figure E.25 The mean synaptic weights when using the reward prediction error plasticity rule simulated flies . . . 248
Figure E.26 The estimated normalised familiarity along the parallel displacements when using the incentive circuit . . . 249
Figure E.27 Detailed view of the polarised-light compass layout . . . 250
Figure E.28 Behavioural simulation for the path integration task—summarised results . . . 251
Figure E.29 Behavioural simulation for the path integration task—detailed results . . . 252
Figure E.30 Behavioural simulation for the path integration task—compensating for time . . . 253
Figure E.31 Real and simulated response of compass neurons—detailed results . . . 254

LIST OF TABLES

Table 3.1 List of intrinsic cell types in the mushroom body . . . 29
Table 3.2 List of extrinsic modulatory cell types in the mushroom body 31
Table 3.3 List of extrinsic output cell types in the mushroom body . . . 33
Table 3.4 Computational models of the mushroom bodies . . . 38
Table 4.1 The different types of olfactory conditioning in insects. . . . 74
Table 4.2 Preference indices of the individual T-maze experiments run for simulated flies . . . 84
Table 5.1 Summary of the steepness changes along the different routes 119
Table D.1 The cross-insect properties of our model. . . . 192
Table D.2 Mean absolute error before and after using the gating function. 208
Table D.3 Properties of dorsal rim ommatidia in different species. . . . 218
Table E.1 Connections among neurons in the fruit-fly mushroom body mapped to the connections of the incentive circuit . . . 221
Table E.2 Additional information on the mushroom body neurons used for the incentive circuit . . . 222
Table E.3 Spherical coordinates of the exact positions of the POL units on the compass dome . . . 223

LIST OF ACRONYMS

ACh	acetylcholine.
ADAC	action-dependent adaptive critic.
ADPF	activity dependent pre-synaptic facilitation.
AI	artificial intelligence.
AL	antennal lobe.
ANN	artificial neural network.
AOT	anterior optic tract.
AOTU	anterior optic tubercle.
APL	anterior paired lateral.
ASOT	anterior superior optic tract.
BP	back propagation.
CAS	continuous action space.
CR	conditioned response.
CRE	crepine.
CS	conditioned stimulus.
CX	central complex.
CZ	convergence zone.
DA	dopamine.
dAC	dorsal accessory calyx.
DAN	dopaminergic neuron.
DAS	discrete action space.
DPM	dorsal paired medial.
DPR	dopaminergic plasticity rule.
DRA	dorsal rim area.
dTrpA1	transient receptor potential A1.
EB	ellipsoid-body.
EPG	ellipsoid-body protocerebral-bridge gall.
FB	fan-shaped body.
GABA	gamma-aminobutyric acid.
Glu	glutamate.

HH Hodgkin-Huxley.

iACT inner antennocerebral tract.
IC incentive circuit.
IF integrate and fire.
IM infomax.
IOC inferior optic commissure.
IW incentive wheel.

KC Kenyon cell.

LA lamina.
LAL lateral accessory lobe.
LH lateral horn.
LI lateral inhibition.
LO lobula.
LOP lobula plate.
LSTM long short-term memory.
LTM long-term memory.

MAM memory assimilation mechanism.
MB mushroom body.
MBON mushroom body output neuron.
MC Monte Carlo.
MCH4 4-methylcyclohexanol.
MDP Markov decision process.
ME medulla.
ML machine learning.
MV mixed-valence.

NO nitric oxide.

OAN octopaminergic neuron.
OCT optical calycal tract.
OCT3 3-octanol.
OL optic lobe.

PAM protocerebral anterior medial.
PB protocerebral bridge.
PCA principle component analysis.
PCPN principle component projection neuron.

PER proboscis extension response.
PFC prefrontal cortex.
PFL2 protocerebral-bridge fan-shape-body lateral-accessory-lobe 2.
PFL3 protocerebral-bridge fan-shape-body lateral-accessory-lobe 3.
PFN protocerebral-bridge fan-shape-body noduli.
PI preference index.
PM perfect memory.
PN projection neuron.
POC posterior optic commissure.
PPL1 protocerebral posterior lateral.

ReLU rectified linear unit.
RL reinforcement learning.
RLM reciprocal long-term memory.
RM restrained memory.
RPE reward prediction error.
RSE root square error.
RSM reciprocal short-term memory.
RSS residual sum of squares.

SD standard deviation.
shi shibire.
SIP superior intermediate protocerebrum.
SLP superior lateral protocebrum.
SM susceptible memory.
SMP superior medial protocerebrum.
SNN spiking neural network.
SOC serpentine optic commissure.
SRDP spike-rate-dependent plasticity.
STDP spike-timing-dependent plasticity.
STM short-term memory.
SVM support vector machines.

TD temporal difference.

UR unconditioned response.
US unconditioned stimulus.
UV ultra violet.

vAC ventral accessory calyx.

VPM4 octopamine-ventral paired 4.
vPN visual projection neuron.
VPR visual place recognition.
VSλ valence-specific.
VUMa2 octopamine-ventral unpaired a2.

WN Willshaw network.
WTA winner-takes-all.

ZM Zernike moment.

1

INTRODUCTION

"Look deep into nature, and then you will understand everything better."

Albert Einstein

Animals can quickly adapt to the changes in their environment and learn how to solve problems that concern their survival without explicitly being informed about their failures. This is a desirable feature for many *artificial intelligence* (AI) systems that remains a challenge, as they usually need a point of reference (like a 'ground truth') that informs them about the correctness of their predictions or actions. *Reinforcement learning* (RL) is a branch of *machine learning* (ML) that uses the experience of the system to update its parameters through trial and error (with the goal of approximating the aforementioned adaptability of animals). This field was inspired by nature, specifically from the way that animals learn how to survive. Initially, the fields of biology and RL were in close collaboration, providing insights both on how animals adapt and achieve complicated tasks, and on how to use this knowledge to improve the performance of the RL systems (Sutton and Barto, 2018; Niv, 2009), resulting in contributions to both fields. This collaboration was important, as it resulted in systems that showed apparent reasoning (clear steps of processing that lead to action; arguably a form of AI), but also helped the biological community to better understand the reasoning in animals, which in-turn provided insights for better artificial systems, and so on. However, the rapid evolution of RL caused a diversion from the original idea. Recent RL systems use neural structures, which are closer to the structure of the brain than older normative approaches, and they are significantly more successful than the previous generations. However, their parameters are optimised in a way that diverges from the learning mechanisms of animals, which results in systems with no apparent reasoning (usually characterised as black boxes). These systems provided few causal insights into how intelligent behaviour is formed or how animals learn. An increasing gap was developed between the two fields, leading to hard-to-interpret RL systems, and limiting understanding of the learning capabilities of animals (and their effect on behaviour). With the aim of bridging this gap, this thesis takes a neuroethological approach; that is, it seeks to explain RL concepts by using computational modelling, and neuroanatomical and behavioural data from insects.

Neuroethology is the study of the behaviour of animals (also known as ethology) as implemented by their neurons. On the behavioural side, it has been argued that intelligence can only be sufficiently demonstrated by closed-loop systems where animals (or agents) interact with their environment (Achacoso and Yamamoto, 1990). On the neural side, despite being difficult to interpret, recent RL approaches use neural representations to produce behaviour; neural processing also better approximates processing in the brain, allowing for a direct comparison between artificial and natural systems. Recent technological advances allow for genetic manipulations that can activate neurons by shining a light on them or cause neurons to emit light when they are active. Such a genetic toolkit allows researchers to procure accurate recordings from freely moving animals (although in most available data the animals are actually fixed); combined, these advances allow researchers to establish neural activation chains from input to behaviour. Such manipulations are currently only practical in limited species of insects (and some fish).

Insects (despite their tiny brains) demonstrate quite complex behaviours; this combined with the powerful genetic toolbox makes insects valuable models for neuroethological work. A large volume of data has been collected from the brain of fruit flies (of the species *Drosophila melanogaster*), providing an excellent basis to explore how RL is implemented in their brain. In terms of their behaviour, there is no obvious purpose for their actions other than feeding and mating. On the other hand, central place foraging insects (such as desert ants and honeybees) have a more obvious goal, which is to find food and bring it back home. This makes them excellent navigators, popular for ethological studies, and interesting study cases for robotics. However, much less neural data are available from the brains of these insects. Thus, a common approach is to infer the function of their brain by using data collected from fruit flies and collectively contribute to studies in neuroethology. The size of the insect brain is on the order of a hundred thousand neurons compared to tens of billions of neurons in the human brain. The structures hypothesised to implement RL in the insect brain are called the *mushroom bodies* (MBs), and they contain less than three thousand neurons in total (small compared to the neural structures used in modern RL approaches). For these reasons, the insect brain seems to be an excellent candidate to study the neural function of RL and try to reproduce the behaviour of insects.

This is not the first time that the MBs have been explored in order to understand how RL is implemented in the insect brain. Several computational models have explored these structures in different ways; for example, proposing different plasticity rules or plausible neural architectures concerning different behaviours and insect species. This thesis has considered these solutions and proposes the novel *dopaminergic plasticity rule* (DPR) that increases learning flexibility, and the novel *incentive*

circuit (IC) as an extension for existing neural architectures. The justification for both the DPR and the IC comes from neuroanatomical and neural activity data taken from the fruit-fly brain. The combination of the rule and circuit was able to reproduce the neural activity of targeted neurons, providing important insights into how different types of memories are formed in the insect brain, and how these memories can affect the behaviour of the animal. The DPR and IC were also applied to problems such as the *visual place recognition* (VPR) and standard RL benchmarks for temporally sparse reinforcements, suggesting mechanisms for collaboration among the different regions in the insect brain, and providing insights into the role of each region in a more standard RL framework.

1.1 INTRODUCTION TO COMPUTATIONAL NEUROETHOLOGY OF INSECTS

Computational neuroethology is a conceptual framework that provides tools for understanding the behaviour of animals through computer simulations and computational modelling of their nervous system (Chiel and Beer, 2009). The resulting models may take the properties of the animal into accounts, such as the details of its neural circuit, sensors, actuators, and behaviour. These models can thus be used to test specific hypotheses for the behaviour or neural processing of the animal; they can also be implemented in artificial agents (or robots), imitating specific aspects of the behaviour of the animals. The approach used in this thesis follows the fundamental principles of computational neuroethology. One of those is that the behaviour of the system depends on its active perception, sensory pre-processing, and actions, which are eventually fused in its neural circuits (Chiel and Beer, 2009). In addition, as animals are situated in an environment, their brain, body, and the environment itself, all contribute to their adaptive behaviour.

The structure of the insect brain is very well documented. A large volume of calcium imaging and electrophysiological data have been collected (especially from fruit flies), providing an excellent ground for computational neuroethology. Insects also exhibit complex (and interesting) behaviours, ranging from olfactory conditioning (attraction or aversion to odours) to solving complicated navigation tasks (requiring the integration of input from multiple sensors and the collaboration among many parts of their brain). Although their level of intelligence is arguable (because highly intelligent animals should be able to consciously reason and make decisions; Achacoso and Yamamoto, 1990), the behaviour of insects seems to be well-optimised to solve tasks that are beyond the abilities of artificial agents. Therefore, understanding how insect brains work could aid in constructing more robust and original AI.

The neuroethological framework also requires neural (or neuronal) models that describe observed behaviours. Neural models mimic neural information processing, which can allow direct comparison with recordings from the brain. Neural models also provide different levels of abstraction (Herz et al., 2006); these range from low-level modelling (for example, detailed modelling of the biochemical and physical properties of the neurons) to high-level modelling (for example, more abstracted algorithmic models). Each of these levels might be more useful for explaining different types of data, like low-level modelling for electrophysiological data, or high-level modelling for behavioural-only data. This thesis uses simple neural models (the output is the weighted sum of the input, optionally transformed by a non-linear function), which lie in the middle of this range.

1.2 SCOPE

The main questions addressed in this thesis concerned the function of the MBs and their parallels to RL frameworks. The first objective was to determine the role of reinforcement, inputs, and outputs of the MBs in the context of RL. The second objective was to examine how *dopamine* (DA) affects the connections among neurons in the MB, as this is usually assumed to be parallel to the reinforcement. The anatomical description of the MBs suggested feedback connections from *mushroom body output neurons* (MBONs) to *dopaminergic neurons* (DANs), which are less common in RL architectures as actions usually affect the reinforcement through the environment and not through direct connections. An exception is the *action-dependent adaptive critic* (ADAC) framework, which calculates the *reward prediction error* (RPE) or the *temporal difference* (TD) error that has not been evident in the DAN or MBON activity of fruit flies. This led to the third objective, which was to determine the role of the MBON $\rightarrow$ DAN connections, and how are they related to the dopaminergic function. The fourth objective was to test the resulting MB model in a wide range of olfactory conditioning paradigms in an attempt to explain the data collected from fruit flies and provide useful predictions for the field of neuroethology. This would also verify the correctness of the model before it was used in more complicated tasks, taking the place of an RL architecture. The final objective was to use this model to solve other insect- and robotic-related tasks (for example, VPR), or popular RL problems (for example, delayed reinforcements) that also concern the field of neuroethology.

The main contributions of this thesis include the development of a biologically justified model for the MB, and its evaluation in a variety of tasks, providing insights for its function with respect to RL and neuroethology. Several more detailed contributions are listed below (grouped by chapter).

1. **The dopaminergic plasticity rule.** A rule to modify the connections among input and output neurons of the MB was developed based on calcium imaging data from the fruit-fly brain. This allowed for several learning phenomena to take place simultaneously, like depression, potentiation, recovery, and saturation of memories. When a higher temporal resolution was used in the simulations, this plasticity rule could also provide insights into the underlying mechanism that causes the backwards conditioning effect (a positive association to stimuli that occurred after punishment ceased).

2. **The incentive circuit.** An initial analysis performed on calcium imaging data collected from MBON and DAN recordings in the fruit-fly MBs (data collected by Li Yan McCurdy from Yale School of Medicine and it was not part of this thesis) helped in the understanding of the structure and function of their circuitry. The analysis of the data (along with the anatomy and function of targeted neurons) led to the exploration of several microcircuits of MBON and DAN interconnections. Different microcircuits facilitated short- and long-term memory and even allowed memories to be transferred from one to the other. The IC emerged by weaving these microcircuits together, which allowed for more complicated memory dynamics that merged the properties from all the explored microcircuits. Each neuron and connection of the IC has an exact parallel in the fruit-fly brain. The individual neural responses produced by the model had similar trends to their real counterparts, as recorded by Li Yan.

3. **Explaining the olfactory conditioning paradigms.** By using the behavioural readout from the MBONs of the IC, different olfactory conditioning paradigms were tested and compared to data reported from fruit flies. These paradigms included classic elemental olfactory conditioning, generic interventions of the involved MBONs and DANs during elemental conditioning, as well as freely moving simulated flies (which are closer to the principles of computational neuroethology). The behaviour produced by the IC was able to explain most of the relevant fruit-fly data and provide insights into the behaviour.

4. **From visual rendering to visual processing, to visual familiarity.** Inspired by the navigation mechanisms of desert ants, the IC was also tested on the VPR task. For this reason, a rendering technique was developed in order to visually capture the world from an insect-like perspective. Two layers of pre-processing were proposed, *principle component analysis* (PCA) whitening (which decorrelated the visual input), and combinatorial sparse encoding (which created a sparse code of input representations that could be better assimilated by the MBs). The IC solved the VPR task with similar performance to simpler models but revealed a unique incremental MBON response when the familiarity of consequent views was consistent.

5. **The marriage of reinforcement learning and insect neuroethology.** The IC could not solve tasks with temporally sparse or delayed reinforcements, suggesting that the MB should be parallel to an associative RL architecture. Alternatively, the function and structure of the MB model suggested that this (alone) might not be parallel to an RL architecture, but it might take the role of an *adaptive critic* in an ADAC architecture. It was finally suggested that the *central complex* (CX) might have the role of the actor, which uses the MBON motivational output to select among the different actions.

2

SYNAPTIC PLASTICITY

"The most useful piece of learning for the uses of life is to unlearn what is untrue."

Antisthenes

Neurons are composed of *dendrites*, *somas* (their bodies), and *axons*; and they form *synapses* wherever they are in contact with other neurons. Synapses are important for establishing communication between neurons, and there are two ways to achieve that: via *neurotransmitters* or *neuromodulators*. Neurotransmitters are messengers that carry information between the neurons through the synapses, while neuromodulators are messengers released from neurons that affect the synaptic strength of the connections. *Synaptic plasticity* is the ability of the brain to form and reorganise such synaptic connections. It is usually assumed to be responsible for memory formation, or changes in the behaviour of the animal. There are alternative ways to form memories or behavioural change mechanisms; for example, the working memory is formed by recurrent (self-) connections of neurons and does not require synaptic plasticity. However, building associations between neural responses is an exclusive feature of synaptic plasticity, and therefore, this chapter works towards understanding the underlying mechanism that builds associations in the brain through neuromodulators.

Dopamine (DA) is a widely accepted neuromodulator, which is occasionally released by the *dopaminergic neurons* (DANs) in different areas of the brain. Although significant work has been done trying to understand the function of plasticity caused by DA in the brain, it is still a great mystery that keeps this topic in the spotlight of research in neuroscience, information theory, and *artificial intelligence* (AI). There is increasing progress in understanding the DA function in the insect brain, enabling more accurate theoretical models. This rapid increase in understanding of the insect brain is mainly based on the fact that they are easier to manipulate (for example, by optogenetics), and easier to access than the mammalian brain. The *mushroom body* (MB) neuropils are known to be centres for associative *reinforcement learning* (RL) in the insect brain, whose synapses are usually assumed to be plastic only in-between a densely connected layer of input and output neurons (however, note that Chapter 3 suggests a more complicated structure and circuit dynamics for the MBs). The input neurons encode multimodal sensory input (for example, olfactory and visual),

while the output neurons encode higher-level behavioural responses (for example, feed or sleep). These apparently well-defined structures are convenient from a computational perspective, as they provide a solid background to study the DA function.

There are quite a few computational models that attempted to depict the function of DA in these structures, emulating plasticity in computational circuits. In AI, the plasticity in the brain is usually formalised with mathematical functions called the *plasticity* (or learning) *rules*. These try to imitate the change in the association between the responses of input and output *units*. Note that units are equivalent to neurons in *artificial neural networks* (ANNs). The synaptic connections between the layers of units are usually represented by a two-dimensional matrix (called the weight matrix), which stores the synaptic strengths from all the input to all the output units. DA is assumed to be the reinforcement signal that affects the synaptic connections stored in the weight matrix. However, in the insect brain, it is evident that no pre-synaptic or post-synaptic activity is needed for the DA to affect the synaptic weights (Hige et al., 2015; Dylla et al., 2017; Berry, Phan, and Davis, 2018). This is something that none of the existing plasticity rules could account for; thus, there is a need for a new plasticity rule that replicates better the function of DA in the brain.

In this chapter, a novel *dopaminergic plasticity rule* (DPR) is proposed, which does not need pre-synaptic or post-synaptic activity to introduce plasticity. The effectiveness of the DPR was tested by modulating the synaptic connections of a simplified MB-like structure during olfactory (backwards or relief) conditioning experiments, based on the work of Handler et al. (2019) with *Drosophila melanogaster* fruit flies. This supported (with calcium-imaging data) the effectiveness of DPR in representing synaptic plasticity in the MBs of the insect brain (correlation of learning effects with experimental data: Pearson correlation coefficient $r = 0.98$, $p < 3.9 \cdot 10^{-4}$).

2.1 RELATED WORK

A common assumption when describing synaptic plasticity is that at a discrete time-step (t) the activity of the pre-synaptic neuron ($x_i \in \{0, 1\}$; of a set of $i \in \{0, ..., X-1\}$ pre-synaptic neurons) affects the activity of the post-synaptic neuron ($a_j \in \{0, 1\}$; of a set of $j \in \{0, ..., A-1\}$ post-synaptic neurons), and this depends on the synaptic weight ($w_{ij} \in \mathbb{R}_+$) and the reinforcement ($r \in \mathbb{R}$). The plasticity rules change the synaptic weights with respect to a goal (for example, to create an association between the pre- and post-synaptic activity) and this depends on the change rate ($\eta \in \mathbb{R}_+$)

and on the salience parameter ($\beta \in \mathbb{R}_+$). Following J. Young et al. (2011), here these parameters were assumed to have the values of $\eta = 0.5$ and,

$$\beta = \begin{cases} 1, & \text{if no salience adjustment,} \\ 1/\sum_k x_k(t), & \text{otherwise.} \end{cases} \tag{2.1}$$

In what follows, there is an overview of the plasticity rules proposed previously for explaining the DA function in the insect brain.

2.1.1 *Reward prediction error rule*

The most popular plasticity rule used to update the weight-matrix is the one proposed by Rescorla and Wagner (1972), which is known as the *reward prediction error* (RPE) or *Rescorla-Wagner* plasticity rule, and it is given by,

$$\Delta w_{ij} = \eta\beta\, x_i(t) \left[r(t) - \sum_{k=0}^{X-1} x_k(t)\, w_{kj}(t) \right] \tag{2.2}$$

$$= \eta\beta\, x_i(t) \Big[r(t) - a_j(t) \Big]. \tag{2.3}$$

In practice, this plasticity rule tries to make the active output unit predict the value of the reinforcement. This is done by changing the synaptic connections between the elements that exist in the input, which is, $x_i(t) = 1$, and the target output. Whether the synaptic connection will become stronger or weaker depends on the sign and magnitude of the error between the predicted and observed reinforcement.

Note that when the input predicts the observed reinforcement—the corresponding element in the output vector has the same value as the observed reinforcement, $a_j(t) = r_j(t)$—there is no change in the weight matrix. In the special case, where two different input vectors predict the same observed reinforcement, and the weight matrix is changed so that one of them exactly predicts this reinforcement, then a linear combination of the two vectors would predict the same reinforcement. As the combined vector can already predict the observed reinforcement, no association between the second input vector and the reinforcement will be learnt. This phenomenon (caused by the RPE) is called 'blocking' (Kamin, 1967).

It is worth noticing that the *temporal difference* (TD) plasticity rule (Sutton and Barto, 1990), which is discussed in Chapter 6, was inspired by a combination of the RPE (expressing classical conditioning) and the learning by 'surprise' theory (Kamin, 1967). This rule can explain puzzling behavioural phenomena (such as blocking, over-

shadowing, and conditioned inhibition) that were inherited by the RPE rule, as well as higher-order conditioning and temporal sensitivity (Niv, 2009).

2.1.2 *Basic rule*

A variation of the RPE rule, named as the *basic* plasticity rule, was used by Pearce (1994) and J. Young et al. (2011), and is given by,

$$\Delta w_{ij} = \eta\beta\, x_i(t)\,[r_j(t) - w_{ij}(t)]. \tag{2.4}$$

Here, the reinforcement depends on the post-synaptic neuron—if $a_j(t) = 1$ then $r_j(t) = r(t)$, otherwise $r_j(t) = 0$. The value of each connection in the weight matrix is independent of each other, as the target value for each of the elements is the reinforcement signal. The idea behind this rule was to lessen the blocking effect caused by the RPE, as such an effect was not clearly observed in insects (and not always in vertebrates either). Thus, they proposed a plasticity rule that avoided blocking and the reinforcement signal could change each synaptic weight irrespective of each other (which was also suggested by direct evidence from insects; Hige et al., 2015; Dylla et al., 2017; Berry, Phan, and Davis, 2018; Schleyer, Fendt, et al., 2018; Schleyer, Weiglein, et al., 2020). The reasoning behind the multidimensional reinforcement signal was that plasticity would occur only for the active post-synaptic neuron, which enforces the association of potentially different types of reinforcements to different output neurons. However, the original version had only one output neuron and one type of reinforcement (J. Young et al., 2011).

2.1.3 *Template rule*

In the above plasticity rules, the input vectors were associated with specific output vectors based on the observed reinforcement. These associations are subsequently 'forgotten' (the vectors are dissociated) when the same input and output pair is provided without the reinforcement. However, if many inputs are associated with the same output, all of them will create new associations without causing any dissociation between the previously learnt inputs and the current output.

A. Balkenius, Kelber, and C. Balkenius (2006) suggested an extension to the basic rule, named the *template*, which introduces forgetting when the reinforcement is present without the previously learnt inputs. Their idea was inspired by multimodal learning in hawkmoths. They suggest that the animal tries to form a template for the stimulus associated with reinforcement by responding to a reinforced input with an

increase in the value of the elements presented in the input and a decrease in the values of all the other elements not present in this input. Formally, this plasticity rule could be written as,

$$\Delta w_{ij} = \begin{cases} \eta\beta\, x_i(t)\,[r_j(t) - w_{ij}(t)], & \text{if } r_j(t) = 1 \text{ and } x_i(t) = 1, \\ -0.2, & \text{if } r_j(t) = 1,\ x_i(t) = 0, \text{ and } w_{ij}(t) > 0, \\ 0, & \text{otherwise,} \end{cases} \tag{2.5}$$

which was formalised by J. Young et al. (2011), and it was able to successfully model the behaviour of hawkmoths as tested by A. Balkenius, Kelber, and C. Balkenius (2006). In their variant, J. Young et al. (2011) replaced the RPE with the basic rule. Also (to avoid any confusion with respect to the change rate and salience) it is worth mentioning that J. Young et al. (2011) used $\eta = 0.5$ and calculated β using Eq. (2.1). Similarly to the basic rule, in order to support plasticity for targeted post-synaptic neurons, multidimensional reinforcement (r_j) was also used here. This plasticity rule results in no learning when the reinforcement is missing, which is different to RPE and basic plasticity rules.

2.1.4 *Hebbian rule*

Another popular plasticity rule was proposed by Hebb (1949). In this rule, the coincidence of an input and an output element strengthens their respective synaptic weight. A variant of this rule uses the reinforcement signal to enable or disable the plasticity (Pennartz, 1997), and it is given by,

$$\Delta w_{ij} = \eta\beta\,[r(t)\, x_i(t)\, a_j(t) - w_{ij}(t)]. \tag{2.6}$$

Note that, this is not the classic Hebbian rule (Hebb, 1949), but a variant that involves the weakening of the synaptic strength when the input, output and reinforcement are contradicted. As opposed to the plasticity rules described before, this variant does not calculate the error between the reinforcement and output (or the reinforcement and the synaptic weight) in order to direct the change of the value. Instead, in the presence of the reinforcement, it strengthens the connection when $r_j(t)\, x_i(t)\, a_j(t) > w_{ij}(t)$, and otherwise weakens it. The synaptic weights are randomly initialised, and the way this is done is crucial for the performance of this rule.

2.1.5 *Neural-modulation rule*

Proposed by J. Young et al. (2011), the *neural-modulation* plasticity rule is an abstraction of the reward model proposed by Izhikevich (2007), which tried to capture some of the current assumptions about the dopaminergic system in the insect brain (Gerber, Tanimoto, and Heisenberg, 2004; Heisenberg, 2003). This rule is given by,

$$\Delta w_{ij} = \eta\beta\, r_j(t)\, x_i(t), \tag{2.7}$$

and it looks similar to the basic rule, but it has some substantial differences. The rule is based on the assumption that the value of an element is increased when reinforcement is present. Therefore, there is no upper limit for the values and no target value, which is in accordance with the Hebbian rule (and different from the basic rule). Unless the reinforcement changes sign, there is no obvious way that the synaptic weight related to an element is decreased. Another difference to the basic rule is that the changes in the synaptic weights do not depend on the difference of the reinforcement from the synaptic weight. Note that, if a global reinforcement signal is in place, this rule could be useful only for a system with one output unit. However, in the context of an MB architecture, it is assumed that there are distinct units that deliver different reinforcements, and each one of them targets a different group of output units. Thus, similar to the basic and template rules, direct information from the output units is unnecessary for this plasticity rule, but its reinforcement is multidimensional (use of r_j instead of r).

2.1.6 *Stochastic Hebbian rule*

In all the plasticity rules above, the changes in the synaptic weights are deterministic. Huerta, Nowotny, et al. (2004) used a stochastic plasticity rule based on the Hebbian principle in order to discriminate odours, and found its performance competitive compared to existing plasticity rules. For this rule, there are two predefined probabilities: the probability of increasing weight, p_+, and the probability of decreasing weight, p_-. Given these probabilities, the plasticity rule is the following,

$$\Delta w_{ij} = \eta\beta \begin{cases} 1 - w_{ij}(t), & \text{with } p_+, \text{ if } x_i = 1,\ a_j = 1, \text{ and } r(t) = 1, \\ -w_{ij}(t), & \text{with } p_-, \text{ if } x_i = 0,\ a_j = 1, \text{ and } r(t) = 1, \\ 0, & \text{otherwise.} \end{cases} \tag{2.8}$$

Note that the reinforcement here was binary ($r \in \{0, 1\}$) and worked as a simple switch for learning.

This rule also inspired Huerta and Nowotny (2009) to test its performance in the hand-written digit classification task, demonstrating that an MB-inspired model was able to compete with state-of-the-art methods. The success of this plasticity rule also led Bazhenov, Huerta, and B. H. Smith (2013) to test it on honeybee decision-making tasks.

2.1.7 *Anti-Hebbian rule*

A relatively different approach was developed by D. Smith, Wessnitzer, and Webb (2008), who tried to model the learning mechanism of the *D. melanogaster* MBs using a *spiking neural network* (SNN) that captured the neural dynamics of the associative learning mechanism in *Aplysia* sea slugs. A simple *integrate and fire* (IF) unit was used along with an anti-Hebbian plasticity rule, which combined the *activity dependent presynaptic facilitation* (ADPF) and Hebbian post-synaptic plasticity. The ADPF pre- and Hebbian post-synaptic rules were inspired by the Kandelian and Hebbian synapses in the MBs respectively (Dubnau and Tully, 2001).

In contrast to the Hebbian rule, the synaptic weights in the anti-Hebbian rule are all initially the same—$w_{ij}(0) = 1$. Then, the weights are reduced on the coincidence of input and output activity with reinforcement, and this can be expressed formally as,

$$\Delta w_{ij} = -\eta\beta \begin{cases} w_{ij}(t), & \text{if } x_i(t) = 1, a_j(t) = 1, \text{ and } r(t) = 1, \\ 0, & \text{otherwise.} \end{cases} \tag{2.9}$$

However, note that with this rule the synaptic weights can only decrease, so everything that is learnt cannot be unlearnt.

2.2 RESULTS

In the insect brain, the biological plasticity mechanism for which there is the most experimental support is that coincidence of the reinforcement and active input units depress the synaptic strength between these units and the output units, regardless of the activity of the output units. However, most of the plasticity rules mentioned (with the exception of the basic, template, and neural-modulation rules) use variations of the Hebbian or RPE rules, which require the activity of the output units.

More recent approaches try to model the activity of DA and output neurons in the insect brain, and they use plasticity rules (C. Zhao et al., 2021) or circuit structures (Springer and Nawrot, 2021; Bennett, Philippides, and Nowotny, 2021; Eschbach et al., 2020) that implement the RPE. For the MB, this plasticity rule is interpreted as the output being the prediction of the reinforcement, so their difference drives the synaptic plasticity (see Section 2.1.1). However, details of neuronal dynamics in fruit flies (Hige et al., 2015; Dylla et al., 2017; Berry, Phan, and Davis, 2018; also in larva: Schleyer, Fendt, et al., 2018; Schleyer, Weiglein, et al., 2020) suggested that the activity of the input units is not required for changes in the synaptic weights, which is in contrast to Hebbian, RPE, basic, and neural-modulation plasticity rules. This suggests that only the template plasticity rule could capture the plasticity dynamics in the MBs, and it highlights the importance of investigating new plasticity rules that are a closer approximation to the function of DA.

2.2.1 *The dopaminergic plasticity rule*

The *dopaminergic plasticity rule* (DPR) is a novel plasticity rule that reflects a recent understanding of the role of DA in the depression and potentiation of synaptic weights in the MB. The DPR is implemented to update the weight-matrix proportionally to the DA level, for both active and inactive input units, with respect to their synaptic weights and default (rest) value. The rule is based on recent findings regarding the role of DA (and co-transmitters) in altering synaptic efficacy in the fruit-fly MBs (see Section 2.2.2). Instead of calculating the error between the reinforcement and its prediction, or using the coincidence of the activity of input and output units, DPR uses the reinforcement to provide a plasticity palette that can be used flexibly by circuit architectures to implement different plasticity functions. Therefore, using this rule, a specific circuit could result in the prediction of reinforcements (see Section 4.3.1), the formation of *short-term memories* (STMs), or *long-term memories* (LTMs), supporting a more flexible range of responses to input-reinforcement contingencies.

The DPR is written formally as,

$$\Delta w_{ij} = \delta_j(t)\,[x_i(t) + w_{ij}(t) - w_{\text{rest}}], \tag{2.10}$$

where $w_{\text{rest}} = 1$ is a constant resting weight. The rule alters the values in the weight-matrix on each time step depending on the *dopaminergic factor*, $\delta_j(t)$, which is determined by the responses of the DANs. For now, this factor is assumed to be defined as $\delta_j(t) = \eta\beta\, r_j(t)$, where η is the change rate and β is the salience parameter of Eq. (2.1); this becomes more complicated in the next section (Section 2.2.2). The dopaminergic

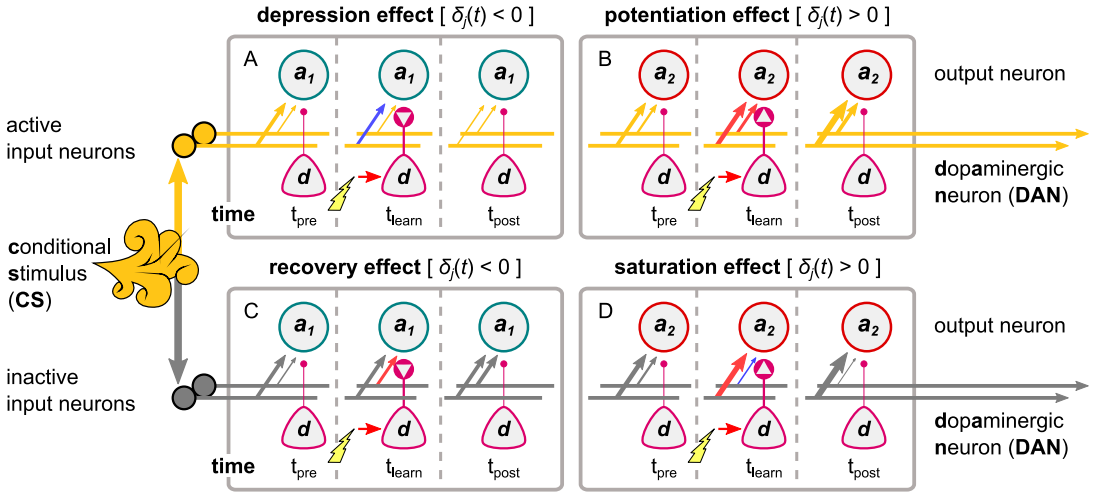

Figure 2.1: The different effects of the *dopaminergic plasticity rule* (DPR), depending on the activity of the input units (yellow indicates active and grey inactive *Kenyon cells*, KC) and the sign of the dopaminergic factor (white arrowheads in dots). The DPR can cause 4 different effects that allow different types of memories to be formed for each target synapse. In each box, time-step $t = t_{\text{pre}}$ shows the initial synaptic weight (line thickness); in time-step, $t = t_{\text{learn}}$, electric shock activates the *dopaminergic neuron* (DAN) causing modulation of the synaptic weight (red = increase, blue = decrease); time-step $t = t_{\text{post}}$ shows the synaptic weights after the shock delivery. (A) Example of the *depression effect*; the synaptic weight decreases when $\delta_j(t) < 0$ and the input is active. (B) Example of the *potentiation effect*; the synaptic weight increases when $\delta_j(t) > 0$ and the input is active. (C) Example of the *recovery effect*; the synaptic weight increases when $\delta_j(t) < 0$ and the input is inactive. (D) Example of the *saturation effect*; when $\delta_j(t) > 0$ and the input is inactive, the synaptic weight increases further when $w_{ij}(t) > w_{\text{rest}}$ or decreases further when $w_{ij}(t) < w_{\text{rest}}$. Adapted and modified from Gkanias, McCurdy, et al. (2022) under CC BY 4.0.

factor can be positive $[\delta_j(t) > 0]$ or negative $[\delta_j(t) < 0]$, which is motivated by recent observations of the differential roles in synaptic plasticity of DopR1 and DopR2 receptors (Handler et al., 2019), as detailed in Section 2.2.2. When combined with two possible states of the input units (active or inactive), this results in four different plasticity effects: *depression*, *potentiation*, *recovery* and *saturation*.

These effects can be inferred directly from Eq. (2.10). If the dopaminergic factor is zero (absence of reinforcement), no learning occurs. If the dopaminergic factor is negative and the input unit is active (positive), the respective synaptic weight is decreased (depression effect of the plasticity rule, see Fig. 2.1A). The recovery effect takes place when the dopaminergic factor is negative and the input unit is inactive $[x_i(t) = 0]$, in which case the synaptic weights tend to reset to the resting weight (see Fig. 2.1C). The potentiation effect is when the dopaminergic factor is positive and the input unit is active, and it causes an increase in the synaptic weights (see Fig. 2.1B). In contrast to the depression effect, as the synaptic weight becomes stronger, it further enhances this effect. If the input unit is inactive, and the dopaminergic factor is

positive then the saturation effect takes control. In this effect, if the current synaptic weight is higher than its resting weight, the synaptic weight continues increasing, but if it is lower then it continues decreasing (see Fig. 2.1D). This effect enhances diversity in the responses of the output units to the different past and current inputs, which is essential for memory consolidation (which is the continued strengthening of memory) and for the formation of LTMs (which are defined by slower acquisition and resistance to further change).

The above plasticity effects can be grouped into active [which are the depression and potentiation; input unit is active, $x_i(t) > 0$] and passive [which are the recovery and saturation; input unit is inactive, $x_i(t) = 0$]. Through influence from multiple DANs, all the plasticity effects can work together in changing the values of the weight-matrix, leading to more complicated effects like the formation of STMs (for example, combining any active and the recovery effects) or LTMs (for example, combining any active and the saturation effects). However, by adding feedback connections from the output units to the DANs, a wide range of circuit properties can be implemented. In Section 3.3.1, a set of microcircuits that have been found in the fruit-fly MBs are introduced, and the DPR is used to explain their function. That chapter also demonstrates how they could interlock and interact in one *incentive circuit* (IC), and how this circuit could control the motivation (and hence the behaviour) of the animal.

2.2.2 *Derivation of the dopaminergic plasticity rule*

The DPR was inspired by electrophysiological studies in the fruit-fly brain, and more specifically in the MBs. For clarity, a brief description of the MBs is provided in this section, while Section 3.1.2 provides a more in-depth description of it. A key factor is that the *Kenyon cells* (KCs) are the input units to the MBs, $\mathbf{x}(t)$, and they are connected to the *mushroom body output neurons* (MBONs), $\mathbf{a}(t)$, with excitatory (positive) connections. The MBON activity indirectly affects the behaviour of the animal. The DANs modulate the KC $\rightarrow$ MBON synaptic weights, $\mathbf{W}(t)$, and they receive reinforcement input, $\mathbf{r}(t)$, as well as feedback from the MBONs.

Handler et al. (2019) suggested that ER$-$Ca^{2+} and cAMP play a decisive role in the dynamics of forward and backward conditioning. *Backward conditioning* is when the reinforcement is delivered just before the input stimuli, and it is based on the time-dependency between the two stimuli. More specifically, they suggested that the synaptic plasticity of the KC $\rightarrow$ MBON connections, Δw_{ij}, is proportional to the combined ER$-$Ca^{2+} and cAMP levels, which can be written formally as,

$$\Delta w_{ij}(t) \propto -(\text{ER}-\text{Ca}^{2+})_{ij}(t) - (\text{cAMP})_{ij}(t). \tag{2.11}$$

A reasonable assumption is that ER$-$Ca^{2+} and cAMP levels are determined by information available in the local area of the target KC axon (pre-synaptic terminal). These are, the DA level emitted by the DANs to the KC synapses of the respective (j^{th}) MBON, $D_j(t) \geq 0$; the activity of the (i^{th}) pre-synaptic KC, $x_i(t) \geq 0$; the respective KC $\rightarrow$ MBON synaptic weight, $w_{ij}(t) \geq 0$ (assumed always positive, exciting the MBON); and the resting synaptic weight, $w_{\text{rest}} \in \mathbb{R}$, which is a constant parameter of the synapse. By shuffling the above quantities in equations, trying to reproduce a signal proportional to the ER$-$Ca^{2+} and cAMP levels, mathematical formulations of the latter levels have been postulated as functions of the available information as,

$$(\text{ER}-\text{Ca}^{2+})_{ij} \propto -D_j^{\triangle}(t)\,[x_i(t) - w_{\text{rest}}] - [D_j^{\triangle}(t) - D_j^{\triangledown}(t)]\,w_{ij}(t) \tag{2.12}$$

$$(\text{cAMP})_{ij} \propto D_j^{\triangledown}(t)\,[x_i(t) - w_{\text{rest}}], \tag{2.13}$$

where $D_j^{\triangledown}(t)$ and $D_j^{\triangle}(t)$ are the depression and potentiation components of the DA respectively [assumed to correspond to DopR1 and DopR2 receptors, Handler et al., 2019, or potentially to involve co-transmitters released by the DAN such as *nitric oxide* (NO), Aso, Ray, et al., 2019]. Two types of DAN terminals are assumed: the depressing and potentiating terminals. In depressing terminals (arrow down), $D_j^{\triangledown}(t)$ makes a higher peak in its activity followed by a faster diffusion than $D_j^{\triangle}(t)$, which seems to be the key for the backward conditioning. The opposite happens in potentiating DAN terminals.

To validate that the DPR shows similar plasticity dynamics to the ones observed in Handler et al. (2019), its plasticity effects were tested during forward and backward conditioning. In this experiment, $0.5\,\text{sec}$ of *conditioned stimulus* (CS) pulse (for example, odour) was provided in different temporal proximity to $0.6\,\text{sec}$ of *unconditioned stimulus* (US) pulse (for example, electric shock). When the dopaminergic factor is negative (as it is in the experiment of Handler et al., 2019), CS delivery prior to the US (forward conditioning) would predict synaptic depression; CS delivery after the US (backward conditioning) should predict synaptic potentiation. Fig. 2.2 shows the ER$-$Ca^{2+} and cAMP levels during forward and backward conditioning for a DAN with a depressing terminal, which are comparable to the data shown in Handler et al. (2019) (also shown in grey in Fig. 2.2). Fig. 2.2A shows the overall plasticity effect for the different forward and backward conditioning setups tested. Note that these overall plasticity effects were the focus here (not the detailed responses of Fig. 2.2B).

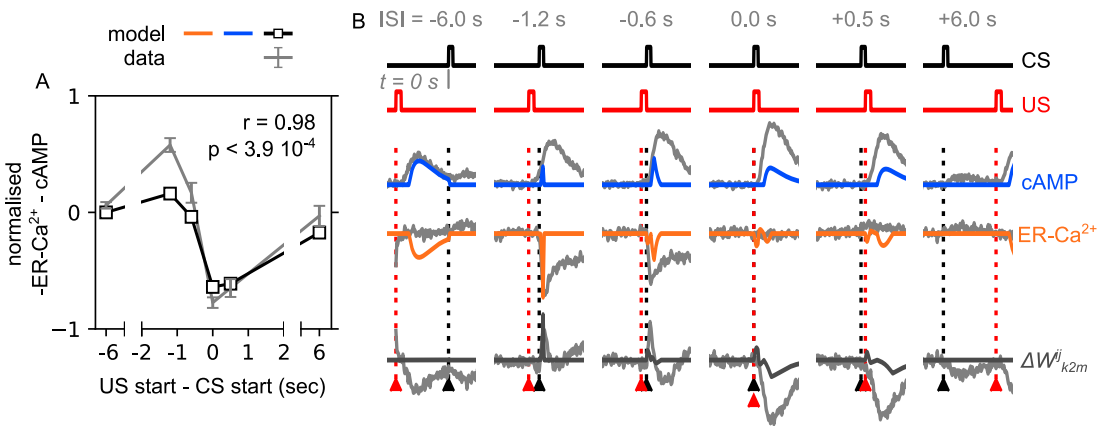

Figure 2.2: The effect of the ER$-$Ca^{2+} and cAMP based on the order of the *conditioned* (CS) and *unconditioned stimuli* (US). (A) Normalised mean change of the (modelled) synaptic weight plotted as a function of the Δs (US start - CS start), similar to Handler et al. (2019)—Figure 5F (blue line). For ease of comparison, the predicted mean values are drawn on the top of the data (mean $\pm$ SEM) from the original paper (Handler et al., 2019)—grey lines and error bars. (B) Detailed ER$-$Ca^{2+} and cAMP (modelled) responses reproduced for the different Δs, and their result synaptic-weight change. The black arrowhead marks the onset time of the CS (duration $0.5\,\text{sec}$); the red arrowhead marks (the onset) time of the US (duration $0.6\,\text{sec}$), similar to Handler et al. (2019)—Figure 5D. For ease of comparison, the predicted responses are drawn on the top of the data from the original paper (Handler et al., 2019; grey lines). Adapted from Gkanias, McCurdy, et al. (2022) under CC BY 4.0.

By replacing Eq. (2.12) and Eq. (2.13) in Eq. (2.11), the update rule can be rewritten as a function of known quantities, forming the DPR of Eq. (2.10), which is rewritten below for convenience,

$$\Delta w_{ij}(t) = \delta_j(t)\,[x_i(t) + w_{ij}(t) - w_{\text{rest}}], \qquad \text{where} \qquad \delta_j(t) = D_j^{\triangle}(t) - D_j^{\triangledown}(t).$$

In Fig. 2.2B (where a depressing DAN terminal is assumed), all four effects of the DPR occur in four out of the six cases, creating complicated dynamics that allow forward and backward learning. Similarly, a potentiating terminal might trigger all the effects in a row but in different order and duration. Also, a negative (inhibiting) feedback from the MBON to the DAN was assumed, as it was suggested by the fruit-fly MBs connectome (Aso, Hattori, et al., 2014) for the KC axons that Handler et al. (2019) recorded from (KC $\rightarrow$ MBON-γ4).

2.3 DISCUSSION

The proposed DPR depends only on information that is plausibly available in the presynaptic area of the KC axon—see Eq. (2.10). This includes the activity of the KC,

the level of DA, and the deviation of the current synaptic weight from a set-point resting weight. Note that it was not possible to obtain good results without this third component of the rule, although the underlying biophysical mechanism is unknown. A speculation could be that it involves synapsin, which has a direct role in regulating the balance of reserve and release vesicle pools, and it is required in the MB for associative learning (Michels et al., 2011). The rule also introduces a bidirectional dopaminergic factor. This is based on the results of Handler et al. (2019), who showed that the combination of DopR1 and DopR2 receptor activity can result in depression or potentiation of the synapse.

In contrast to the DPR, the RPE plasticity rule requires that the difference (error) between the post-synaptic MBON activity and the DA level is somehow calculated in the pre-synaptic KC axon. On a different note, the Hebbian and stochastic Hebbian plasticity rules need to calculate the correlation between the activities of the KCs and MBONs. However, both of these are inconsistent with the observation that learning is generally unaffected by silencing the MBONs during acquisition (Hige et al., 2015; Krashes, Keene, et al., 2007; Dubnau, Grady, et al., 2001; McGuire, Le, and Davis, 2001). In addition, the RPE, basic, Hebbian, neural-modulation, and anti-Hebbian plasticity rules predict no synaptic changes when the input units (KCs) are inactive. This is inconsistent with findings that activity of the pre-synaptic neurons is not required for the synaptic plasticity (Hige et al., 2015; Berry, Phan, and Davis, 2018). Based on the above constraints, only the template plasticity rule—Eq. (2.5)—seems plausible from the set of rules described in Section 2.1.

There are some important differences between the template plasticity rule and the DPR. First, in the template plasticity rule, explicit reinforcement is required for synaptic plasticity, which is not the case in the DPR. The dopaminergic factor depends on the responses of the DAN terminals, which are influenced by other factors in addition to the reinforcement (for example, feedback from the MBONs) or their potentiating/depressing property. Both plasticity rules have a conductor of plasticity. In DPR, this is the dopaminergic factor, and in the template rule, it is the activity of the KCs. However, the dopaminergic factor as the conductor provides more learning flexibility, as it can take both positive and negative values, forming the four different plasticity effects. A final difference is that the template plasticity rule assumes a constant decay of the synaptic strength, while the decay (recovery effect) of the DPR is proportional to the difference between the current synaptic strengths and the resting weight.

From a functional perspective, when the dopaminergic factor is negative, the DPR is similar to a more flexible version of the anti-Hebbian plasticity rule. This is, when the input is active and the respective synaptic weight is high, the synaptic weight is reduced to zero. The extra flexibility comes from the fact that when the input is

inactive the synaptic weight will restore its original value (resting weight) at the same rate. This happens by descending the error between the activity of the respective input and its synaptic strength. However, note that when the dopaminergic factor is positive, the synaptic weight ascends this error. This means that the responses of the input unit are now accumulated in the respective synaptic weight, which essentially integrates the responses to a memory that saturates (consolidation). The plasticity effects of the DPR can be used by specialised circuits of connections between output and DA units to implement a larger variety of plasticity functions, a subset of which is explored in the next chapter (Chapter 3). These circuits could also combine the plasticity effects of the DPR to produce any desirable learning phenomena, and allow learning and unlearning of input patterns in different ways. Chapter 4 shows that the experience-based flexible activity of the output units (caused by these functions) can explain the behaviour of fruit flies in elemental learning conditions.

2.4 METHODS

2.4.1 *Backward conditioning experiments*

Inspired by Handler et al. (2019), the DPR was challenged with the backward conditioning experiment (presented in Fig. 2.2), which involves different timings between the occurrence of the CS and US. For this experiment, the CS always occurred in time-step $t_{\text{CS}} = 0.0\,\text{sec}$, while the US occurrence varied, $t_{\text{US}} \in \{-6.0\,\text{sec}, -1.2\,\text{sec}, -0.6\,\text{sec}, 0.0\,\text{sec}, 0.5\,\text{sec}, 6.0\,\text{sec}\}$, resulting in the different conditioning cases shown in Fig. 2.2—forward (US occurs along or after the CS) or backward conditioning (US occurs before the CS). Each case was run for $T = 1,000$ discrete time-steps $t \in \{-7\,\text{sec}, ..., 8\,\text{sec}\}$ sampling at $66.67\,\text{Hz}$ ($\text{d}t = 0.015\,\text{sec}$). Following the set-up of Handler et al. (2019), in each experiment the CS was present for $0.5\,\text{sec}$ and the US for $0.6\,\text{sec}$.

The CS was activating a (single) KC, which represented the pre-synaptic activity,

$$k(t) = (1 - \frac{\text{d}t}{\tau_{\text{KC}}})\, k(t - \text{d}t) + \frac{\text{d}t}{\tau_{\text{KC}}}\,\text{CS}(t), \tag{2.14}$$

where $\tau_{\text{KC}} = 0.5\,\text{sec}$ is an exponential decay time time-constant, $\text{CS} \in \{0, 1\}$ is either active or inactive, and the KC activity is bounded in $k(t) \in [0, 2]$. The post-synaptic activity was represented by the MBON,

$$m(t) = k(t)\, w(t - \text{d}t), \tag{2.15}$$

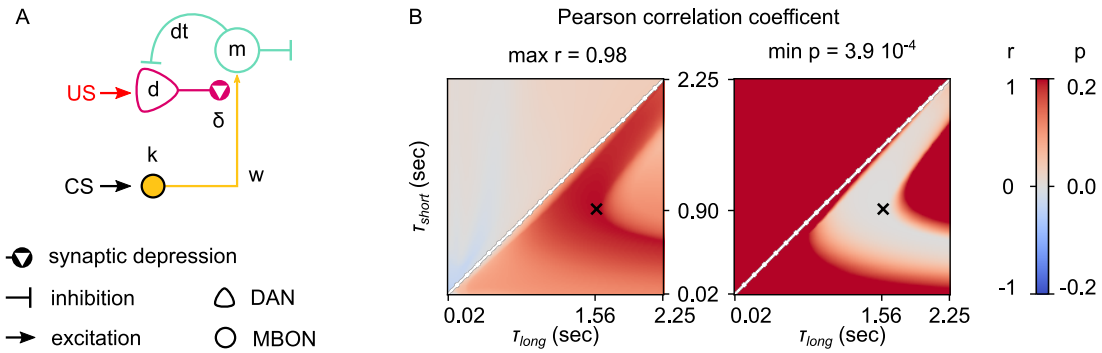

Figure 2.3: (A) Schematic of the microcircuit of a *Kenyon cell* (KC), a *dopaminergic neuron* (DAN), and a *mushroom body output neuron* (MBON) that was used to model the backward conditioning paradigm. (B) Parameter exploration for τ_{short} and τ_{long} in Eq. (2.17) and Eq. (2.18). For combinations of $\tau_{\text{short}} \in [0.02, 2.25]$ and $\tau_{\text{long}} \in [0.02, 2.25]$, the Pearson correlation coefficient was calculated between the predicted normalised mean change of the synaptic weight and the one extracted from the data of Handler et al. (2019), and the r and p values were reported. 'x' marks the pair of parameters with the highest correlation. Adapted from Gkanias, McCurdy, et al. (2022) under CC BY 4.0.

which was also bounded in $m(t) \in [0, 2]$; $w(t) \geq 0$ is the (positive) KC to MBON synaptic weight. Similarly, the the DAN activity is calculated as,

$$d(t) = \text{US}(t) - m(t - \text{d}t). \tag{2.16}$$

Note that the DAN activity depends on (inhibited by) the MBON activity in the previous times=-step. This was introduced to approximate the set-up of Handler et al. (2019) better, as the neurons they recorded from were part of a *susceptible memory* (SM) microcircuit (see Fig. 2.3A for a schematic and Section 3.3.2 for more details).

Following the above, the potentiating and depressing components of the dopaminergic factor were calculated as,

$$D^{\triangledown}(t) = \frac{\text{d}t}{\tau_{\text{short}}}\, d(t) + (1 - \frac{\text{d}t}{\tau_{\text{short}}})\, D^{\triangledown}(t - \text{d}t), \tag{2.17}$$

$$D^{\triangle}(t) = \frac{\text{d}t}{\tau_{\text{long}}}\, d(t) + (1 - \frac{\text{d}t}{\tau_{\text{long}}})\, D^{\triangle}(t - \text{d}t), \tag{2.18}$$

where $\tau_{\text{short}} = 0.9\,\text{sec}$ and $\tau_{\text{long}} = 1.56\,\text{sec}$ are the exponential decay time-constants that define the short (main) and long (secondary) duration of the dopamine effect. The longer the time constant, the slower the diffusion but also the lower the peak of the effect. The values of these two parameters were the result of a grid exploration illustrated in Fig. 2.3B. The difference between the potentiating and depressing compo-

nents results in the dopaminergic factor, which is used to update the synaptic weight through the DPR,

$$\delta(t) = D^{\triangle}(t) - D^{\triangledown}(t), \tag{2.19}$$

$$w(t) = \delta(t)\,[k(t) + w(t - \mathrm{d}t) - w_{\text{rest}}]. \tag{2.20}$$

Finally, the ER$-$Ca^{2+} and cAMP levels were calculated using Eq. (2.12) and Eq. (2.13), which are repeated here for convenience,

$$\text{ER}-\text{Ca}^{2+} = -D^{\triangle}(t)\,[k(t) - w_{\text{rest}}] - [D^{\triangle}(t) - D^{\triangledown}(t)]\,w(t - \mathrm{d}t),$$

$$\text{cAMP} = D^{\triangledown}(t)\,[k(t) - w_{\text{rest}}].$$

The detailed activities from all the neurons and chemical levels computed during the simulation are illustrated in Fig. E.1. A subset of them is also illustrated in Fig. 2.2B.

2.4.2 *Normalised mean change*

In Fig. 2.2A, the normalised mean change of the synaptic weight is reported, which is calculated using the computed ER$-$Ca^{2+} and cAMP levels, and the formula below,

$$\langle \Delta w \rangle \propto \frac{1}{T} \sum_{t=0}^{T-1} -\text{ER}-\text{Ca}^{2+}(t) - \text{cAMP}(t), \tag{2.21}$$

where t is a discrete time-step, T is the total number of recorded responses in a single trial, and $\langle \cdot \rangle$ denotes the mean over time.

3

MEMORY DYNAMICS

> *"Everything must be made as simple as possible. But not simpler."*
>
> Albert Einstein

Memory is an ambiguous term. In neurobiology, it is defined as a neuro-chemical process for storing experiences (Zlotnik and Vansintjan, 2019), but it can also be a neural circuit in the brain that allows for the acquisition and retrieval of experiences. This function is not trivial as most animals learn continuously (Sokolowski, 2001), which indicates complicated and highly dynamic systems that are able to assimilate memories by moving them to the *long-term memory* (LTM) and recall them later if needed.

The stimulus that triggers the memory acquisition (which is the process of storing) is usually called *reinforcement*, and it is highly correlated with the satisfaction state of the animal. This could be either a pleasant or unpleasant stimulus, which is associated with a joyful or irritating experience respectively. *Reinforcement learning* (RL) was inspired by these principles of satisfaction in animals and tried to optimise the performance of systems by trial-and-error experience (Barto, 1997). Similarly to animals, RL agents collect experiences and update their policy, which is used for making decisions about future actions. Theoretical advances in RL have greatly enhanced our understanding of biological systems (Niv, 2009; Gershman and Daw, 2017); for example, the *temporal difference* (TD) plasticity rule brought insights into how to interpret the activity of *dopaminergic neurons* (DANs) in the mammalian brain (Sutton, 1988). This suggests that studying the animal brain and using it as inspiration for computational models could be particularly fruitful to multiple disciplines. However, the size and complexity of the mammalian brain is an important barrier in progress. On the other hand, the smaller brains of insects allow for faster exploration and analysis of brain functions with the added benefit of higher resolution.

The fruit fly *Drosophila melanogaster* is able to form, retain, and forget olfactory associations with reinforcers (for example, electric shock). The key neural substrate is known to lie in neuropils of their brain called the *mushroom bodies* (MBs) (Davis, 1993; Heisenberg, 2003; Busto, Cervantes-Sandoval, and Davis, 2010). A variety of experimental paradigms have been developed to explore the function and structure

of these neuropils. Most of these paradigms are based on olfactory conditioning in fruit flies, where researchers try to create associations between an odour (*conditioned stimulus*, CS) and a reinforcer (*unconditioned stimulus*, US), imitating the *classical* (or *Pavlovian*) conditioning (Pavlov, 1949; Tully and Quinn, 1985). Some other paradigms are based on the visual navigation mechanism in honeybees and desert ants, where a view (experience) is associated with a behaviour, imitating *operant* (or *instrumental*) conditioning (Skinner, 1938). Many theoreticians have tried to model the MBs, concluding that their function is of similar complexity to neuropils in the mammalian brain, but with a more clearly organised structure. For their models, they usually use principles from RL, which has been the most powerful in tying together the three layers of Marr (1982) (which are, computation, algorithm, and implementation) into one coherent framework that is used for experimental investigations.

This chapter explores these highly dynamic systems in the insect brain and proposes a novel *incentive circuit* (IC), which utilises the *dopaminergic plasticity rule* (DPR) (proposed in the previous chapter) and exploits continuous memory dynamics. In addition, a computational model of the proposed circuit shows that its increased complexity is necessary for the explanation of the observed memory dynamics. The IC is anatomically accurate, it provides useful predictions on how memories are formed in the MBs and on how these memories can affect the behaviour of the animal. However, the predicted responses do not always match the ones observed in flies (due to unaccounted factors), highlighting that more accurate (and probably complicated) models of the MBs are yet to come.

3.1 THE INSECT BRAIN

Different insect species share many characteristics in their mid-brain structure. In particular, they all have a *central complex* (CX), two MBs (one per hemisphere), and some other *general neuropils* surrounding them. In this thesis, the general neuropils are defined as any neuropil in the mid-brain that is not part of the CX, MBs, or *antennal lobes* (ALs). These neuropils are less studied than the CX and MBs, but they have a role in controlling the insect's behaviour (Heisenberg, 2003). Fig. 3.1 gives an illustration of the mid-brain anatomy of fruit flies, while the following sections provide in-depth information about each neuropil, with a focus on the MBs.

3.1.1 *The mid-brain*

The CX is a neuropil in the insect brain responsible for navigation and orientation (shown in Fig. 3.1 with dark yellow colour). Pre-processed visual, tactile, and other

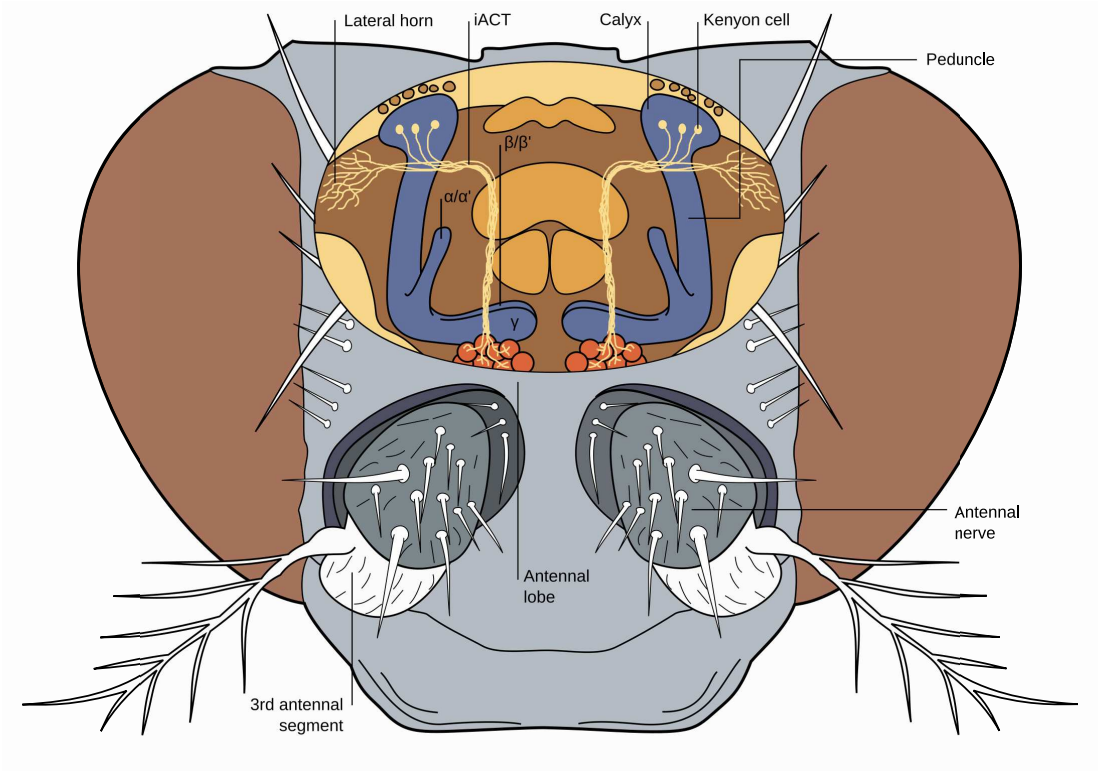

Figure 3.1: The mid-brain of fruit flies. The *antennal lobes* (ALs) receive an odour signal from the third antennal segments and maxillary palps (not shown). Glomeruli in the ALs organise the odour signals according to their chemical composition and project it to the *lateral horns* (LHs, placed at the sides of the mid-brain) and the calyces of the *mushroom bodies* (MBs). The projection happens through the *inner antennocerebral tract* (iACT). In the centre of the mid-brain (in dark yellow; moustache-shape, kidney-shape, and the two other neuropils below) is the *central complex* (CX), while the brown area that includes all these structures is the general neuropil (which includes the *convergence zones*, CZs, and the LHs), which interact with the MBs and CX. Adapted by permission from Springer Nature Customer Service Centre GmbH: Copyright © Heisenberg (2003).

sensory inputs enter the CX from the general neuropils surrounding it (shown in Fig. 3.1 with brown colour) through the tangential and columnar neurons, carrying orientation and localisation information. The CX uses this information to update the ring attractor of its *protocerebral bridge* (PB) (dark yellow moustache-like shape in Fig. 3.1), which indicates global orientation information. This information is used by its *path integrator* in the *fan-shaped body* (FB) (dark yellow bean-like shape in the centre of Fig. 3.1) to update its relative (to a point of interest) location information, and subsequently, it is pushed back to the general neuropils, which are responsible for controlling the behaviour of the animal.

The MBs are shown in Fig. 3.1 with blue, and the most favoured hypothesis for their function is that they contribute to the olfactory learning task (Heisenberg, 2003). They are associative RL centres that learn to associate pre-processed sensory input

(like odours) to behaviours. Recent studies showed that their function contributes to sensory-motor control, particularly visuomotor and that they guide behaviours through memories. In practice, MBs are able to manage associative memories with respect to sensory signals and (similarly to the CX) pass them to the general neuropils modulating the behaviour of the animal.

It is easier to think of the brain structure and its components as separate systems running in parallel; the sensory input creates a continuous information stream that flows to each system simultaneously. For example, when a fly is exposed to an odour, information about this odour is captured by the antennae of the animal and transferred to the ALs through the antennal nerve (see Fig. 3.1). This information is then transferred to the *lateral horns* (LHs) of the dorsolateral protocerebrum through the *inner antennocerebral tract* (iACT). Note that the LHs (which are parts of the general neuropils) are assumed to guide the innate behaviour of the animal. The iACT passes through the mid-brain of the animal, crossing the CX, MBs, and other general neuropils, which implies that information carried in these axons may be available in all of these neuropils. The MBs and CX consequently process this information in parallel and provide output to the general neuropils. This output is augmented with new input information coming from the sensors and iACT and is processed by the CX and MBs as before. This creates a recurrent loop that integrates sensory input and internal states into motivations that drive the behaviour of the animal.

Zooming in on the MBs, the *projection neurons* (PNs) transfer information from the iACT to the Calyx of the MBs, where the dendrites of the *Kenyon cells* (KCs) are placed. The KC axons transfer the information to the lobes of the MBs and finally to the *mushroom body output neurons* (MBONs). The axon terminals of the MBONs target different areas of the general neuropils, modulating different behaviours of the animal, which were initiated by the LHs. The next section discusses the anatomy and function of the MBs in more detail.

3.1.2 *The mushroom body*

The general structure of the *mushroom bodies* (MBs) is usually described as a set of intrinsic and extrinsic neurons. A single group of neurons, the KCs, are the only intrinsic neurons in the system. There are a number of groups of extrinsic neurons, including the DANs (whose output alters the connections among neurons) and the MBONs, which are both grouped in a number of different clusters.

Getting into the detailed structure, the dendrites of the KCs form the calyx in the dorsocaudal part of the mid-brain, which is the main sensory input to the system. The parallel axon fibres of the KCs move towards the frontoventral part of the mid-brain

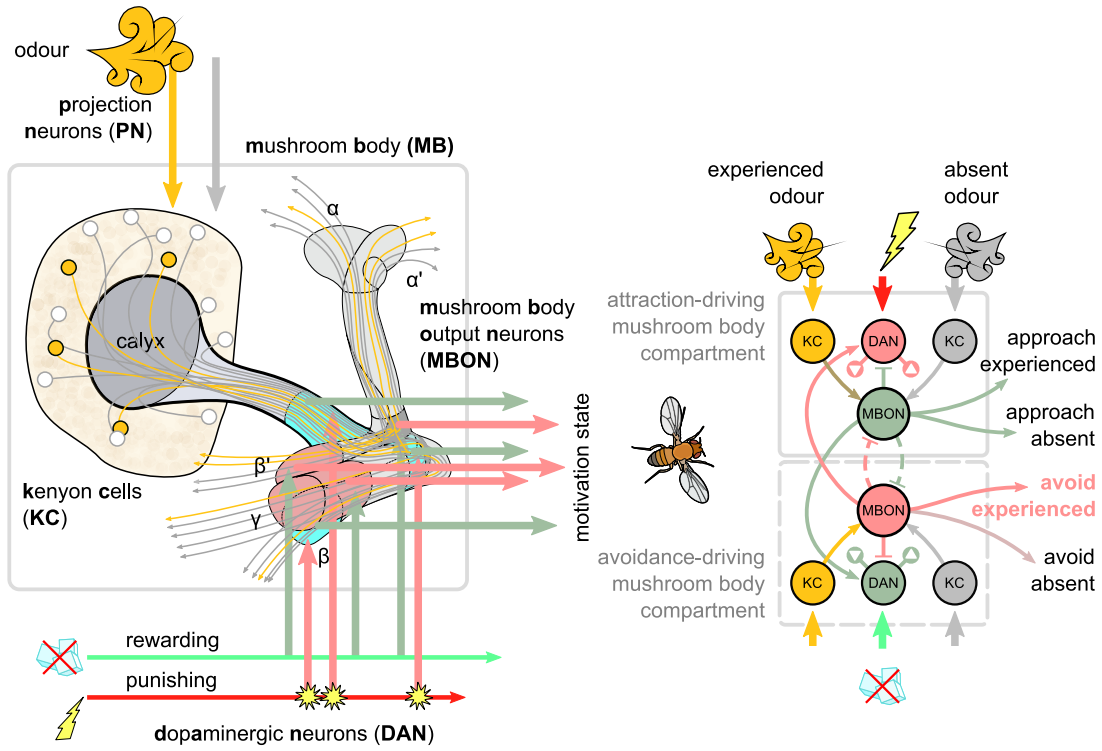

Figure 3.2: Overview of the *mushroom body* (MB) circuit. Left: the main anatomical pathways. In the illustration, the presented odour activates the *Kenyon cells* (KCs) through the *projection neurons* (PNs). The parallel axons of KCs propagate this signal to the lobes of the MB. The *mushroom body output neurons* (MBONs) extend their dendrites in the MB lobes, receiving input from the KCs. Electric shock creates a punishing signal that excites some *dopaminergic neurons* (DANs), whose axons terminate in the lobes and modulate the synaptic weights between KCs and MBONs. Right: schematic of potential connections between punishment/reward DANs and approach/avoidance MBONs. Note that although DANs transferring punishing signals modulate the KC activation of MBONs that encode positive motivations (decreasing attraction to the presented odour and increasing attraction to odours not present—see Chapter 2), MBONs that encode negative motivations will also gain higher responses due to release of inhibition between MBONs, and the feedback connections from MBONs to other DANs. In the *incentive circuit* (IC), these functions are further decomposed by three DANs and three MBONs for each motivation (positive or negative), and these units are mapped to specific identified neurons and microcircuits in the brain of *Drosophila melanogaster*. These circuits include some direct (but not mutual) MBON-MBON connections (dashed inhibitory connections). Adapted from Gkanias, McCurdy, et al. (2022) under CC BY 4.0.

through a stalk-like structure called the *pedunculus*, where they split in the medial (β/β′ and γ) and vertical (α/α′) lobes by turning the axons towards the central and dorsal parts of the brain respectively (see Fig. 3.2). *Dopaminergic* (DA), *octopaminergic* (OA), and other aminergic neurons project their axons onto the calyx, pedunculus, and MB lobes. Their dendrites are mainly extended in the *convergence zones* (CZs), which are a sub-group of neuropils from the general neuropils (see below). The aminergic neurons usually carry reinforcement information (and occasionally

sensory input) that is assumed to drive the learning mechanism of the MBs. Among others, there are two main clusters of DANs targeting the MBs, the *protocerebral anterior medial* (PAM), and the *protocerebral posterior lateral* (PPL1), both of which communicate reinforcement signals. The MBONs extend their dendrites in the calyx, pedunculus, and lobes, while they project their axon terminals to the CZ. These neurons are usually grouped by their neurotransmitter—which could be one of *glutamate* (Glu), *gamma-aminobutyric acid* (GABA), or *acetylcholine* (ACh)—or by the location where they extend their dendrites in the MBs. This grouping of the MBONs results in fifteen discrete compartments in the lobes of the MB, each of which is assumed to have its own role in modulating the behaviours of the animal.

Kenyon cells

As described before, the somas of the *Kenyon cells* (KCs) are densely packed above and behind the calyces, while their dendrites live in the calyces, and their long axons form the pedunculi which are then split in the medial and vertical lobes. There are around two thousand KCs in each MB of the adult flies, which are sequentially generated from four neuroblasts (K. Ito et al., 1998; Aso, Hattori, et al., 2014). There are three high-level groups of KCs: α/β, α'/β', and γ, which correspond to the lobes where they terminate their axons to. Table 3.1 summarises the details of the MB intrinsic neurons including KCs. Regarding the topological properties of KCs, the dendrites of the γ-KCs are claw-shaped and they occupy mostly the centre of the calyces. In contrast, the α/β and α'/β' have spiny and varicose shapes, and they are localised to a peripheral zone of the calyces (Heisenberg, 2003; Waddell, 2010). This suggests that the γ-KCs receive input from a broader number of PNs, while the other ones are more selective.

In the pedunculus, the long axons of the KCs often appear as concentric rings with the α/β and α'/β' KCs being in the outer layers, while the γ ones in the centre (Heisenberg, 2003). This is in contrast with their layout in the lobes, where they appear as laminae and form seven distinct layers: γ *main* and *dorsal* (d); α'/β' *middle* (m) and *anterior-posterior* (ap); and α/β *posterior* (p), *core* (c), and *surface* (s) (see Table 3.1). All layers have their dendrites in the main calyx, predominantly receiving olfactory information, except for the γd and α/βp, which have their dendrites in the *ventral accessory calyx* (vAC) and *dorsal accessory calyx* (dAC) respectively, and they receive information from different modalities (Aso, Hattori, et al., 2014). At the heel of the MB, the three groups of KCs form five lobes. The γ group forms the γ medial lobe, the α/β group forms the α vertical and β medial lobes by branching its axons and creating two separate terminals for each neuron, and the α'/β' group similarly forms the α' vertical and β' medial lobes as it is illustrated in Fig. 3.3.

categories	transmitter	cell type name	# cells	other names
Kenyon cells		γd	~ 75	
		γmain	~ 600	
		α′/β′ap	~ 210	α′/β′a, α′/β′p
		α′/β′m	~ 140	
		α/βp	~ 90	
		α/βs	~ 500	
		α/βc	~ 400	
Modulatory neurons	5HT amn	MB-DPM	1	DPM
	GABA	MB-APL	1	APL

Table 3.1: List of intrinsic cell types in the *mushroom body* (MB). There are three types of intrinsic neurons in the MBs: the *Kenyon cells* (KCs), *anterior paired lateral* (APL) neuron and the *dorsal paired medial* (DPM). Adapted and modified from Aso, Hattori, et al. (2014) under CC BY 4.0.

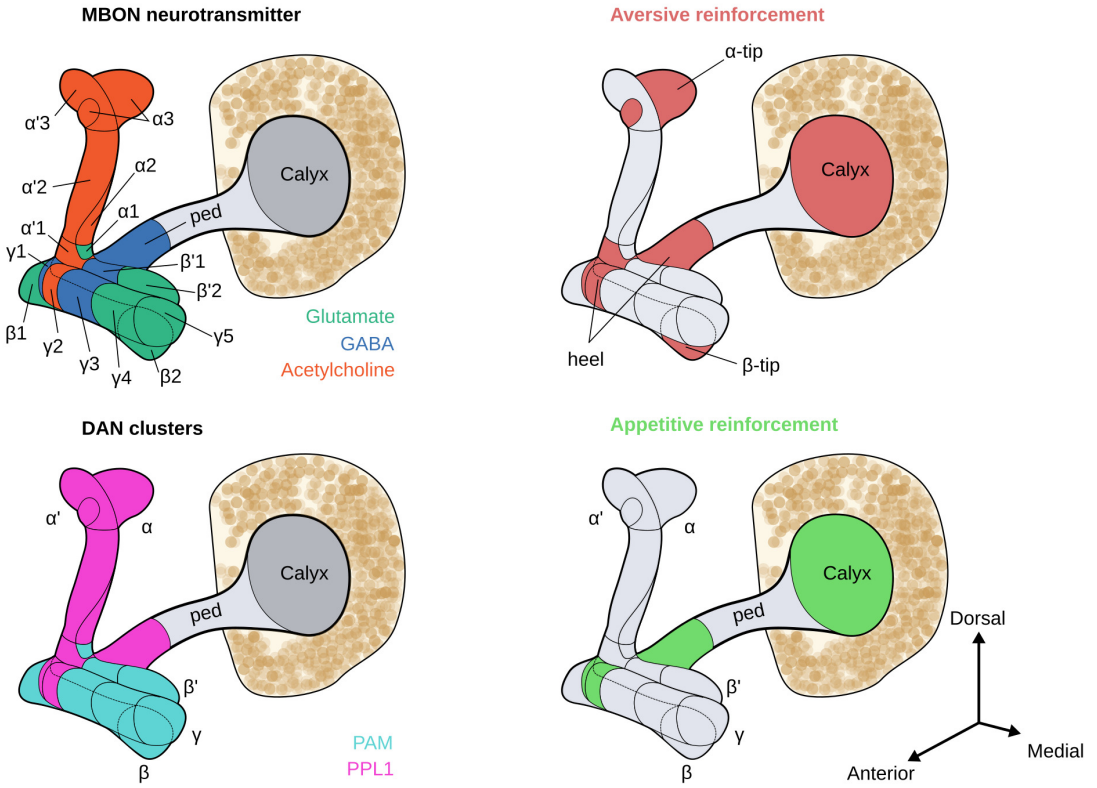

Figure 3.3: The involvement of *mushroom body output* (MBONs) and *dopaminergic neurons* (DANs) in aversive and appetitive reinforcement. Top-left: *mushroom body* (MB) compartments, grouped by the neurotransmitter of their respective MBONs. Bottom-left: MB compartments grouped by the cluster of DANs that terminate their axons there. Top-right: MB compartments involved in aversive reinforcements (in red). Bottom-right: MB compartments involved in appetitive reinforcements (in green).

Dopaminergic neurons

Extrinsic neurons provide input to and output from the MBs. *Dopaminergic neurons* (DANs) have various extrinsic target areas, typically in circumscribed regions that lie medial and lateral to the lobes (which are the CZs) and occasionally in the LH, where the PNs terminate (Heisenberg, 2003). Their axons project mainly in specific compartments of the MB lobes and occasionally in the pedunculus and calyx. There are around six hundred DANs in the brain of fruit flies (Waddell, 2010), but only around a hundred and thirty of them project to each of the MBs (Aso, Hattori, et al., 2014). More specifically, they project their axons in specific compartments within the lobes transmitting information about the reinforcement and guiding learning (Schwaerzel et al., 2003; Claridge-Chang et al., 2009; Mao and Davis, 2009; Aso, Siwanowicz, et al., 2010; Aso, Herb, et al., 2012; Burke et al., 2012; C. Liu et al., 2012). Aso, Hattori, et al. (2014) identified twenty distinct types of DANs, which are grouped in the PAM and PPL1 clusters (summarised in Table 3.2). There are six types of DANs in the PPL1 cluster and each one is composed of one to two neurons per hemisphere, while there are fourteen DAN types in the PAM cluster composed of around fifteen neurons on average. Each type of DAN responds to a different type of reinforcement (Riemensperger et al., 2005; Mao and Davis, 2009; Burke et al., 2012; C. Liu et al., 2012), suggesting that they may selectively communicate reinforcement information to the MBONs, and modulate the synaptic strength between the KC axons and the MBON dendrites. Fig. E.2 shows all the twenty types of DANs and how they interact with the compartments in the MB lobes.

As *dopamine* (DA) is responsible for modulating the KC $\rightarrow$ MBON synaptic strength, a reasonable conclusion would be that it is equivalent to the reinforcement signal. Instead of signalling the actual reward, there is a hypothesis that these responses could signal the 'error' between the expected and actual reward that the animal received (Waddell, 2010). Although this hypothesis has been extensively supported by the TD framework, and there is evidence in the mammalian brain that this framework can sufficiently explain the DAN activity, the same claim is not universally accepted for the insect brain (see Chapter 2).

From a behavioural aspect, Waddell (2010), Claridge-Chang et al. (2009), and Aso, Siwanowicz, et al. (2010) supported that aversive reinforcement is mainly transmitted by neurons in the PPL1 cluster. More specifically, neurons innervating the lower pedunculus, close to the junction and the α-tip show a strong reaction to shock (Mao and Davis, 2009; also see Fig. 3.3). However, artificially activating the PAM-β2β'2a neurons (three neurons of the PAM cluster mainly innervating the β-tip) substituted for an aversive stimulus in conditioning (Aso, Siwanowicz, et al., 2010). As for ap-

transmitter	short name	cell type name	# cells	other names
dopamine	PAM-01	PAM-γ5	$8-21$	aSP13, MB-M1?
	PAM-02	PAM-β′2a	$6-9$	
	PAM-03	PAM-β2β′2a	>3	MB-M3, MB-M1?
	PAM-04	PAM-β2	$8-19$	MB-M8 subset
	PAM-05	PAM-β′2p	$14-17$	MB-M5?, MB-AIM?
	PAM-06	PAM-β′2m	$12-15$	MB-M5?, MB-AIM?
	PAM-07	PAM-γ4<γ1γ2	$13-17$	subset of MB-AIM?
	PAM-08	PAM-γ4		subset of MB-AIM?
	PAM-09	PAM-β1ped	$1-3$	subset of MB-MVP1
	PAM-10	PAM-β1	$4-6$	subset of MB-MVP1 and MB-M8
	PAM-11	PAM-α1	>6	subset of MB-MVP1, MB-VP1
	PAM-12	PAM-γ3	$9-23$	MB-M2
	PAM-13	PAM-β′1ap	$13-14$	subset of MB-AIM?
	PAM-14	PAM-β′1m		subset of MB-AIM?
	PPL1-01	PPL1-γ1ped	$1-2$	MB-MP1, MP
	PPL1-02	PPL1-γ1	1	
	PPL1-03	PPL1-γ2α′1	1	MB-MV1
	PPL1-04	PPL1-α′3	1	
	PPL1-05	PPL1-α′2α2	1	
	PPL1-06	PPL1-α3	1	
		PPL2-ab	>1	
octopamine		OA-VPM3	1	
		OA-VPM4	1	
		OA-VPM5	1	
		OA-VPM2a	2	
5HT		CSD	1	
SIFamide		SIDamide	4	MB-C2?
GABA		MB-C1	>2	

Table 3.2: List of extrinsic modulatory cell types in the *mushroom body* (MB). Most extrinsic modulatory neurons in the MB are *dopaminergic* (DA). Adapted and modified from Aso, Hattori, et al. (2014) under CC BY 4.0.

petitive reinforcement, there is some support that it is mediated by the *octopaminergic neurons* (OANs) system (Schwaerzel et al., 2003; Heisenberg, 2003; Schroll et al., 2006). Evidence shows that a subset of these neurons, called the *octopamine-ventral unpaired a2* (VUMa2), innervate the calyx, while the *octopamine-ventral paired 4* (VPM4) innervate the heel of the MB (see Fig. 3.3). Their activation has been shown to sub-

stitute for sugar in the *D. melanogaster* larvae (Schroll et al., 2006) and honeybees (Hammer, 1993). A summary of the properties of these neurons can be found in Table 3.2 and in Busch et al. (2009). Waddell (2010) supported that OANs indirectly affect appetitive learning by relieving a specific group of DANs in the PPL1 cluster (the PPL1-γ1pedc, which project to the MB heel) and the ones in the calyx that communicate punishment. According to Krashes, DasGupta, et al. (2009), PPL1-γ1pedc neurons are inhibited by hunger, which suggests that they aim to retrieve appetitive memories. This could be implemented by OANs that relieve PPL1-γ1pedc neurons and aim to acquire appetitive memories.

Interestingly, despite the fact that activating the PPL1-γ1pedc neurons can trigger aversive learning (Aso, Siwanowicz, et al., 2010), silencing them (or any other PPL1 neurons) did not impair aversive or appetitive learning in *D. melanogaster* (Krashes, DasGupta, et al., 2009). More specifically, when these neurons were silenced the animal expressed attraction to the emitted odour, mimicking food deprivation. Consequently, Waddell (2010) concluded that PPL1-γ1pedc neurons provide reinforcement and motivational control.

In summary, irrespective of the cluster they belong to (PAM or PPL1), some DANs promote appetitive and some others aversive learning in the MBs. Their interaction with different groups of OANs might affect the type of promoted valence, but it could also be independent. The exact interaction between the different aminergic neurons is still unclear.

Mushroom body output neurons

The axons of the *mushroom body output neurons* (MBONs) are projected outside of the MBs and they usually terminate in the same regions where the DANs extend their dendrites (which is in the CZs and more sparsely in the LHs). Moreover, they extend their dendrites in the lobes, pedunculi, and calyces of the MBs, establishing connections with the axons of the KCs. Aso, Hattori, et al. (2014) found thirty-four MBONs in the adult fruit-fly brain, grouped by twenty-one types (as summarised in Table 3.3 and illustrated in Fig. E.2), each of which defining one to two neurons, except one which defines eight neurons. Another popular grouping of the MBONs is based on their neurotransmitter (ACh, Glu, or GABA). Neurons with the same neurotransmitter extend their dendrites towards neighbouring compartments in the lobes. In particular, most of the cholinergic neurons extend their dendrites towards the vertical lobes, glutamatergic neurons occupy the medial lobes, and GABAergic neurons prefer the pedunculus and the intersection between the lobes (see Fig. 3.3).

Their dendrites barely overlap with the ones of other MBONs when they are extended in different compartments. This sets the borders of the fifteen compartments

transmitter	short name	cell type name	# cells	other names
glutamate	MBON-01	MBON-γ5β′2a	1	MB-M6
	MBON-02	MBON-β2β′2a	1	
	MBON-03	MBON-β′2mp	1	MB-M4
	MBON-04	MBON-β′2mp_bilateral	1	
	MBON-05	MBON-γ4>γ1γ2	1	
	MBON-06	MBON-β1>α	1	MB-MV2
	MBON-07	MBON-α1	2	
GABA	MBON-08	MBON-γ3	1	
	MBON-09	MBON-γ3β′1	1	
	MBON-10	MBON-β′1	8	
	MBON-11	MBON-γ1pedc>α/β	1	MB-MVP2
acetylcholine	MBON-12	MBON-γ2α′1	2	
	MBON-13	MBON-α′2	1	MB-V4
	MBON-14	MBON-α3	2	MB-V3
	MBON-15	MBON-α′1	2	MB-V2
	MBON-16	MBON-α′3ap	1	MB-V2α′
	MBON-17	MBON-α′3m	2	MB-V2α′
	MBON-18	MBON-α2sc	1	MB-V2α′
	MBON-19	MBON-α2p3p	2	MB-V2
N. D.	MBON-20	MBON-γ1γ2	1	
	MBON-21	MBON-γ4γ5	1	
	MBON-22	MBON-calyx	1	MB-CP1

Table 3.3: List of extrinsic output cell types in the *mushroom body* (MB). *Mushroom body output neurons* (MBONs) are usually grouped by their neurotransmitter—*glutamate* (Glu), *gamma-aminobutyric acid* (GABA), and *acetylcholine* (ACh). Adapted and modified from Aso, Hattori, et al. (2014) under CC BY 4.0.

in the lobes, revealing a physical organisation of the outputs as illustrated in Fig. 3.3. As Tanaka, Tanimoto, and K. Ito (2008) proposed, five compartments have been allocated to each of the α/β, α′/β′, and γ lobes (see Fig. E.2). It has been suggested that the twenty-one different types of MBONs translate the odour identity encoded by the KCs to a multi-dimensional odour value. Aso, Hattori, et al. (2014) suggested that the thirty-four individual MBONs provide a representation that biases the behavioural responses; this can be interpreted as a behavioural control mechanism. Although (by definition) all MBONs terminate their axons outside the MB, three of them also have terminals in it. More specifically, the MBON-γ4>γ1γ2 have terminals in the γ1 and γ2 compartments, the MBON-γ1pedc>α/β in all the compartments in the α and β lobes, and the MBON-β1>α in all the compartments of the α lobe. This creates a

feed-forward MBON-network of four layers, where each MBON is connected to their corresponding KCs and MBONs from the previous layer, as shown in Fig. 3.4.

Interestingly, one type of DANs, the PAM-γ4<γ1γ2, shows the reversed effect: they extend their dendrites inside the MB (and more specifically in the γ1 and γ2 compartments) as well as outside. Along with MBON-γ4>γ1γ2, this type of PAM neurons creates a recurrent connection to the MBON-network which involves dopaminergic modulations (see Fig. E.2). The layers of the network described above are organised as follows (from deeper to shallower layers): α-lobe, β-lobe, γ1γ2pedc, and then the rest for the compartments (see Fig. 3.4). As the representations in deeper layers of the neural network are more complex, the population of the MBONs in the α lobe is expected to be able to achieve the highest level of representations (Aso, Hattori, et al., 2014).

The anterior paired lateral neuron

A special type of MB intrinsic neurons, the *anterior paired lateral* (APL), extends its dendrites in all the compartments of the MB (including the calyx, pedunculus, and lobes), and also projects its axon to the same regions (see Fig. 3.4). There is only one of these neurons for each hemisphere of the brain, located lateral to the MB calyx and close to the LH, and it releases the inhibitory GABA neurotransmitter in the calyx of the respective MB (Tanaka, Tanimoto, and K. Ito, 2008). Its responses are correlated to both the sensory input and the reinforcement signal in experiments involving an odour and electric shock (X. Liu and Davis, 2009).

Masuda-Nakagawa et al. (2014) also identified the APLs in the larval stage of fruit flies, reporting that (in contrast to the adult flies) they innervate only certain compartments of the lobes, and they do not innervate the pedunculus. On the other hand, in line with the adult flies, they innervate the whole calyx of the animal. It was noticed that the larval stage APL neuron appears quite polarised (which is, dendrites in the lobes and axon terminals in the calyx). In the adult stage, the same neuron loses its polarity, developing dendrites and axon terminals in all the regions of the MB (C.-L. Wu et al., 2013). This suggests that in early stages it may function as a source of global inhibition in the calyx, suppressing the learning behaviour, but later it is split in local inhibitory circuits all over the MB.

Optical imaging experiments have provided some evidence that there is mutual suppression between this neuron and olfactory learning (X. Liu and Davis, 2009). More specifically, silencing the activity of the APL neurons enabled faster response of the fruit flies to a learned odour (when compared to not silencing it), suggesting that the activity of this neuron suppresses the learning procedure. A potential mechanism to explain this is that this neuron controls the activation threshold of the KCs.

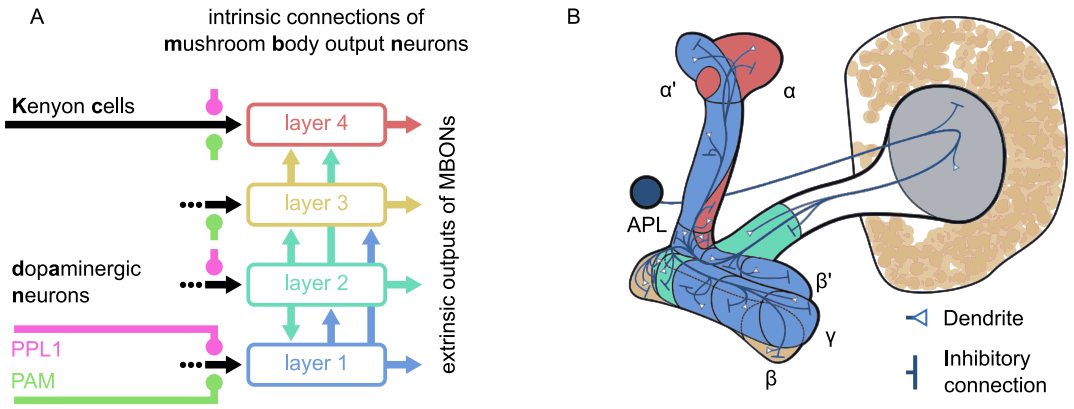

Figure 3.4: The MBON-network layers and the *anterior paired lateral* (APL) neuron. (A) Due to forward and feedback connections among *dopaminergic* (DANs) and *mushroom body output neurons* (MBONs), a four-layered neural network is formed in the *mushroom bodies* (MBs), with forward and recurrent connections. (B) The APL neuron connects to all these layers and creates local inhibition circuits. The different compartments of the MB are colour-coded with the same colours as the layers: (1) blue, (2) turquoise, (3) golden yellow, and (4) red.

In addition, the APL neurons showed a decrease in their response to learnt odours, which suggested a mutual suppression between the APL neuron and olfactory learning. On the other hand, Y. Wu et al. (2012) illustrated the opposite effect in reversal learning (which is, to learn that the odour does not predict the reinforcement after being trained otherwise). This means that, instead of suppressing it (as shown before), activating the APL neurons promoted reversal learning for the fruit flies, which suggested that APL neurons might be reinforcing for the animal. This was in contrast to the effect described by X. Liu and Davis (2009) and introduced some confusion about the function of these neurons in the MB.

A similar neuron to the APL is the *dorsal paired medial* (DPM) neuron (Waddell et al., 2000), which also exists as a single unit per hemisphere. Unlike the APL, this neuron innervates all compartments except the pedunculus, and its function has been related to memory consolidation and transfer between the MB compartments (Yu et al., 2005; Keene, Krashes, et al., 2006; Pitman et al., 2011; Haynes, Christmann, and Griffith, 2015).

Connections between the hemispheres

At the level of perception, signals from both antennae are projected to both ALs and then through the PNs to the KCs of the MBs (R. F. Stocker et al., 1990). At the KC level, there are limited connections between the two hemispheres of the brain (K. Ito et al., 1998). Interestingly, most of the MBONs innervate both the right and left MBs

(Tanaka, Tanimoto, and K. Ito, 2008). Exceptions are the MB-DPM, MB-C1, MB-C2, MBON-α3 and MBON-α′2 groups of neurons, which are hemisphere specific.

3.1.3 *Convergence zones*

The MBON axons terminate in five general neuropils. One of those is the LH (which is responsible for the innate behaviour of the animal), but 90 % of the terminals converge in the other four neuropils. These are the *crepine* (CRE), the *superior medial protocerebrum* (SMP), the *superior intermediate protocerebrum* (SIP), and the *superior lateral protocebrum* (SLP), which constitute the *convergence zones* (CZs) (Aso, Hattori, et al., 2014). Conditioned and unconditioned stimuli (like sensory input and internal reinforcements) also converge in these neuropils, along with information about the nature of the signal. DAN dendrites, axon terminals from neurons in the CX, and dendrites of neurons in the FB of the CX are also extended towards these regions (Hanesch, Fischbach, and Heisenberg, 1989; J. M. Young and Armstrong, 2010; Li et al., 2020). This suggests that the MBONs may indirectly modulate the activity of motor-related neurons in the CX and (through them) drive the behaviour of the animal (Strauss, 2002).

3.2 RELATED WORK

Several computational models have tried to capture the structure and function of the MBs, usually abstracting the common features of this structure across various insect species. Modellers have treated the MBs as performing odour discrimination (Huerta, Nowotny, et al., 2004), olfactory conditioning (A. Balkenius, Kelber, and C. Balkenius, 2006; D. Smith, Wessnitzer, and Webb, 2008; Finelli et al., 2008; J. Young et al., 2011; Wessnitzer et al., 2012; Peng and Chittka, 2017; Faghihi et al., 2017; Eschbach et al., 2020; Bennett, Philippides, and Nowotny, 2021; C. Zhao et al., 2021; Springer and Nawrot, 2021; Jiang and Litwin-Kumar, 2021), or calculating the scene familiarity (Z. Wu and Guo, 2011; Baddeley et al., 2012; Arena et al., 2013; Bazhenov, Huerta, and B. H. Smith, 2013; Ardin, Peng, et al., 2016). However, it seems like the MBs can subserve all these functions, depending on context (or experience), which is what drives the activity of the KCs (Cohn, Morantte, and Ruta, 2015). This suggests that the output neurons of the MB do not just inform the animal whether an odour is known or attractive, or if a scene is familiar, but they actually motivate the animal to take an action like approach, avoid, escape, or forage. There is emerging evidence supporting this idea of the MBONs driving non-binary but antagonistic motivations (Schwaerzel et al., 2003; Krashes, DasGupta, et al., 2009; Gerber, R. Stocker, et al.,

2009; Waddell, 2010; Lin et al., 2014; Perisse et al., 2016; Senapati et al., 2019), which has started to be explored in recent models.

In this section, the most important of the above computational models are summarised, which are those using properties of the MBs in order to build more biologically plausible computational models. Some of them were more constrained by the MB structure, while others made stronger assumptions about it in an attempt to produce reasonable behaviour. Different types of neural models were used for this purpose (firing rate, spiking, or neuromorphic), each with its own advantages and disadvantages. The most important characteristics of the models that are described in this section are summarised in Table 3.4.

3.2.1 *Firing-rate models*

The first model constrained by the structure of the MBs was built by Huerta, Nowotny, et al. (2004), who used the circuit of Fig. 3.5 in order to solve the odour discrimination task. More specifically, their model has the form of a three-layered neural network with the AL as the input layer, the KCs as the hidden layer, and the MB lobes as the output layer. These correspond to the three consecutive processing stages of their model: random non-linear mapping, linear classification [with similar characteristics to the *support vector machines* (SVMs)], and mutual inhibition [implemented using *winner-takes-all* (WTA)].

The AL $\rightarrow$ KC non-linear transformation aimed at producing a sparse KC representation and it did not involve learning. The respective transformation matrix, $\mathbf{C}$, had constant values, and the AL $\rightarrow$ KC transformation was given by,

$$x_i(t) = \Theta\Big[\sum_{j=0}^{n_{\text{AL}}-1} s_j(t)\, c_{ji} - \theta_{\text{KC}}\Big], \qquad i \in \{0, ..., n_{\text{KC}} - 1\},$$

where θ_{KC} is the threshold of the KC neurons, and $\Theta[\cdot]$ is a threshold function that ensured that the KC representation is a binary code—similar to the McCulloch and Pitts (1943) neurons. In the above equation, $\mathbf{s}(t) \in \mathbb{R}^{n_{\text{AL}}}$ is the AL neural vector code, $\mathbf{x}(t) \in \{0, 1\}^{n_{\text{KC}}}$ is the KC neural vector code, n_{AL} is the population of the AL neurons, and n_{KC} is the population of the KCs, where $n_{\text{AL}} << n_{\text{KC}}$.

The next transformation (KC $\rightarrow$ MBON) aimed at linear odour discrimination and inherited the properties of SVMs, which are to maximise the margin between the different classes and the lines separating them (Vapnik, 1995). The weight matrix, $\mathbf{W} \in \mathbb{R}^{n_{\text{KC}} \times n_{\text{MB}}}$, connecting these two layers was plastic, as Huerta, Nowotny, et al. (2004) assumed that this is where learning happens in the MB, and its plasticity was

	PN → KC		KC → MBON		
	learning rule	normalised KC	learning rule	MBON interact	neurons type
Huerta, Nowotny, et al. (2004)	—	LI	SH	WTA	FR (binary)
Z. Wu and Guo (2011)	H	LI	—	LI	FR (logistic)
Bennett, Philippides, and Nowotny (2021)	—	—	RPE	—	FR
C. Zhao et al. (2021)	—	—	RPE	—	FR
Springer and Nawrot (2021)	—	—	RPE	—	FR
Jiang and Litwin-Kumar (2021)	BP	—	RPE	—	FR
Finelli et al. (2008)	SRDP/ STDP	—	H	—	HH
D. Smith, Wessnitzer, and Webb (2008)	—	—	aH (ADPF)	—	IF
Wessnitzer et al. (2012)	—	—	3F (STDP)	—	IZ
Arena et al. (2013)	H	LE & GI	H (STDP)	—	IZ
Ardin, Peng, et al. (2016)	—	—	3F (STDP)	—	IZ
Peng and Chittka (2017)	H	Yes	3F (STDP)	—	IZ
Faghihi et al. (2017)	H	—	H	—	IZ
Schmuker, Pfeil, and Nawrot (2014)	—	LI	H	WTA	NM

Table 3.4: Summary of computational models of the *mushroom bodies* (MBs) that account for its anatomical properties. LI: lateral inhibition, LE: local excitation, GI: global inhibition, RPE: reward prediction error, H: Hebbian, aH: anti-Hebbian, SH: stochastic Hebbian, 3F: three-factor, BP: backpropagation, FR: firing rate, WTA: winner-takes-all, IF: integrate and fire, HH: Hodgin-Huxley, IZ: Izhikevich, NM: neuromorphic.

assumed to follow the Hebbian plasticity rule (Hebb, 1949). Using the above weight matrix, the values of the MBON vector were given by,

$$a_i(t) = \Theta\Big\{ \sum_{j=0}^{n_{\text{KC}}-1} x_j(t)\, w_{ji} - \text{WTA}_\mu\Big[\sum_{k=0}^{n_{\text{KC}}} x_k(t)\, w_{k\mu} \Big] \Big\}, \quad i \in \{0, ..., n_{\text{MBON}} - 1\},$$

where $\text{WTA}_\mu\Big[\sum_{k=0}^{n_{\text{KC}}} x_k(t)\, w_{k\mu} \Big]$ is the μ^{th} largest value of the respective n_{MBON}-sized vector. This worked as a non-constant threshold that introduced mutual inhibition in

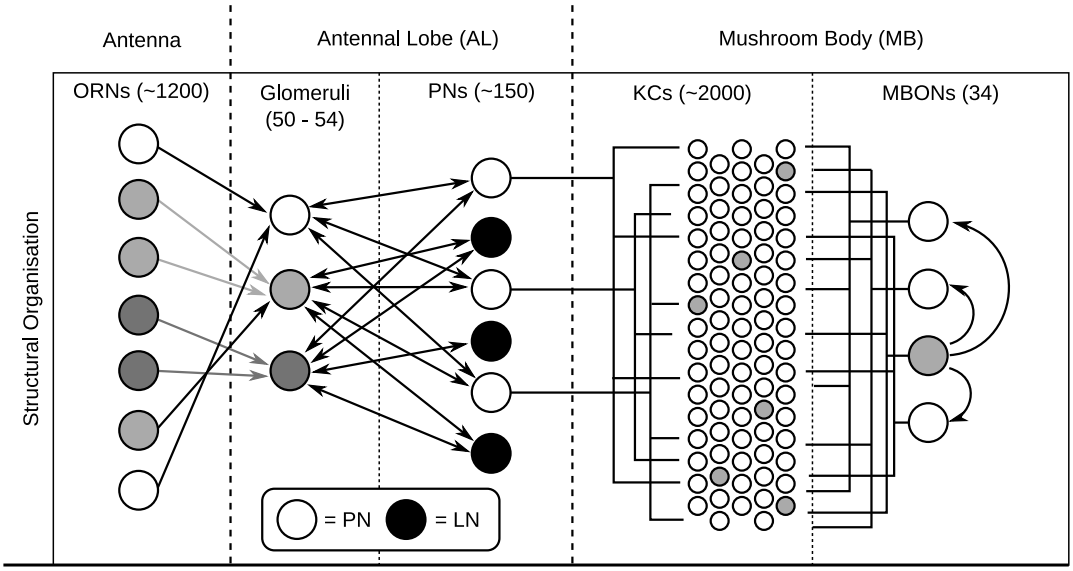

Figure 3.5: Schematic representation of the olfactory processing layers of insects related to the *mushroom bodies* (MBs). Input from the *olfactory receptor neurons* (ORNs) is grouped by the glomeruli in the *antennal lobes* (ALs). Those are projected onto the *Kenyon cells* (KCs) via the *projection neurons* (PNs), entering the MBs; they also interact within the ALs via the *local neurons* (LNs). Most computational models of the MBs start modelling from the PN layer. Adapted and modified from Huerta, Nowotny, et al. (2004) under CC BY 4.0.

the system, following the ideas of O'Reilly (2001)—namely, that the combination of mutual inhibition and Hebbian learning is biologically plausible. Note that, in a WTA configuration, this mechanism allows only μ winners to fire. Although in the above feed-forward system the neurons are deterministic, the plasticity rule used was the stochastic Hebbian rule of Section 2.1.6.

A rather different approach was taken by Z. Wu and Guo (2011), who tried to model how the visual information flows in the mid-brain of the fruit fly by using five modules [equivalent to layers in an *artificial neural network* (ANN)]. They assumed that the decision-making happens in the MBs (M-module) and the danger detection in the CX (D-module). A 'binding' or B-module represented indirect connections between the CX and MB, which were implied to be located in the CZs. Although in their model DA was released in the MB (M-module), learning happened only between the B-module and the CX (D-module). The individual firing-rate neurons were modelled using the logistic function, and the learning rule was given as,

$$\Delta w_{ij} = b_i(t)\, d_j(t) - 0.1\, d_i(t)\, d_j(t)\, w_{ij}(t),$$

where b_i and d_j are the responses of the i^{th} unit of B- and j^{th} unit of D-module respectively, and w_{ij} is the weight of the synapse connecting these units. Note that the

D-module carries the reinforcement signal (which is similar to the DAN responses) while the B-module must be similar to the MBON responses. Also, note that $w_{ij} = 0$ for $i \neq j$ and that all modules (except for the C-module, which represents the colours) have the same number of units (which is eighty), which in practice results in eighty parallel circuits of one unit per module.

Recently, a series of firing-rate MB models with MBON $\rightarrow$ DAN feedback connections were suggested almost in parallel. Bennett, Philippides, and Nowotny (2021) were the first to introduce such feedback connections. They ignored the processing before the KC layer of neurons, shifting their focus towards the processing thereafter. They indirectly used the *reward prediction error* (RPE) plasticity rule to update the synaptic weights, arguing that the combination of negative dopaminergic effect and positive feedback from the MBONs, or positive dopaminergic effect and negative feedback implement this rule. They proposed two models: *valence-specific* (VSλ) and *mixed-valence* (MV). In the VSλ model, MBONs excited or depressed their opposing-valence DANs, and each DAN either depressed or potentiated the synapses of their respective opposite-valence MBONs. In the MV model, DANs potentiated the synapses of their same-valence MBON; at the same time, they depressed the synapses of their opposite-valence MBON. Similarly, MBONs directly excited, and indirectly inhibited their opposite- and same-valence DANs respectively. They showed that both models could accurately explain the learning dynamic in a range of olfactory-conditioning-related behavioural experiments done with fruit flies.

Springer and Nawrot (2021) proposed a model that looks similar to the MV model. Two MBONs drove attraction (specifically, MBON-11 and the MB-V2 group; these are MBON-15 to MBON-19—see Table 3.3) and two avoidance (MBON-06 and MBON-01). Additionally, MBON-11 inhibits MBON-01, and MBON-06 inhibits the MB-V2 group; and the inhibited neurons excite their respective DANs. The PAM and PPL1 clusters of DANs mediated rewards and punishments respectively, depressing the KC $\rightarrow$ MBON synaptic weights of the avoidance- and attraction-driving MBONs. When the opposite reinforcement was delivered (with respect to each cluster) the target synaptic weights were gradually recovering, which introduced the extinction of memories (related to the findings of Felsenberg, Jacob, et al., 2018).

C. Zhao et al. (2021) also explored MBON $\rightarrow$ DAN connections as standalone microcircuits. More specifically, they explored the role of DANs as error- (DAN is the difference between the MBON activity and the electric-shock input) or target-encoding neurons (DAN is the electric-shock input) by also using the RPE plasticity rule. They suggested that a combination of the Hebbian and anti-Hebbian plasticity rules can result in the implementation of the RPE rule, and they demonstrated that none of the correlation-based plasticity rules can explain plasticity in the MBs. They finally

argued that there is experimental evidence supporting the predictions from both the error- (Hige et al., 2015) and the target-encoding (Aso, Sitaraman, et al., 2014) DAN models.

Jiang and Litwin-Kumar (2021) modelled the MB in the form of a deep neural network; calculated the MBON → DAN and MBON → MBON synaptic weights (among others) by using *back propagation* (BP), while the KC → MBON synapses remained plastic (using a similar rule to the MV model of Bennett, Philippides, and Nowotny, 2021). The model was trained to explain the backward conditioning data (Handler et al., 2019) and supported transfer learning and second-order conditioning. They suggested that the summarised DAN responses were implementing the RPE and allowed for valence (and reward) predictions.

3.2.2 *Spiking neural network models*

D. Smith, Wessnitzer, and Webb (2008) took a theoretical approach, exploring the neurochemical properties of the anti-Hebbian plasticity rule (defined in Section 2.1.7) by using a *spiking neural network* (SNN). This approach assumed that the correlation of a reward signal (r) with pre-synaptic (KC, p) activity was a key mechanism for learning, which is in line with the neural-modulation plasticity rule. However, in their equations this is multiplied by the post-synaptic activity, resulting in the anti-Hebbian rule. The actual rule they used for the update of the synaptic weights was given by the equations below,

$$\frac{\mathrm{d}g_{syn}}{\mathrm{d}t} = r(t)\,p(t)\,\eta_{syn} - [g_{syn} - g_{base}(t)]\,\kappa_{syn},$$
$$\frac{\mathrm{d}g_{base}}{\mathrm{d}t} = [g_{syn}(t)\,\alpha_{syn} - g_{base}]\,q(t)\,\eta_{base} - g_{base}\,\kappa_{base},$$

where $g_{syn}(t)$ and $g_{base}(t)$ are the pre- and post-synaptic conductances respectively, η_{syn} and η_{base} are the growth rates, κ_{syn} and κ_{base} are the decay rates, and α_{syn} influences the target value towards which $g_{base}(t)$ will grow; $p(t)$ and $q(t)$ are the pre- and post-synaptic history traces respectively.

The work of Huerta, Nowotny, et al. (2004) and D. Smith, Wessnitzer, and Webb (2008) inspired Wessnitzer et al. (2012) to construct an SNN in order to explain the experimental paradigms described by J. Young et al. (2011). This SNN used a simplified version of the *Hodgkin-Huxley* (HH) neuron model (Izhikevich, 2007) along with *spike-timing-dependent plasticity* (STDP) to model the network. Due to STDP, this network has some kind of *short-term memory* (STM) which works in a similar way to the eligibility traces in RL (Sutton and Barto, 1990). Because of the integration of

the STDP tag, aminergic reinforcement, and Hebbian learning, this is known as the *three-factor* plasticity rule and it is given by,

$$\begin{aligned}
\frac{\mathrm{d}g}{\mathrm{d}t} &= c(t)\, d(t), \\
\frac{\mathrm{d}c}{\mathrm{d}t} &= \frac{-c}{\tau_c} + \mathrm{STDP}(t_{\mathrm{pre}} - t_{\mathrm{post}})\, \delta(t - t_{\mathrm{pre/post}}), \\
\frac{\mathrm{d}d}{\mathrm{d}t} &= \frac{-d}{\tau_d} + \mathrm{DA}(t),
\end{aligned}$$

where g is the synaptic conductance, c is the synaptic 'tag' that maintains the eligibility trace, d is the DA extracellular concentration, $\mathrm{DA}(t)$ is the amount of DA in time-step t, $\delta(t)$ is the Dirac delta function, and τ_c and τ_d are time constants. In contrast to D. Smith, Wessnitzer, and Webb (2008), the above plasticity rule seems closer to the neural-modulation rule of Section 2.1.5 as it does not involve the responses of the output neurons—despite the fact that Hebbian learning was suggested as one of the three factors (Wessnitzer et al., 2012).

Ardin, Peng, et al. (2016) followed up the work of Wessnitzer et al. (2012), which was developed for olfactory conditioning, and they merged it with the work of Baddeley et al. (2012), who built a structure-free model for visual navigation to create a visual navigation memory based on the *D. melanogaster* olfactory system. In practice, they used the general concept and problem formulation of the visual navigation task and the more biologically constrained model of Wessnitzer et al. (2012) to build a system that rated the familiarity of panoramic views with respect to its experience. They then scanned the environment using the most familiar views to orientate an agent while navigating. They demonstrated that such a system was sufficient for navigation and they analysed its capacity in terms of memory. This work showed that a more complicated sequential RL system is not necessary for some simple navigation scenarios; associative RL (with the three-factor rule) was enough to describe a subset of behaviours.

Further development of this system focused on its capability for positive and negative patterning in the olfactory conditioning task (Peng and Chittka, 2017), which was first shown in Wessnitzer et al. (2012). This variation added plasticity in the PN $\rightarrow$ KC connections and introduced two classes of KC neurons: generalisation (I) and discrimination (II). The dendrites of the KC.I expanded towards more PNs (forty-five to fifty-five), while the dendrites of the KC.II expanded towards fewer (five to fifteen). Peng and Chittka (2017) suggested that this detail was important in order to model the peak-shift behaviour of honeybees.

Modelling the synaptic plasticity of PN $\rightarrow$ KC connections was not first seen in Peng and Chittka (2017). Working with the olfactory system of locusts, Finelli et al.

(2008) compared models with STDP and *spike-rate-dependent plasticity* (SRDP) in the MB calyx, concluding that STDP works better than SRDP in sparse representations. In both cases, the plasticity rule was described as,

$$\Delta g = L_r^{\pm}\, C_i\, F_p,$$

where $L_r^{\pm}$ is the plasticity rule factor, C_i is a linear function of instantaneous synaptic conductance, and F_p is a binary function of pairing frequency.

Arena et al. (2013) also modelled complicated dynamics in the MB, introducing synaptic plasticity in both the PN $\rightarrow$ KC and KC $\rightarrow$ MB connections (similar to Peng and Chittka, 2017). In their approach, they use simple STDP instead of the three-factor learning rule introduced by Wessnitzer et al. (2012) for the KC $\rightarrow$ MB connections, which was formalised as,

$$\Delta \mathbf{w} = \begin{cases} A^+ \exp(\Delta t/\tau^+), & \text{if} \quad \Delta t < 0, \\ -A^- \exp(\Delta t/\tau^-), & \text{if} \quad \Delta t > 0, \end{cases}$$

where the positive constants A^+ and A^- represent the maximum variation of the synaptic weight. Four extrinsic neurons were considered: reinforcement (RN), matching (SN), no-matching (DN), and premotor neurons (PmN). Two neuron types had dendrites in the lobes (PmN and SN), but only the PmN was an MBON that modulated the motor commands. In addition, they split the lobes into two clusters (α/β and α'/β') with the latter working as a 'backup copy' of the perceived information. Neurons from one cluster were connected to the neurons of the other via plastic synapses that followed the Hebbian principle. Neurons were also connected to neurons of the same cluster, exciting their immediate neighbours but inhibiting those further away. Each layer of the model (which are the AL, KC, and lobes) was based on lattices of spiking neurons; the purpose of the two MB lobes was to enable comparison between temporarily different stimuli. Their results showed that this model could recognise a previously presented stimulus, which they suggest matches the abilities of flies and bees. However, it seems that the model itself was designed to explicitly solve this task rather than being a clear mapping from neuroanatomy.

Faghihi et al. (2017) followed up the work of Wessnitzer et al. (2012) in using the simplified spiking neurons (Izhikevich, 2007), but they replaced the three-factor plas-

ticity rule with a Hebbian one for both PN $\rightarrow$ KC and KC $\rightarrow$ MBON connections. For the PN $\rightarrow$ KC connections, they use the following plastic synaptic weights,

$$w_i^{p2k} = \frac{\left(\sum_{j=1}^{N} R_{i,j}\right)^2}{1 + \left(\sum_{j=1}^{N} R_{i,j}\right)^2},$$

which model the long-range modulatory effect of retro-axonal signalling. Note that retro-axonal signalling is the phenomenon in that information can travel backwards along the axons. Here, the $R_{i,j}$ is the received retrograde signal by the synapse connecting neurons i and j. The KC $\rightarrow$ MBON plastic connections were modelled in two steps. First, the Hebbian rule was applied in the input as,

$$\frac{\mathrm{d}C}{\mathrm{d}t} = \frac{C}{\tau_c} + \Delta(\mathrm{KC}, \mathrm{MBON})\,\delta(t - t_c), \qquad C \geq 0,$$

where $\tau_c = 20\,\mathrm{ms}$ is a time-constant, and $\delta(t - t_c)$ is the Dirac function. $\Delta(\mathrm{KC}, \mathrm{MBON})$ follows the Hebbian principle; it is positive when the KC and MBON fire together, negative when only one of them fires, and zero when neither of them fires. This quantity was then used by the second step, which updates the KC $\rightarrow$ MBON synaptic weights by using the following equation,

$$\frac{\mathrm{d}w^{k2m}}{\mathrm{d}t} = Cd - \alpha\sqrt{w^{k2m}\frac{C}{d}},$$

where d is the extra-cellular concentration of DA, and $\alpha = 0.1$ the 'extinction' parameter. Faghihi et al. (2017) used this model to control the behaviour of a simulated fly, which was placed in an arena of multiple odours that were sparsely paired with an electric shock.

3.2.3 *Neuromorphic models*

The MBs also attracted the interest of neuromorphic computing modellers. The approach taken by Schmuker, Pfeil, and Nawrot (2014) was inspired by the insect olfactory pathway, and it used a neuromorphic circuit to build a system able to classify handwritten digits, similarly to Huerta and Nowotny (2009). Their plasticity rule was based on the Hebbian eligibility constraint, adding a fixed amount of potential to the eligible synapses when the classification is correct while removing it otherwise,

$$\Delta w_{ij} = \begin{cases} \epsilon_i\, c, & \text{if classification was correct,} \\ -\epsilon_i\, c, & \text{if classification was incorrect,} \end{cases}$$

where c is a constant value, and $\epsilon \in \{0, 1\}$ indicates whether the firing-rate of the presynaptic neuron i has exceed the threshold. Here, reinforcement was based on the correct or incorrect classification, given by predefined labels (which is essentially supervised learning). Finally, they compare their system with the *D. melanogaster* MBs in terms of function and computational speed, and they claimed to see many similarities between them.

3.3 RESULTS

The previous section provided a summary of the most successful computational models that tried to describe the function and role of the MBs. These included feed-forward and WTA circuits as well as recent advances in exploring MBON → DAN connections. In most of these models, the update in the synapses is based on variations of RPE or Hebbian plasticity, either via the plasticity rule itself or via feedback connections. This section describes a novel (firing-rate) model of the MB, the IC, which is characterised by MBON → DAN connections that have been shown to exist in the MB of *D. melanogaster*. Despite the simplicity of the neural model (which is non-spiking), this circuit combined with the DPR from the previous chapter enables elegant memory dynamics that have also been observed in fruit flies.

3.3.1 *The incentive circuit*

The *incentive circuit* (IC) is a circuit in the MB of the fruit fly *D. melanogaster* that enables complicated memory dynamics through self-motivation. Gkanias, McCurdy, et al. (2022) identified and modelled this circuit (illustrated in Fig. 3.6), which consists of six MBONs that receive KC input, and six DANs that modulate the KC → MBON connections. The circuit includes some inhibitory MBON → MBON connections and some (both inhibitory and excitatory) MBON → DAN feedback connections. All the neurons and connections in this circuit were mapped to identified connectivity in the MB, as summarised in Table E.1. The IC is composed of a number of known microcircuits (described in detail in the following sections); a specific function for each microcircuit was proposed, which was supported by biological justification. In this section, an abstract overview of the IC and its function is presented.

For each motivational state (attraction or avoidance) the IC has three types of MBONs (*susceptible*, *restrained*, and *LTM*) and three types of DANs (*discharging*, *charging*, and *forgetting*; see Fig. 3.6). Working from the outer edges of the model, the discharging DANs respond to punishment (left side) or reward (right side), and they influence the susceptible MBONs, which (by default) respond to all the KC

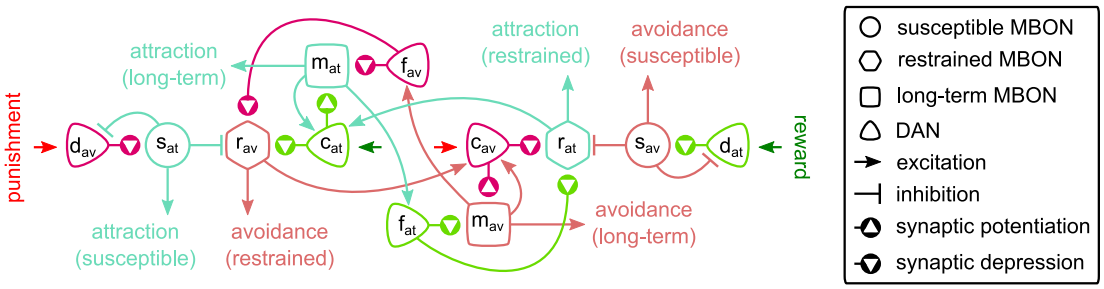

Figure 3.6: The *incentive circuit* (IC) integrates the different microcircuits of the *mushroom body* (MB) into a unified model allowing the expression of more complicated behaviours and memory dynamics. It combines the *susceptible* (SM), *restrained* (RM), *reciprocal short-* (RSM) and *long-term memories* (LTM and RLM) and the *memory assimilation mechanism* (MAM) microcircuits in one circuit that is able to form, consolidate and forget different types of memories that motivate the animal to take actions. d_{av} and d_{at}: avoidance- and attraction-driving discharging *dopaminergic neurons* (DANs). c_{av} and c_{at}: avoidance- and attraction-driving charging DANs. f_{av} and f_{at}: avoidance- and attraction-driving forgetting DANs. s_{av} and s_{at}: avoidance- and attraction-driving susceptible *mushroom body output neurons* (MBONs). r_{av} and r_{at}: avoidance- and attraction-driving restrained MBONs. m_{av} and m_{at}: avoidance- and attraction-driving *long-term memory* (LTM) MBONs. Adapted from Gkanias, McCurdy, et al. (2022) under CC BY 4.0.

inputs (not shown). The susceptible MBONs in turn inhibit the responses of the restrained MBONs of the opposite valence. When the discharging DANs depress the response of the opposite susceptible MBONs, this releases the response of the restrained MBONs of the same valence and also decreases the inhibitory feedback to the discharging DANs. The restrained MBONs activate their respective (same valence) charging DANs, which start to potentiate the LTM MBONs of the same valence, while also depressing the response (to KC input) of the restrained MBON of opposite valence. Similarly, the LTM MBONs enhance the activity of the charging DANs, increasing the momentum of LTM, while simultaneously activating their respective forgetting DANs to decrease the momentum of the opposite valence LTM. The forgetting DANs also depress the restrained MBONs which makes space for the acquisition of new memories while preserving old ones.

The following sections show in detail how each simulated neuron of this circuit responds during acquisition and forgetting in the aversive olfactory conditioning paradigm shown in Fig. 3.7. These (modelled) responses were compared to (calcium imaging) responses observed in the corresponding neurons of the fly under the same experimental paradigm. Chapter 4 provides more details on the different olfactory conditioning paradigms used to test plasticity mechanisms in fruit flies.

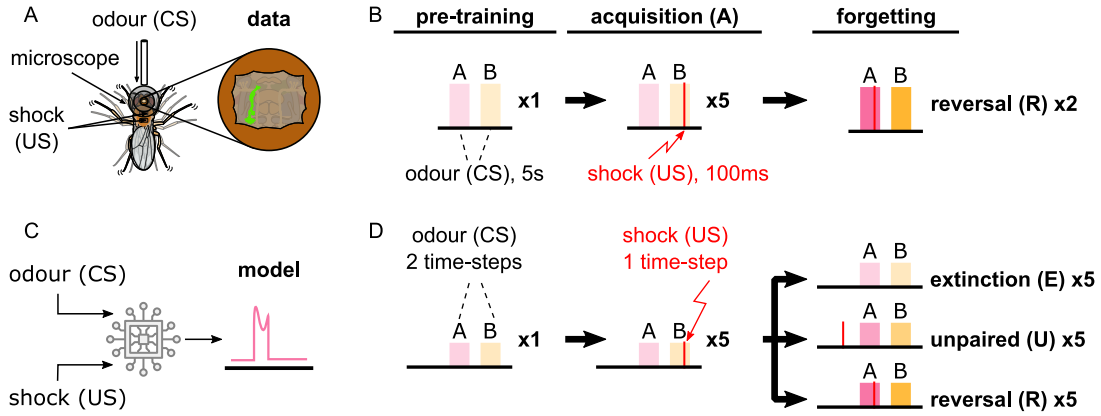

Figure 3.7: Description of the experimental setup and the aversive olfactory conditioning paradigms. (A) Setup for visualising neural activity via Ca^{2+} imaging during aversive olfactory memory acquisition and reversal. Flies were head-fixed and cuticle dissected for ratiometric imaging of Ca^{2+}-sensitive $GCaMP_6f$ and Ca^{2+}-insensitive tdTomato. (B) The aversive olfactory conditioning experimental paradigm. 5 sec presentations of odours A (*3-octanol*, OCT3; coloured pink) and B (*4-methylcyclohexanol*, MCH4; coloured yellow) continuously alternate, separated by fresh air, while the shock input (100 ms of 120 V) forms the different phases: 1 repeat of pre-training, where no shock is delivered; 5 repeats of acquisition, where shock (thin red line) is delivered in the last second of odour B; 2 repeats of reversal where a shock is paired with odour A. (C) Abstract representation of the computational model as an electronic chip. The model receives the conditioned (odour) and unconditioned stimuli (electric shock) and produces the *dopaminergic* (DAN) and *mushroom body output neuron* (MBON) responses using the *incentive circuit* (IC) and the *dopaminergic plasticity rule* (DPR). (D) The aversive olfactory conditioning experimental paradigm modified for testing the model. Odours A (coloured pink) and B (coloured yellow) are presented for two time-steps each, in alternation, separated by 1 time-step fresh air, while the shock input forms the different phases and forgetting conditions: 1 repeat of pre-training, where no shock is delivered; 5 repeats of acquisition, where a shock is delivered in the second time-step of odour B; 5 repeats of forgetting that can be either extinction (the lightest shade of odour colour) where the shock is not presented, unpaired (mid shade of odour colour) where shock (thin red line) is paired with the fresh air 'break', or reversal (dark shade of odour colour) where a shock is paired with odour A. Adapted from Gkanias, McCurdy, et al. (2022) under CC BY 4.0.

3.3.2 *Susceptible & restrained memories*

Pavlowsky et al. (2018) identified a microcircuit in the MB, where a punishment-encoding DAN (PPL1-γ1pedc) depresses the KC synapses onto an attraction-driving MBON (MBON-γ1pedc>α/β), which in turn inhibits the same DAN. They argue this is a memory consolidation mechanism, as the drop in the MBON response will reduce its inhibition of the DAN, and this would enhance the formation of the memory if the same odour-punishment pairing is continued. Felsenberg, Jacob, et al. (2018) further showed that the same MBON directly inhibits an avoidance-driving MBON (MBON-γ5β'2a), such that its activity increases (driving avoidance) after punish-

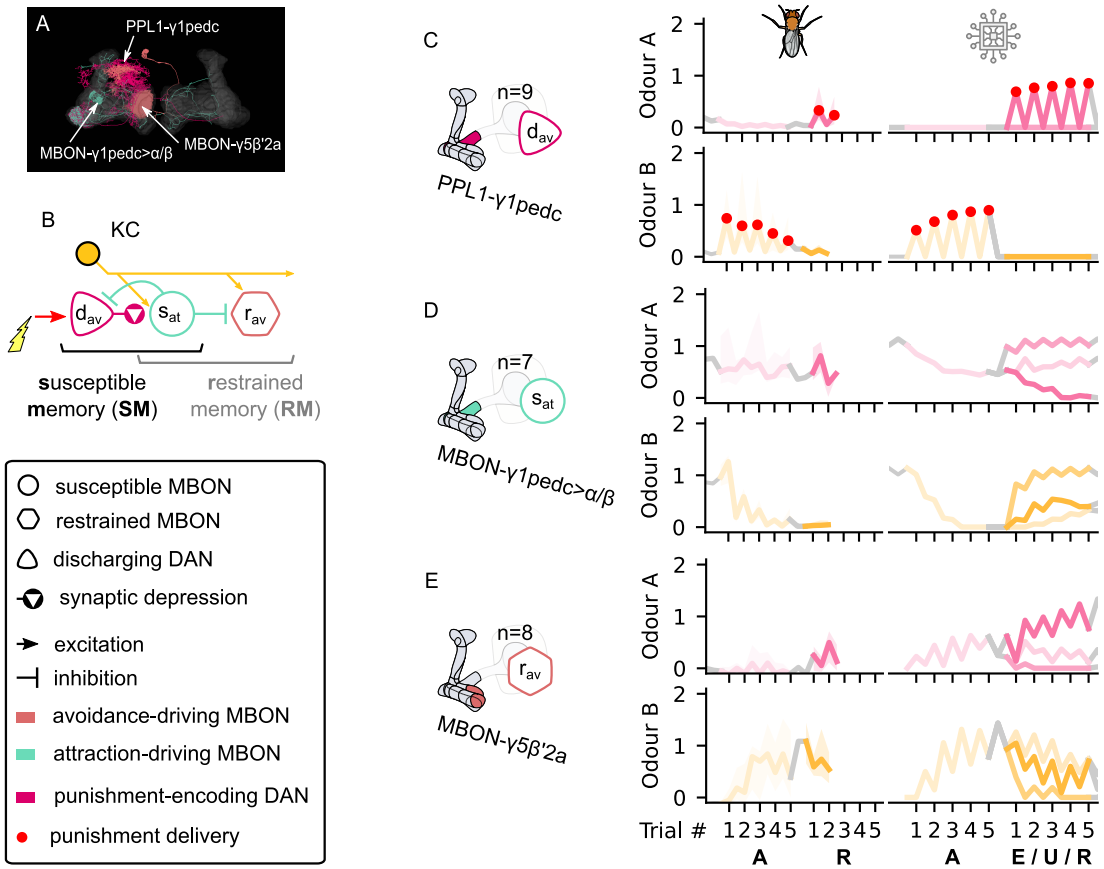

Figure 3.8: The *susceptible* (SM) and *restrained memory* (RM) microcircuits of the *mushroom body* (MB). (A) Image of the attraction-driving susceptible and avoidance-driving restrained memory microcircuits made of the PPL1-γ1pedc, MBON-γ1pedc and MBON-γ5β′2a neurons—created using the Virtual Fly Brain software (Milyaev et al., 2012). (B) Schematic representation of the susceptible and restrained microcircuits connected via the susceptible *mushroom body output neuron* (MBON). The responses of (C) the punishment-encoding discharging *dopaminergic neuron* (DAN), d_{av}, (D) the attraction-driving susceptible MBON, s_{at}, and (E) the avoidance-driving restrained MBON, r_{av}, generated by experimental data (left) and the model (right) during the olfactory conditioning paradigms of Fig. 3.7D. The lightest shades denote the extinction, the mid shades the unpaired and the dark shades the reversal phase. For each trial two consecutive time-steps are reported: the off-shock (odour only) followed by the on-shock (paired odour and shock) when available (odour B in acquisition and odour A in reversal phase) otherwise a second off-shock time-step (all the other phases). Adapted from Gkanias, McCurdy, et al. (2022) under CC BY 4.0.

ment. Fig. 3.8A shows these neurons in the MB and Fig. 3.8B a schematic representation of their interconnections. Note that the MBON → MBON inhibition is not reciprocal; rather, it is assumed (see Fig. 3.6 and below) that there is a different microcircuit in which an avoidance-driving MBON inhibits an attraction-driving MBON. Fig. 3.8C-E show the responses of these neurons from experimental data (left) and from the modelled IC (right), captured during aversive conditioning (the paradigm shown in Fig. 3.7). The comparison between the responses of the flies and those of

the model was done qualitatively (rather than quantitatively), as the model does not always provide a good fit for the data (but is still better compared to other models). This was done by visually comparing the relative (to the first trial) trend of the neural responses produced by the flies and the model (up to the second reversal trial—dark shades—as there are no data from thereafter), weighing more the off-shock time-step (the first of the two responses per trial) as this demonstrates better the response of the neuron without the direct influence of the reinforcement. Fig. 3.8C-E clearly demonstrate that the recorded and modelled responses follow a similar pattern.

Both in the experimental data and the model, learning in this circuit is shown by the sharp drop of the response of MBON-γ1pedc>α/β (Fig. 3.8D) to odour B. This is already apparent from the second trial of the acquisition phase. There is a similar drop in the response to odour A in the reversal phase. This rapid decrease is due to the depressing effect of the DAN on the KC → MBON synaptic weight. Note that this DAN is named 'discharging' because the target synaptic strengths are high (or charged) by default. However, due to the recovery effect of the DPR, if the reinforcement subsequently occurs without the KC activity (see unpaired phase in the model, for which experimental data are missing), the MBON synaptic weights reset to their resting weight (see Fig. E.6A—Odour B). This is consistent with the high learning rate and low retention time observed in Aso and Rubin (2016), and it results in a memory that is easily created and erased: a *susceptible memory* (SM). The response of MBON-γ5β′2a (Fig. 3.8E) can be observed to have the opposite pattern; it starts to respond to odour B from the second trial of acquisition as it is no longer restrained. Note however that, when the restraint is removed, the response it expresses also depends on its own synaptic weights for KC input, which may be affected by other elements in the IC. The experimental data of the DAN (PPL1-γ1pedc, Fig. 3.8C) shows a slight drop in the shock response during the whole experiment (first paired with odour B, then with odour A). This drop may reflect a sensory adaptation to shock, which is not included in the model of the IC. Consequently, the model shows a positive feedback effect; the DAN causes depression of the MBON response to the odour, which reduces inhibition of the DAN. This increases the DAN response and causes further depression in the MBON. Note that this is opposite to the expected effects of RPE.

Similar microcircuits in the MB can be extracted from the connectome described in Aso, Hattori, et al. (2014) and Li et al. (2020) (also identified in larvae; Eichler et al., 2017). This leads to the assumption that there may be exactly corresponding susceptible and *restrained memory* (RM) microcircuits with opposite valence; a reward-encoding DAN that discharges the response to odour of an avoidance-driving MBON, which in turn releases its restraint on an attraction-driving MBON (see Fig. E.4B and right side of the IC in Fig. 3.6, which mirrors the left side but with opposite valence).

Some suggestions of specific identities for the neurons forming this circuit are the PAM-γ4<γ1γ2 as the reward-encoding discharging DAN, the MBON-γ4>γ1γ2 as the avoidance-driving susceptible MBON, and MBON-γ2α′1 as the attraction-driving restrained MBON (see Fig. E.4A). The latter identification is based on the possibility of inhibiting connections from MBONs in the γ4 compartment to the ones in the γ2 compartment, as suggested by Aso, Hattori, et al. (2014) and Cohn, Morantte, and Ruta (2015). Although MBON-γ4>γ1γ2 is characterised by the Glu neurotransmitter, it is possible that it can inhibit MBON-γ2α′1 though Glu-gated chloride channels (Cleland, 1996; W. W. Liu and Wilson, 2013; McCarthy et al., 2011). Although the susceptible MBONs are by definition inhibitory, they could still contribute to the behaviour of the animal (promote attraction or avoidance). However, throughout this thesis, Dale's law was assumed (Dale, 1935); it states that a neuron can be either inhibitory or excitatory (not both). The susceptible MBONs could still promote attraction or avoidance (indirectly) by disinhibiting other neurons that promote these behaviours (possibly located in the CZs).

3.3.3 *Reciprocal short-term memories*

McCurdy et al. (2021) suggested that the attraction-driving restrained MBON in the previous circuit (MBON-γ2α′1) indirectly decreases the synaptic weights from KCs to the avoidance-driving restrained MBON (MBON-γ5β′2a) via an attraction-encoding DAN (PAM-β′2a). This microcircuit was also supported by Felsenberg, Jacob, et al. (2018) and Berry, Phan, and Davis (2018). Cohn, Morantte, and Ruta (2015) and Li et al. (2020) suggested that the corresponding avoidance-driving restrained MBON (MBON-γ5β′2a) excites an avoidance-encoding DAN (PPL1-γ2α′1), which closes the loop by affecting the connections from the KCs to the attraction-driving restrained MBON and forms the *reciprocal short-term memory* (RSM) microcircuit (actual neurons in the MBs shown in Fig. 3.9A; a schematic representation of the described connections shown in Fig. 3.9B).

The charging DANs, PAM-β′2a and PPL1-γ2α′1 (named after their LTM charging property—see Section 3.3.4), should be activated directly by reinforcement, as well as by the restrained MBONs; memories of a given valence are also affected by the expression of memories of the opposite valence. The latter feature keeps the balance between the memories by automatically erasing a memory when a memory of the opposite valence starts building up, and results in the balanced learning rate and retention time observed in Aso and Rubin (2016). As the memories in this pair of restrained MBONs are very fragile, these MBONs can be thought to store STMs.

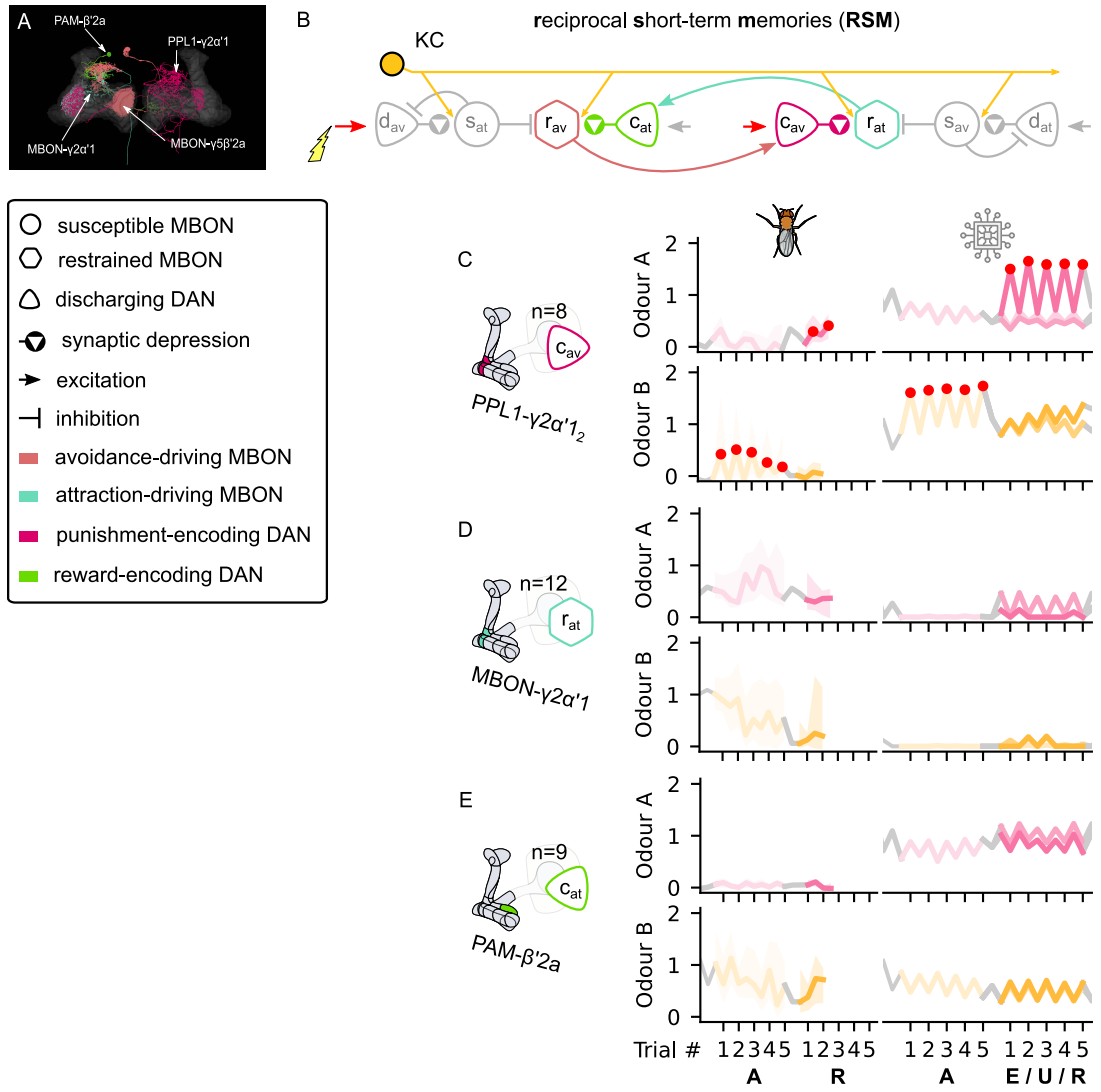

Figure 3.9: The *reciprocal short-term memories* (RSM) microcircuit of the *mushroom body* (MB). (A) Image of the reciprocal short-term memories microcircuit made of the MBON-γ5β′2a, PAM-β′2a, PPL1-γ2α′1 and MBON-γ2α′1 neurons—created using the Virtual Fly Brain software (Milyaev et al., 2012). (B) Schematic representation of the reciprocal short-term memories microcircuit (coloured) connected to the susceptible memories via the restrained *mushroom body output neurons* (MBONs). The responses of (C) the punishment-encoding charging *dopaminergic neuron* (DAN), c_{av}, the (D) attraction-driving restrained MBON, r_{at}, and (E) the reward-encoding charging DAN, c_{at}, generated by experimental data (left) and the model (right) during the olfactory conditioning paradigms of Fig. 3.7D. The lightest shades denote the extinction, the mid shades the unpaired and the dark shades the reversal phase. For each trial two consecutive time-steps are reported: the off-shock (odour only) followed by the on-shock (paired odour and shock) when available (odour B in acquisition and odour A in reversal phase) otherwise a second off-shock time-step (all the other phases). Adapted from Gkanias, McCurdy, et al. (2022) under CC BY 4.0.

The effects of this circuit (as shown in Fig. 3.9C-E) are relatively subtle. During acquisition, the shock activates the punishment-encoding charging DAN (see Fig. 3.9C),

which decreases the synaptic weights of the KC onto the attraction-driving restrained MBON (see Fig. E.7C). This cannot be seen in Fig. 3.9D because the restrained MBON is already strongly inhibited (by the avoidance-driving susceptible MBON). The low response of this MBON means that the opposing reward-encoding charging DAN is largely unaffected in this conditioning paradigm (see Fig. 3.9E). The non-zero activity level of this DAN is a consequence of the LTM microcircuit—described in the following section. The responses are similar for both odours because the circuit starts in a balanced state (no preference for either odour). The different responses to the two odours seen in the experimental data might therefore represent an unbalanced starting state of its LTM for these odours due to previous experiences of the fly. It is also possible that influence from other MB-related neurons might be in place, and that some missing component could allow for a better fit.

3.3.4 *Long-term memory*

Ichinose et al. (2015) described a microcircuit where a reward-encoding DAN (PAM-α1) potentiates the KC → MBON synapses of MBON-α1, and MBON-α1 in turn excites the PAM-α1. Data from Li et al. (2020) suggested numerous similar microcircuits; some of them involved the PAM-β′2a and PPL1-γ2α′1, which are the charging DANs of the RSM microcircuit introduced in the previous section. Specifically, the reward-encoding charging DAN (PAM-β′2a) can potentiate the response of the attraction-driving MBON-β2β′2a; and similarly, the punishment-encoding charging DAN (PPL1-γ2α′1) can potentiate the avoidance-driving MBON-α′1 (see Fig. 3.10A and C, while Fig. 3.10B shows these connections schematically with the KCs omitted for convenience). Crucially, these connections form positive feedback circuits; the DAN potentiates the response of the MBON to the odour, which increases its excitation of the DAN. As a consequence, even when the reinforcement ceases, the learning momentum can continue (saturation effect of the DPR—see Fig. 2.1D), resulting in *long-term memory* (LTM) consolidation and enhancement.

Fig. 3.10D—right demonstrates the charging of the avoidance-driving LTM MBON during memory acquisition (for odour B), and its continued increase during the forgetting phases. However, these trends are not evident in the experimental data as illustrated in Fig. 3.10D—left, which is an inaccuracy of the model. Alternatively, this might be because the responses of LTM neurons depend on the overall experience of the animal; thus they are hard to predict during one experiment. For example, it could be the case that the animal has already built some long-term avoidance memory for odour A, such that its presentation without reinforcement continues its learning momentum; this may lead to the observed increasing response. Note that the decreasing

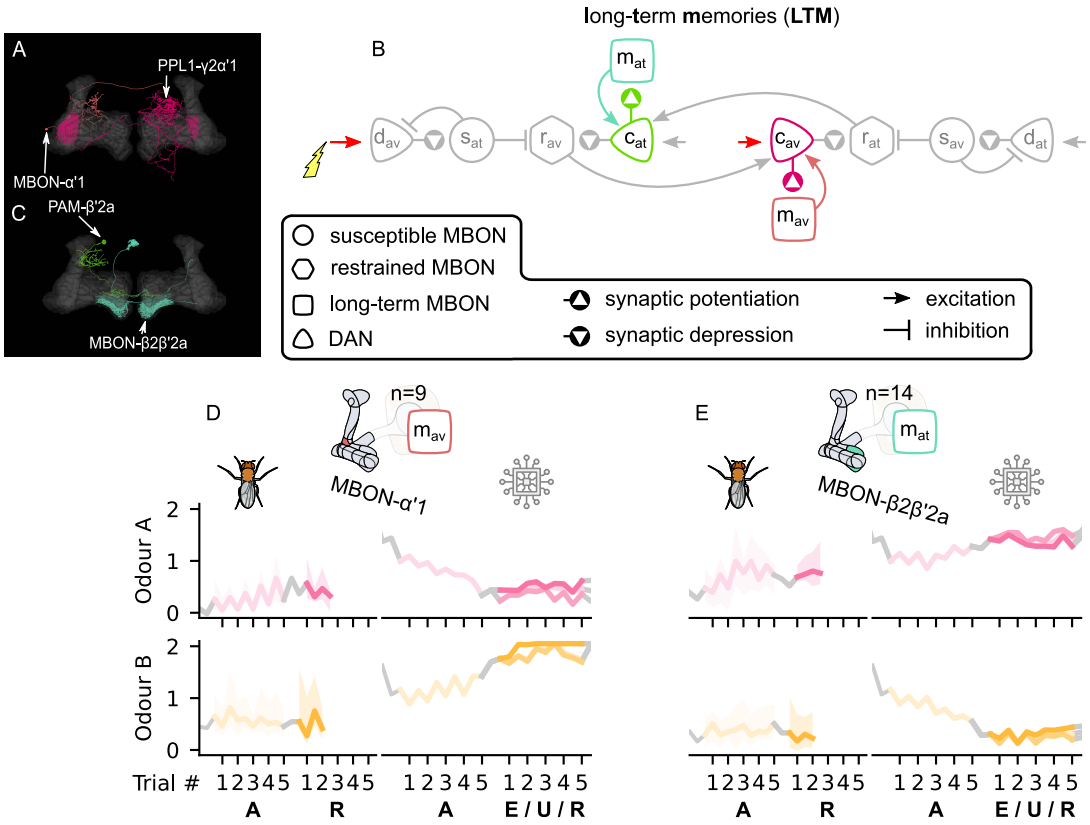

Figure 3.10: The *long-term memory* (LTM) microcircuits of the *mushroom body* (MB). (A) Image of the avoidance-encoding LTM microcircuit made of the MBON-α'1 and PPL1-γ2α'1—created using the Virtual Fly Brain software (Milyaev et al., 2012). (B) Schematic representation of the LTM microcircuits (coloured) connected to the *reciprocal short-term memory* (RSM) microcircuit via the charging *dopaminergic neurons* (DANs). (C) Image of the attraction-encoding LTM microcircuit made of the MBON-β2β'2a and PAM-β2a—created using the Virtual Fly Brain software (Milyaev et al., 2012). The responses of the (D) avoidance-driving LTM *mushroom body output neuron* (MBON), m_{av}, and (E) the attraction-driving LTM MBON, m_{at}, generated by experimental data (left) and the model (right) during the olfactory conditioning paradigms of Fig. 3.7D. The lightest shades denote the extinction, the mid shades the unpaired and the dark shades the reversal phase. For each trial two consecutive time-steps are reported: the off-shock (odour only) followed by the on-shock (paired odour and shock) when available (odour B in acquisition and odour A in reversal phase) otherwise a second off-shock time-step (all the other phases). Adapted from Gkanias, McCurdy, et al. (2022) under CC BY 4.0.

response to odour A during memory acquisition in the model and the observed effects for the attraction-driving LTM MBON (Fig. 3.10E) are due to influence from additional microcircuits (to be described in the following section). Fig. E.9 shows the responses of these neurons using only the microcircuits that have been introduced so far. In this case, the responses of both LTM MBONs saturate instantly, indicating that there must be another mechanism to regulate them in order to generate useful behaviour for the animal.

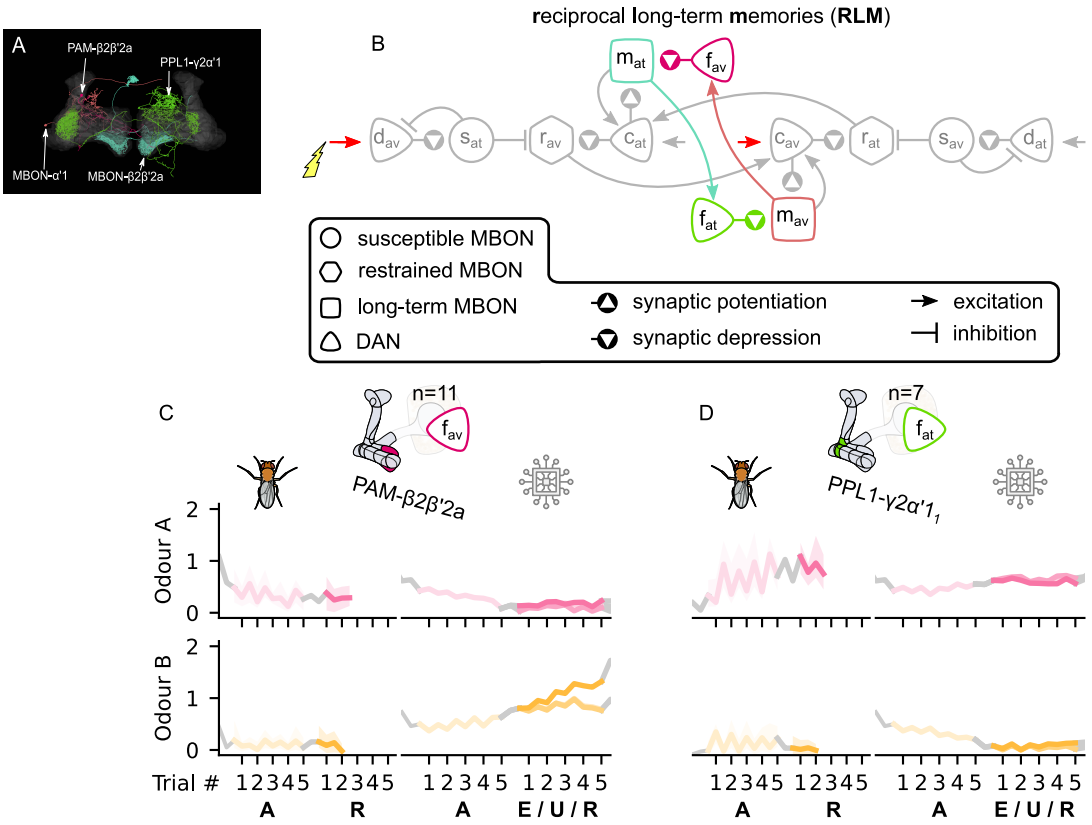

Figure 3.11: The *reciprocal long-term memories* (RLM) microcircuit of the *mushroom body* (MB). (A) Image of the RLM microcircuit in the MB made of the MBON-α′1, PAM-β2β′2a, MBON-β2β′2a and PPL1-γ2α′1—created using the Virtual Fly Brain software (Milyaev et al., 2012). (B) Schematic representation of the RLM microcircuit (coloured) The responses of (C) the punishment-encoding forgetting *dopaminergic neuron* (DAN), f_{av}, the (D) reward-encoding forgetting DAN, f_{at}, generated by experimental data (left) and the model (right) during the olfactory conditioning paradigms of Fig. 3.7D. The lightest shades denote the extinction, the mid shades the unpaired and the dark shades the reversal phase. For each trial two consecutive time-steps are reported: the off-shock (odour only) followed by the on-shock (paired odour and shock) when available (odour B in acquisition and odour A in reversal phase) otherwise a second off-shock time-step (all the other phases). Adapted from Gkanias, McCurdy, et al. (2022) under CC BY 4.0.

3.3.5 *Reciprocal long-term memories*

As described in the previous section, once the LTM microcircuit begins to charge, there is a self-sustaining increase in the synaptic weights during odour delivery, which prevents any subsequent adaptation to altered reward contingencies. To allow these weights to decrease in response to charging of the LTM of opposite valence, the two LTM MBONs must be connected via respective forgetting DANs (see Fig. 3.11B). Note, these forgetting DANs do not have to receive any direct reinforcement signal. Instead, as long as an LTM MBON is active, its respective forgetting DAN should be active, which should cause synaptic depression for the opposite LTM MBON (forget-

ting the learnt memory, as shown in Fig. 3.11C and D). This counteracts any potentiation effect due to the respective charging DAN of the LTM MBON (see Fig. E.12E and F). As a consequence, sustained reinforcement of one valence can gradually overcome the positive feedback of the LTM circuit of the opposite valence, and this can cause the charging momentum to slow down and eventually reverse (shifting to a discharging state). The LTMs are thus in long-term competition.

The above *reciprocal long-term memory* (RLM) microcircuit (shown in Fig. 3.11B) has been identified in the descriptions of Aso, Hattori, et al. (2014) and Li et al. (2020), where MBON-α′1 is the avoidance-driving LTM MBON, MBON-β2β′2a is the attraction-driving LTM MBON, PAM-β2β′2a is the avoidance-driving forgetting DAN, and PPL1-γ2α′1 is the attraction-driving forgetting DAN, as shown in Fig. 3.11A. One problem with this identification is that there is only one PPL1-γ2α′1 per hemisphere, and this has already been suggested to be the punishment-encoding charging DAN of the RSM and LTM microcircuits. However, there are multiple axon terminals of this neuron in the MB ($MB296B_1$ and $MB296B_2$) and each one of them seems to communicate a different response (see Fig. E.3—row 5, columns 6 and 7). Interestingly, the responses communicated by the $MB296B_1$ terminal are close to the ones produced by the punishment-encoding charging DAN (see Fig. 3.9C), and the ones of the $MB296B_2$ are close to the ones produced by the attraction-driving forgetting DAN (see Fig. 3.11D). This implies that different axons of the same DA neuron might create responses that depend on where the axon terminates, and actually work as separate processing units. Fig. 3.11C and D show that the reconstructed responses of these neurons from the modelled IC are surprisingly similar to the ones observed in the data.

3.3.6 *Memory assimilation mechanism*

The forgetting DANs allow the developing LTM of one valence to induce forgetting of the LTM of the opposite valence. However, the forgetting DANs can also be used for another critical function that maintains flexibility for future learning, which is to erase the memory of the same valence from their respective restrained MBONs. Thus, the forgetting DANs could also suppress the KC synaptic weights of their respective restrained MBONs, forming the *memory assimilation mechanism* (MAM) microcircuit (see Fig. 3.12C). This effectively allows memory transfer between the restrained and the LTM MBONs, enhancing both the adaptability and the capacity of the circuit. This effect can be observed in the difference of the responses of the same neurons in Fig. 3.10, Fig. 3.11, and Fig. E.11, where the restrained memory becomes weaker as the LTM becomes stronger (driven by the respective forgetting and charging DANs).

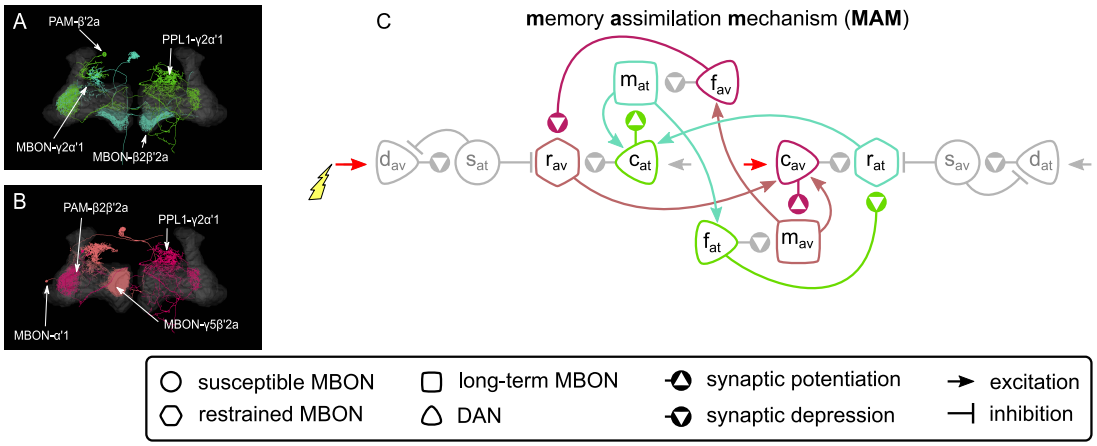

Figure 3.12: The *memory assimilation mechanism* (MAM) microcircuit of the mushroom body (MB). (A) Image of the avoidance-specific MAM microcircuit in the MB made of the MBON-γ5β′2a, PPL1-γ2α′1, MBON-α′1 and PAM-β2β′2a—created using the Virtual Fly Brain software (Milyaev et al., 2012). (B) Image of the attraction-specific MAM microcircuit in the MB made of the MBON-γ2α′1, PAM-β′2a, MBON-β2β′2a and PPL1-γ2α′1—created using the Virtual Fly Brain software (Milyaev et al., 2012). (C) Schematic representation of the MBM microcircuit (coloured). The forgetting *dopaminergic neurons* (DANs) connect to the restrained *mushroom body output neurons* (MBONs) of the same valence, hence increasing *long-term memory* (LTM) strength reduces (assimilates) the restrained memory, constituting the MAM microcircuits. Adapted from Gkanias, McCurdy, et al. (2022) under CC BY 4.0.

Anatomical data supports the notion that forgetting DANs depress KC → MBON synapses of the restrained MBONs of the same valence (Aso, Hattori, et al., 2014). More specifically, the avoidance-driving forgetting DAN (identified as PAM-β2β′2a) modulates the KC → MBON-γ5β′2a synapses, while the attraction-driving forgetting DAN (PPL1-γ2α′1) modulates the KC → MBON-γ2α′1 synapses—shown in Fig. 3.12A and B respectively.

3.4 DISCUSSION

The combination of the novel DPR and IC as a model of the MB was able to generate comparable neural responses to flies in associative learning and forgetting paradigms. All the hypothesised connections of the IC were identifiable in the brain of the fruit fly *D. melanogaster*, and it was suggested that (based on their functionality) at least three types of MB output (susceptible, restrained, and LTM), and three types of dopaminergic neurons (discharging, charging, and forgetting) exist in the circuit. This is shown to form a unified system for rapid memory acquisition and transfer from STM to LTM. The observed memory transfer mechanism could underlie the ability to make exploration/exploitation trade-offs, which will be explored in Chapter 4.

Other than the responses to electric shock, the model showed a relatively good fit to the data collected from neurons in the SM and RM microcircuits during olfactory conditioning. Although the responses of the attraction-driving restrained MBON ($\gamma2\alpha'1$) seemed inconsistent between the model and the data, the fit was actually not bad. The model predicted a drop in the (already very low) activity during acquisition followed by an increase during the reversal for both odours (see pre-training phase in Fig. 3.9D—right), which was also apparent in the data (however, note the very high variance in Fig. 3.9D—left). The overall responses of the punishing charging and forgetting DANs were also in accordance with the data, as opposed to their rewarding equivalents. Finally, the LTM MBONs did not show a particularly good fit to the data, although the attraction-driving one seems to fit less badly.

3.4.1 *Predictions of the computational model*

The model yields predictions that can be tested using established experiment protocols. MBON-$\gamma2\alpha'1$ and MBON-$\gamma5\beta'2a$ should exhibit STMs, while MBON-$\alpha'1$ and MBON-$\beta2\beta'2a$ LTMs. Also, MBON-$\gamma1pedc{>}\alpha/\beta$ and MBON-$\gamma4{>}\gamma1\gamma2$ should exhibit memories susceptible to extinction. Susceptible MBONs should show more consistent responses across flies. Restrained and LTM MBONs should have more variable responses because they are strongly affected by previous experiences of the animal. Activation of MBON-$\gamma2\alpha'1$ or MBON-$\beta2\beta'2a$ should increase the responses rate of PAM-$\beta'2a$, and similarly activating MBON-$\gamma5\beta'2a$ or MBON-$\alpha'1$ should excite PPL1-$\gamma2\alpha'1$. This would verify the excitatory STM reciprocal and LTM feedback connections of the circuit. Activating the LTM MBONs (MBON-$\alpha'1$ and MBON-$\beta2\beta'2a$) should also excite the forgetting DANs (PAM-$\beta2\beta'2a$ and PPL1-$\gamma2\alpha'1$ respectively). This would verify the excitatory LTM reciprocal connections of the circuit.

By consistently activating one of the LTM MBONs while delivering a specific odour, the LTM MBON should show an increased response to that odour even without the use of reinforcement. This would verify the saturation effect of the DPR and the charging momentum hypothesis. On the other hand, if we observe a reduced response rate, this would show that the MBON-DAN feedback connection is inhibitory and that RPE is most probably implemented by the circuit. Blocking the output of charging DANs (PPL1-$\gamma2\alpha'1$ and PAM-$\beta'2a$) could reduce the acquisition rate of LTM MBONs, while blocking the output of LTM MBONs would prevent memory consolidation. Blocking the reciprocal connections of the circuit should prevent generalising among opposing motivations (unable to make short- or long-term alterations of responses to odours once memories have formed). Blocking the output of forgetting DANs would additionally lead to hyper-saturation of LTMs, which could cause inflexible

behaviour. Activation of the forgetting DANs should depress the KC-MBON synaptic weights of the restrained and LTM MBONs of the same and opposite valence respectively, and as a result suppress their response to KC activation. Activation of the same DANs should cause increased activity of these MBONs for silenced KCs at the time.

3.4.2 *Additional mushroom body connections*

In the proposed model of the IC, only KC $\rightarrow$ MBON, MBON $\rightarrow$ DAN, and modulatory DAN $\multimap$ MBON connections were explored, all of which were suggested to be essential for successful learning in the MBs. However, there are a number of additional known connections in the MBs, such as excitatory KC $\rightarrow$ APL, inhibitory APL $\rightarrow$ KC, DAN $\rightarrow$ MBON, axo-axonic KC $\rightarrow$ KC, and KC $\rightarrow$ DAN connections, all of which have been neglected in this model and they need further consideration.

As mentioned in Section 3.2, there are two APL neurons in the larval brain, one for each MB. They extend their dendrites to the lobes of the MBs and terminate their axons in the calyces, releasing the inhibitory GABA neurotransmitter (Tanaka, Tanimoto, and K. Ito, 2008). Although there are still two of them in the adult brain, both their dendrites and axons innervate the calyx and the lobes (C.-L. Wu et al., 2013) suggesting that they function as both global and local inhibitory circuits. Moreover, mutual inhibitory DAN $\rightarrow$ APL (X. Liu and Davis, 2009) and APL $\rightarrow$ DAN connections have been proposed (Y. Wu et al., 2012), but there is no clear description of what their function is. Several previous models have demonstrated that a potential function for this global/local inhibition network is gain control, such that the total number of KCs firing to different stimuli remains similar (Peng and Chittka, 2017; Delahunt, Riffell, and Kutz, 2018); indeed, the same effect can be implemented using a flexible threshold for KC firing (Saumweber et al., 2018; Zhu, Mangan, and Webb, 2020; F. Zhao et al., 2020). In the description of the IC, the KC input was simplified, representing just two odours as different patterns across a small number of KCs (see Section 3.5). It was also assumed that a fixed number of KCs are active at all times, so the hypothesised gain control function of the APL was not useful here. However, it remains an interesting question whether there is learning between the KCs and APL in the lobes (Zhou et al., 2019), or between the APL and the KCs in the calyx, and what role this might play in the overall dynamics of memory acquisition.

Eichler et al. (2017) suggested that most of the KC inputs (60 %) are from other KCs. These connections (together with the ones from the APL) might create local WTA networks, which force a limited number of KCs per compartment to be active at one time. This predicts that it is possible for the same KC axon to be active in one

compartment but inactive in another, which is consistent with recent data from Bilz et al. (2020). It also predicts that an almost fixed number of KCs might be active at all times, even when no odour is delivered (for example, fresh air only), which enables acquisition and forgetting at all times. I. Ito et al. (2008) showed that KCs can be active even in the absence of odours (but with no consistent spiking), which is a characteristic of WTA networks when the underlying distribution of spikes across the neurons is almost uniform.

Eichler et al. (2017) also observed (from electron microscopy reconstruction in larvae) that within a compartment (in what they call a 'canonical microcircuit') KCs make direct synapses to the axons of DANs and DAN pre-synapses often simultaneously contact KCs and MBONs. The same connections have been observed in adult *D. melanogaster* by Takemura et al. (2017). The extent to which KCs (and thus odour inputs) might be directly exciting DANs remains unclear. Cervantes-Sandoval et al. (2017) showed that stimulating KCs resulted in increased DAN responses, and that DANs are activated through the ACh neurotransmitter. However, note that in the IC, such an effect could be explained without assuming a direct connection. For example, in the LTM microcircuit, activating the KCs resulted in increased activity of the LTM MBON, which excites the respective charging DAN. The DAN from which Cervantes-Sandoval et al. (2017) provided evidence was PPL1-$\alpha 2\alpha'2$, which is excited by MBON-$\alpha 2\alpha'2$ neurons that are characterised by the ACh neurotransmitter (Aso, Hattori, et al., 2014). In the context of the IC, this could be an LTM MBON that excites its respective charging DAN (PPL1-$\alpha 2\alpha'2$; Li et al., 2020); this would explain the presence of ACh at PPL1-$\alpha 2\alpha'2$. More generally, the altered activity of DANs in response to odours (which has been observed during learning) can be also observed in the IC without requiring direct KC $\rightarrow$ DAN connections or their modification. Nevertheless, such connections may possibly play a role in enhancing the specificity of dopamine-induced changes in KC $\rightarrow$ MBON connectivity. Interestingly, the depression of KC $\rightarrow$ DAN synapses (in parallel with KC $\rightarrow$ MBON synapses) could provide an alternative mechanism for implementing RPE induced plasticity (Takemura et al., 2017).

Takemura et al. (2017) demonstrated that the direct synapses observed from DANs to MBONs alter the MBON post-synaptic current to DAN activation and this alteration did not depend on KCs activity. This could be a mechanism by which learnt responses to reinforcements are coordinated with the current presence or absence of the reinforcement (Schleyer, Weiglein, et al., 2020; Schleyer, Saumweber, et al., 2011; Gerber and Hendel, 2006). Another possibility is that post-synaptic as well as pre-synaptic changes might be involved in learning at the KC $\rightarrow$ MBON synapse (Pribbenow et al., 2021).

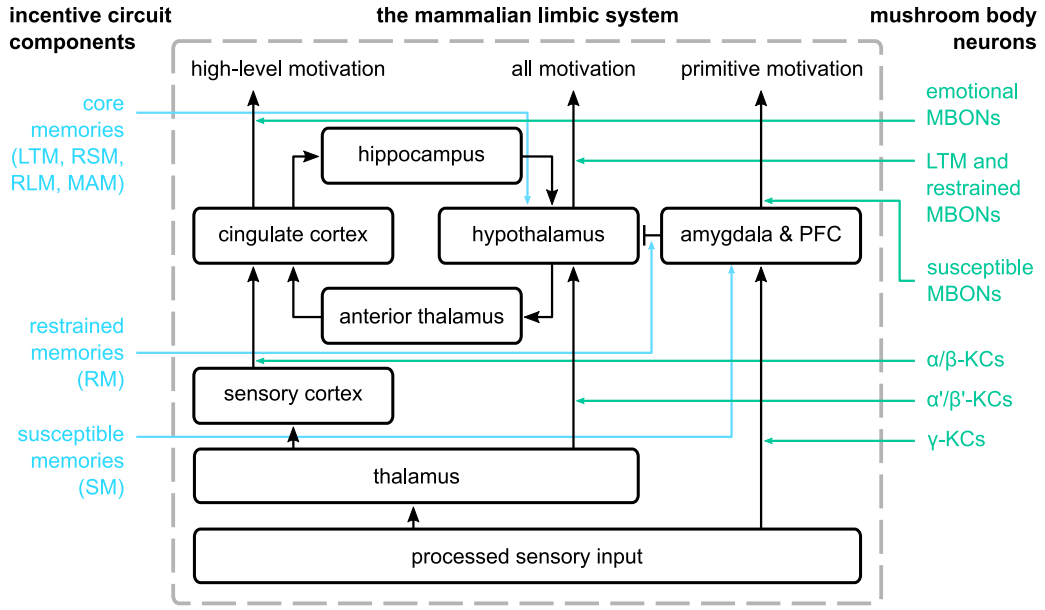

Figure 3.13: The mammalian limbic system as described by Papez (1937) and MacLean (1949), and the suggested parallels in the proposed *incentive circuit* (IC). On the left (in blue), there are the *mushroom body* (MB) microcircuits that correspond to the different structures in the mammalian limbic system. In the centre, there are connections among the different structures of the limbic system. On the right (in green), there are the groups of MB neurons that were suggested to have similar functions to the ones in the limbic system. Adapted from Gkanias, McCurdy, et al. (2022) under CC BY 4.0.

3.4.3 *Beyond attraction and aversion*

The IC consists of six MBONs and six DANs that link a pair of antagonistic motivations, attraction and avoidance. However, there are around thirty-four MBONs and one-hundred-and-thirty DANs in the MB of the adult fruit-fly brain, within which the IC is an identifiable motif. It is possible that this motif is repeated, representing additional opposing motivations. Also, some neurons in these repeated motifs might have multiple roles depending on the motivational context (as proposed by Cohn, Morantte, and Ruta, 2015). For example, a neuron could work simultaneously as a restrained and LTM MBON or as a discharging and forgetting DAN, depending on the identity of the reinforcer. Appendix A demonstrates this concept of a unified system of motivations as the *incentive wheel* (IW). This could explain how PAM-β2β′2a is a sugar-encoding discharging DAN in the appetitive olfactory conditioning context (MB301B; May et al., 2020), but it is also an avoidance-driving forgetting DAN in a different context (for example, aversive olfactory conditioning). In addition, two MBONs of the IC do not interact with the α′/β′ KCs of the MB. MBON-γ4>γ1γ2 and MBON-γ1pedc>α/β are part of two autonomous microcircuits (which are the SMs) and are working under the context provided by the almost six-hundred-and-seventy-

five γ-KCs relative to the task. This makes it possible that the KCs from the γ lobe connect to all the SMs of the flies for the eight available motivations illustrated in Fig. A.1.

From a functional point of view, the MBs seem to be involved in the motivation and behaviour of the animal, especially when it comes to behaviours essential for survival. In the mammalian brain, this function is subserved by the *limbic system*, which is composed of a set of complicated structures, such as the *thalamus*, *hypothalamus*, *hippocampus*, and *amygdala* (Dalgleish, 2004; Roxo et al., 2011). According to Papez (1937), sensory (mostly olfactory) input comes in the limbic system through the thalamus, which connects to both the *cingulate cortex* (through the sensory cortex) and the hypothalamus (Roxo et al., 2011; Dalgleish, 2004). Responses in the cingulate cortex guide emotions, while responses in the hypothalamus guide behaviour (bodily responses). Finally, the hypothalamus connects with the cingulate cortex through the *anterior thalamus* (forward), and the hippocampus (backward stream). MacLean (1949) augmented this model by adding the amygdala and *prefrontal cortex* (PFC) structures, which encode primitive emotions (for example, anger and fear), and connect to the hypothalamus (Roxo et al., 2011; Dalgleish, 2004). Fig. 3.13 suggests that some of the functions that have been identified in the IC could be mapped to limbic system structures.

More specifically, the α'/β'-KCs could have a similar role to the neurons in the thalamus, α/β-KCs represent a higher abstraction of the input stimuli and have a similar role to the ones in the sensory cortex, while the γ-KCs represent relatively unprocessed stimuli. The susceptible MBONs would then parallel neurons in the amygdala, which creates responses related to primitive motivations and connects to (inhibits) the restrained MBONs. The restrained MBONs would be analogous to the hypothalamus, providing primary behavioural control. As suggested previously, the same MBON could act as an LTM or restrained MBON in different ICs of an IW (see Appendix A); thus, the LTM could also correspond to the hypothalamus (with input from the α'/β'-KCs). Therefore, the RSM, RLM, LTM, and MAM microcircuits could correspond to hypothalamus functions. To continue the analogy, the function of the cingulate cortex is represented by the α/β MBONs, which suggests that they encode the 'emotions' of the animal towards reinforced stimuli and potentially control more sophisticated decision making. This mapping suggests that the connections among the restrained/LTM (α'/β') MBONs and the 'emotional' (α/β) MBONs are similar to the hippocampus and anterior thalamus pathways.

While it might seem startling to suggest that a compact circuit of single identified neurons in the insect MB mimics in miniature these far larger and more complex structures in the mammalian brain, the justification comes from the similarity in the

behavioural demands common to all animals: surviving and adapting in a changing world.

3.5 METHODS

The implementation of the IC as a computational model was based on firing-rate neurons and used simple ANN properties. Thus, the connections between neurons were represented by synaptic-weight matrices and the non-linear transformations of information passing from one neuron to another by activation functions. In the following sections, these parameters and relevant properties of the computational model are defined; they are consistent throughout all the experiments in this and the following chapters.

3.5.1 *Parameters of the model*

In the experiments done in this chapter and in Chapter 4, it was assumed that the odour identity passes through the PNs into the MB and its KCs. For now, the odour signal is assumed to be represented by $n_{\text{PN}} = 2$ PNs (one for each odour) and that these form distinct activations in a set of $n_{\text{KC}} = 10$ KCs in the MB (the subset of KCs that respond to the specific odours that were used in the experiments). Different representations were explored in Section 5.5.4 and 6.2.1. Therefore, the vector $\mathbf{p}_{\text{A}} = [1, 0]$ represented the activity of the PNs when odour A was detected, $\mathbf{p}_{\text{B}} = [0, 1]$ when odour B was detected, $\mathbf{p}_{\text{AB}} = [1, 1]$ when both odours were detected, and $\mathbf{p}_{\varnothing} = [0, 0]$ when none of them was detected. The responses of the KCs were calculated by,

$$\mathbf{k}(t) = \text{WTA}_{0.5}[\mathbf{p}^T(t) \cdot \mathbf{W}^{p2k} + \eta], \qquad \eta \sim \mathcal{N}(0, 0.001), \tag{3.1}$$

where η is Gaussian noise, $\mathbf{W}^{p2k} \in \mathbb{R}_{+}^{2\times 10}$ is the weight matrix that allows the transformation of the two-dimensional odour signal into the ten-dimensional KC responses, and t is the current time-step. The $\text{WTA}_{0.5}[x]$ is an activation function that keeps the top $50\,\%$ of KCs active based on the strength of their activity. Note that the number of neurons used for PNs and KCs is not very important, and any combination of PN and KC populations could be used. However, for larger KC populations, a lower percentage of KCs should be active at a given time. The PN $\rightarrow$ KC synaptic weights used here were,

$$\mathbf{W}^{p2k} = \begin{bmatrix} 0.8 & 0.8 & 0.8 & 0.8 & 0.8 & 0.8 & 0.8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.8 & 0.8 & 0.8 & 0.8 & 0.8 & 0.8 \end{bmatrix}. \tag{3.2}$$

The odours were represented by different firing patterns across 10 KCs: 4 fire only for odour A, 3 fire only for odour B, and the remaining 3 fire to either odour. This configuration was selected to show the effects of the DPR when there is an overlap in the KCs that respond to the two odours that were used in the conditioning paradigms. This assumption (overlap in responses) also created the best fit with the data, suggesting that there might be overlapping KCs encoding the real odours tested in the fly experiments.

The delivered reinforcement, $\mathbf{u}(t) \in \{0,1\}^2$, was transformed into an input for the DANs by using the weight matrix $\mathbf{W}^{u2d} \in \mathbb{R}_+^{2\times n_{\text{DAN}}}$. The activity of the DANs was represented by a six-dimensional vector, $\mathbf{d}(t) \in \mathbb{R}^6$, where each dimension represents a different neuron in the IC. Specifically,

$$\mathbf{d}(t) = \begin{bmatrix} d_{\text{at}}(t) & d_{\text{av}}(t) & c_{\text{at}}(t) & c_{\text{av}}(t) & f_{\text{at}}(t) & f_{\text{av}}(t) \end{bmatrix}. \tag{3.3}$$

The reinforcement was represented by a two-dimensional vector, $\mathbf{u}$, where the first dimension denoted rewarding signal, $\mathbf{u}_{\text{sugar}} = [1,0]$, and the second dimension denoted punishment, $\mathbf{u}_{\text{shock}} = [0,1]$. The contribution of this vector to the responses of the DANs was given by,

$$\mathbf{W}^{u2d} = \begin{bmatrix} 2 & 0 & 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 2 & 0 & 0 \end{bmatrix}. \tag{3.4}$$

In line with the DANs' vector representation, there was a similar vector for MBONs, $\mathbf{m}(t) \in \mathbb{R}^6$, where each dimension represented the response of a specific neuron in time t, as it is shown in the equation below,

$$\mathbf{m}(t) = \begin{bmatrix} s_{\text{at}}(t) & s_{\text{av}}(t) & r_{\text{at}}(t) & r_{\text{av}}(t) & m_{\text{at}}(t) & m_{\text{av}}(t) \end{bmatrix}. \tag{3.5}$$

The weight matrix that encodes the contribution of KCs to the MBON responses, $\mathbf{W}^{k2m}(t) \in \mathbb{R}^{10\times 6}$, was initialised as,

$$\mathbf{W}^{k2m}(t=0) = \mathbf{1}(n_{\text{KC}}, n_{\text{MBON}}) = \mathbf{1}(10,6), \tag{3.6}$$

which effectively is a 10×6 matrix of ones. In other words, all KCs connect to all MBONs, and their initial weight is positive and the same for all connections. As these are plastic weights, their value depends on the time step, and therefore the time variable, t, was provided as input. Note that also $w_{\text{rest}} = 1$, which initially results in the absence of memory—$w_{ij}^{k2m}(t=0) - w_{\text{rest}} = 0$. Thus, any deviation of the

synaptic weights from their resting value represents a stored memory with strength $v_{ij}^{\text{mem}}(t) = ||w_{ij}^{k2m}(t) - w_{\text{rest}}||$.

There were also MBON $\rightarrow$ DAN connections, $\mathbf{W}^{m2d} \in \mathbb{R}^{6\times6}$, and MBON $\rightarrow$ MBON connections, $\mathbf{W}^{m2m} \in \mathbb{R}^{6\times6}$, which were given by,

$$\mathbf{W}^{m2d} = \begin{bmatrix} 0 & -0.3 & 0 & 0 & 0 & 0 \\ -0.3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0.3 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0.3 & 0 & 0.5 \end{bmatrix}, \tag{3.7}$$

$$\mathbf{W}^{m2m} = \begin{bmatrix} 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{3.8}$$

The above matrices summarise the excitatory (positive) and inhibitory (negative) connections between MBONs and DANs (or other MBONs) as defined by the IC (see Fig. 3.6 and Fig. 3.14). The sign of the weights was fixed and based on the known anatomical properties of the MBs of *D. melanogaster*; the magnitude of the weights was hand-tuned in order to get the desired result. This was done by taking into account the constraint that equivalent types of connections should have the same weight (for example, in the reciprocal microcircuits). The magnitude of the synaptic weights specifies the effective strength of each of the described microcircuits in the overall circuit. There were also bias parameters to the responses of DANs, $\mathbf{b}^d$, and MBONs, $\mathbf{b}^m$, which were fixed as,

$$\mathbf{b}^d = \begin{bmatrix} -0.5 & -0.5 & -0.15 & -0.15 & -0.15 & -0.15 \end{bmatrix}, \tag{3.9}$$

$$\mathbf{b}^m = \begin{bmatrix} -2 & -2 & -0.5 & -0.5 & -0.5 & -0.5 \end{bmatrix}. \tag{3.10}$$

This bias can be interpreted as the resting value of the neurons, or some external input from other neurons that are not included in the model.

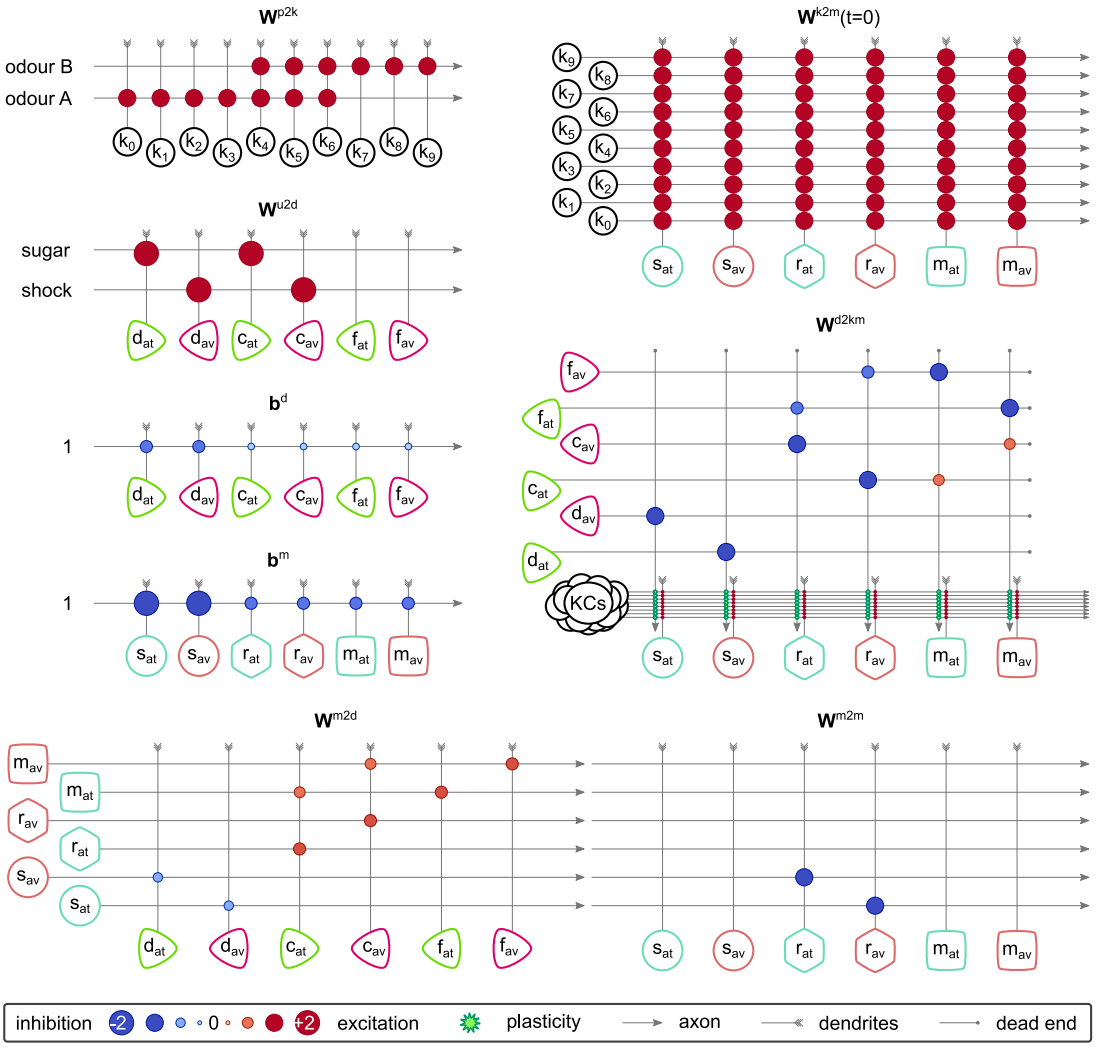

Figure 3.14: The synaptic weights and connections among the neurons of the *incentive circuit* (IC). Each panel corresponds to a different synaptic weights matrix of the circuit. The size of the circles when the pre-synaptic axon crosses the post-synaptic dendrite shows how strong a connection is, and the colour shows the sign (blue for inhibition, red for excitation). Light-green stars show where the synaptic plasticity takes place and how the *dopaminergic neurons* (DANs) modulate the synaptic weights between *Kenyon cells* (KCs) and *mushroom body output neurons* (MBONs). Adapted from Gkanias, McCurdy, et al. (2022) under CC BY 4.0.

Finally, the DAN-function matrix was defined, $\mathbf{W}^{d2km} \in \mathbb{R}^{n_{\text{DAN}} \times n_{\text{MBON}}}$, which transforms the responses of the DANs into the dopaminergic factor that modulates the $\mathbf{W}^{k2m}(t)$ synaptic weights, and it is given below,

$$\mathbf{W}^{d2km} = \begin{bmatrix} 0 & -1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0.3 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0.3 \\ 0 & 0 & -0.3 & 0 & 0 & -1 \\ 0 & 0 & 0 & -0.3 & -1 & 0 \end{bmatrix}. \tag{3.11}$$

All the parameters described above are illustrated in Fig. 3.14. Fig. E.16, E.17, and E.18 illustrate how each of these parameters affects the responses of the neurons in the IC. Finally, the activation function used in order to generate the DAN and MBON responses, is given by,

$$\varrho(x) = \begin{cases} 2, & \text{if} \quad x \geq 2, \\ x, & \text{if} \quad 0 < x < 2, \\ 0, & \text{if} \quad x \leq 0, \end{cases} \tag{3.12}$$

which is the *rectified linear unit* (ReLU) function, bounded in $\varrho(x) \in [0, 2]$. The reason for this bound was to avoid having extremely high values that grow exponentially during the charging of the LTM.

3.5.2 *Forward propagation*

For each discrete time step, t, the information from the environment was propagated through the model, and the responses of the neurons and the synaptic weights were updated. This process is called *forward propagation*, and it was repeated for all the iterations of the experiment.

First, the odour input, $\mathbf{p}(t)$ (which was encoded by the PNs), and the reinforcement, $\mathbf{u}(t)$, were read from the environment, and the KC responses were calculated using Eq. (3.1). To calculate the updates for the DANs and MBONs, the following equations were used,

$$\mathbf{d}(t) = \varrho\Big\{(1-\frac{\mathrm{d}t}{\tau})\,\mathbf{d}(t-\mathrm{d}t) + \frac{\mathrm{d}t}{\tau}\Big[\mathbf{u}^T(t)\cdot\mathbf{W}^{u2d} + \mathbf{m}^T(t-\mathrm{d}t)\cdot\mathbf{W}^{m2d} + \mathbf{b}^d\Big]\Big\}, \tag{3.13}$$

$$\mathbf{m}(t) = \varrho\Big\{(1-\frac{\mathrm{d}t}{\tau})\,\mathbf{m}(t-\mathrm{d}t) + \frac{\mathrm{d}t}{\tau}\Big[\mathbf{k}^T(t)\cdot\mathbf{W}^{k2m}(t-\mathrm{d}t) + \mathbf{m}^T(t-\mathrm{d}t)\cdot\mathbf{W}^{m2m} + \mathbf{b}^m\Big]\Big\}, \tag{3.14}$$

where $\mathrm{d}t = 1\,\mathrm{sec}$ is the duration of a time step, $\tau = 3\,\mathrm{sec}$ is a time-constant that is defined by the number of time-steps associated in each trial, and T denotes the transpose operation of the matrix or vector.

Finally, the dopaminergic factor, $\delta(t) \in \mathbb{R}^6$, was calculated, and the KC $\rightarrow$ MBON synaptic weights were updated as,

$$\delta(t) = \mathbf{d}^T(t) \cdot \mathbf{W}^{d2km}, \tag{3.15}$$

$$\mathbf{W}^{k2m}(t) = \max[\mathbf{W}^{k2m}(t - \mathrm{d}t) + \delta(t) * [\mathbf{k}^T(t) + \mathbf{W}^{k2m}(t + \mathrm{d}t) - w_{\text{rest}}], 0], \tag{3.16}$$

where '$*$' denotes element-wise multiplication and $w_{\text{rest}} = 1$ is the resting value of the weights. Element-wise multiplication is when each element of the $\delta(t)$ vector is multiplied with each column of the $\mathbf{W}^{k2m}(t)$ matrix. Element-wise addition of the transposed vector, $\mathbf{k}^T(t)$, to the $\mathbf{W}^{k2m}(t + \mathrm{d}t)$ matrix is when each element of $\mathbf{k}(t)$ is added to the corresponding row of $\mathbf{W}^{k2m}(t + \mathrm{d}t)$. The above procedure is repeated for all iterations of the simulation.

3.5.3 *Modelling the neural responses*

To emulate the acquisition and forgetting paradigms used for flies, the simulated circuit was run for $T = 73$ *time steps*. Each time step comprises 4 *iterations* of the forward propagation process described above to smooth out any bias due to the order of computations (value vs weights update). After an initial time-step at $t = 0$, there were 24 *trials* where each trial consisted of 3 time-steps.

Within each trial, the first time-step had no odour, while the odour was presented during the second and third time-steps: odour A on even trials and odour B on odd trials. A trial could have no shock (Fig. 3.15A), unpaired shock presented in the first time-step (Fig. 3.15B), or paired shock presented in the third time-step (Fig. 3.15C). The first 2 trials comprised the *pre-training phase*, where the model was exposed to the two odours alternately (odour A in trial 1 and odour B in trial 2) without shock delivery. The *acquisition phase* followed, where the shock was delivered with odour B for 10 trials (5 trials per odour; Fig. 3.15D). Before proceeding to the forgetting phases, 2 empty trials occurred (1 per odour), which are called the *resting trials*. The *forgetting phases* lasted for another 10 trials (5 trials per odour; Fig. 3.15E, F and G). There were three possible forms for the forgetting phase: in extinction, no shock was delivered while alternation of the odours continued (see Fig. 3.15E); in unpaired forgetting, the shock was delivered unpaired from odour A (see Fig. 3.15F); and in reversal, the shock was paired with odour A (Fig. 3.15G).

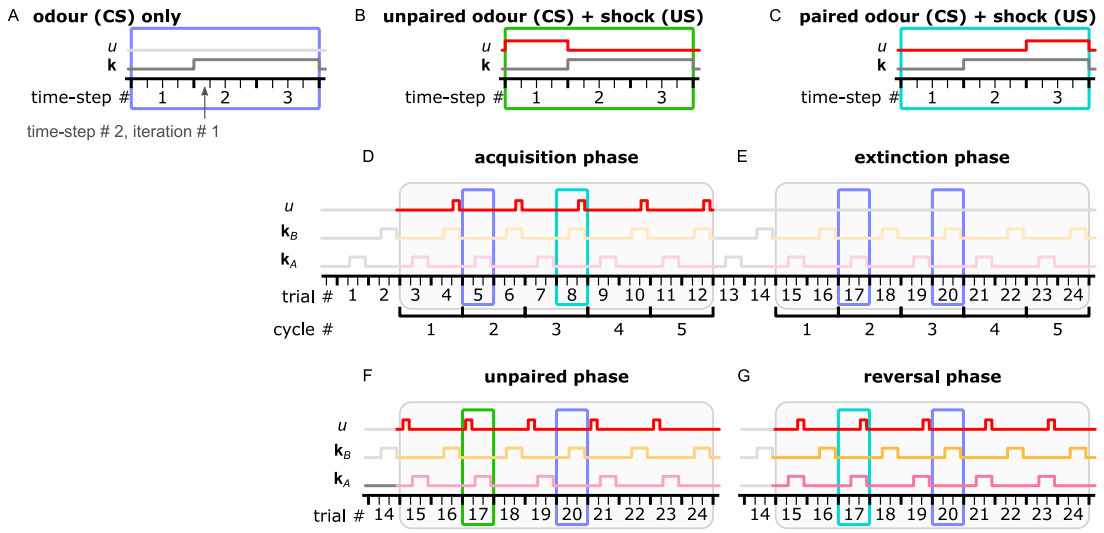

Figure 3.15: Description of the simulation process from the experiments. A single trial is composed of 3 time-steps and each time-step by 4 iterations. The odour was provided only during the 2nd and 3rd time-steps of each trial, while shock delivery was optional. (A) In an extinction trial (marked as blue), only odour (*conditioned stimulus*, CS) was delivered. (B) During an unpaired trial (marked as green), shock (*unconditioned stimulus*, US) was delivered during time-step 1. (C) During a paired trial (marked as cyan), shock is delivered along with odour delivery and during time-step 3. (D) The acquisition phase had 5 odour A only trials and 5 paired odour B trials alternating. (E) The extinction phase had 5 odour A only trials and 5 odour B only trials alternating. (F) The unpaired phase had 5 odour A unpaired trials and 5 odour B only trials alternating. (G) The reversal phase has 5 odour A paired trials and 5 odour B only trials alternating. The colours used in this figure for the CS and US match the ones in Fig. 3.7. The coloured boxes in D, E, F, and G provide examples of the odour-only (A; blue), unpaired odour+shock (B; green), and paired odour+shock (C; cyan). Note that each trial consists of 3 time-steps and 12 iterations in total. Adapted from Gkanias, McCurdy, et al. (2022) under CC BY 4.0.

The classic unpaired conditioning paradigm

In this case, during the acquisition phase, the shock was delivered in odd trials (omission of odour B), followed by an extinction phase as described above. Note that this method is relevant to Section 4.3.2.

3.5.4 *Validation of the dopaminergic factor*

In Eq. (3.15), the dopaminergic factor was derived from the matrix in Eq. (3.11), which captured the effects of dopamine release in an abstract and temporarily-independent form. For the experiments in Fig. 3.6 (and later in Fig. 4.3), this could be derived by

Eq. (2.17) and Eq. (2.18) with $\tau_{\text{short}} = \tau_{\text{long}} = \mathrm{d}t$, which removes the dynamics induced by the relation between $\mathbf{D}_{\triangledown}(t)$ and $\mathbf{D}_{\triangle}(t)$, and Eq. (3.15) could emerge from,

$$\delta(t) = \mathbf{D}^{\triangle}(t) - \mathbf{D}^{\triangledown}(t) \tag{3.17}$$

$$= \mathbf{d}^T(t) \cdot \mathbf{W}_{+}^{d2km} + \mathbf{d}^T(t) \cdot \mathbf{W}_{-}^{d2km}, \qquad \text{for } \tau_{\text{short}} = \tau_{\text{long}} = \mathrm{d}t \tag{3.18}$$

$$= \mathbf{d}^T(t) \cdot \mathbf{W}^{d2km}. \tag{3.19}$$

This is because each update represents a time-step that is longer than the effective period of backward conditioning for the responses shown in Section 3.3.1 (and later Section 4.2), where sampling frequency was low ($\leq 0.5\,\mathrm{Hz}$ and $1\,\mathrm{Hz}$ respectively), resulting in the ToToEq. (3.15).

3.5.5 *Data collection*

To verify the plausibility of the IC, McCurdy et al. (2021) recorded the neural activity in genetically targeted neurons during aversive olfactory conditioning. They simultaneously expressed the green GCaMP$_6$f Ca^{2+} indicator, and red Ca^{2+}-insensitive td-Tomato in neurons of interest to visualise the Ca^{2+} changes that reflect the neural activity. They collected data from 357 female flies (flies were 5- to 8-days-old; data from 2-14 flies per neuron), and for 43 neurons, which are illustrated in Fig. E.3.

Each fly was head-fixed for simultaneous delivery of odours and electric shock while recording the neural activity. Their proboscis was also glued, while their body and legs were free to move (see Fig. 3.7A). The flies were allowed to recover from the glueing process for $15\,\mathrm{min}$ before placing them under the microscope. McCurdy et al. (2021) used green ($555\,\mathrm{nm}$) and blue ($470\,\mathrm{nm}$) lights to record GCaMP and Tomato signals. They also used $0.1\,\%$ *3-octanol* (OCT3) and $0.1\,\%$ *4-methylcyclohexanol* (MCH4) for odour A and B respectively, and the flow rate was kept constant at $500\,\mathrm{mL/min}$ for each odour. The flies were allowed to acclimate to the airflow for at least one minute before starting the experiment.

During the experiments, trials were alternated, where $5\,\mathrm{sec}$ of each odour was presented $5\,\mathrm{sec}$ after the (green or red) light was on. It started with 2 pre-training trials (1 per odour), followed by 5 acquisition trials per odour. During acquisition, flies received alternating $5\,\mathrm{sec}$ pulses of OCT3 (odour A) and MCH4 (odour B) paired with electric shock, and repeated for 5 trials. During a reversal, OCT3 was presented with shock and MCH4 without, and this was repeated for 2 trials. On trials where the electric shock was delivered, it was presented $4\,\mathrm{sec}$ after odour onset for $100\,\mathrm{ms}$ at $120\,\mathrm{V}$.

The collection of these data was entirely done by McCurdy et al. (2021) and it was not part of this thesis.

3.5.6 *Calculating off- and on-shock values*

From the data collection process described above, trials of 100 time-steps were extracted at $5\,\mathrm{Hz}$ ($20\,\mathrm{sec}$ each). Odour was delivered between time-steps 25 and 50 (between $5\,\mathrm{sec}$ and $10\,\mathrm{sec}$); shock was delivered during time-step 45 (at $9\,\mathrm{sec}$). In Section 3.3, two values were reported for each trial: the *off-shock* and *on-shock*. These values represented the average response to the odour before and during the period in which shock delivery could have occurred (even if the shock was not delivered).

For the off-shock value, the values from time steps between 28 ($5.6\,\mathrm{sec}$) and 42 ($8.4\,\mathrm{sec}$) were collected from the recorded activities of each neuron. This resulted in a matrix of $n_{\mathrm{fly}} \times 15$ values, whose average and standard deviation was the reported off-shock value. Similarly, for the on-shock values, the values in time-steps between 44 ($8.6\,\mathrm{sec}$) and 48 ($9.6\,\mathrm{sec}$) were collected, which resulted in a matrix of $n_{\mathrm{fly}} \times 5$ values, whose average and standard deviation was the on-shock value. Therefore, the time window from $8.6\,\mathrm{sec}$ to $9.6\,\mathrm{sec}$ was defined as 'on-shock', where shock onset occurred at $t = 9\,\mathrm{sec}$.

4

OLFACTION AS A CASE STUDY

"Nothing awakens a reminiscence like an odour."

Victor Hugo

The insect *antennal lobes* (ALs) can represent billions of odourants (components of smells) with a limited number of neurons (Bargmann, 2006; Bear et al., 2016; Robertson, 2018). Compared to visual and auditory information, which are characterised by well-studied wavelengths, odours are characterised by chemical bindings which are less intuitive to humans. However, they seem to provide a rich-in-information and memorable representation of experiences, which forms the basis for many of the behaviours of animals. In fact, strong memories and behaviours in response to odours have been observed in most animals and this is likely related to the similar circuits that they inherit from a common ancestor (Naumann et al., 2015). In the insect brain, the understanding of odour processing is continuously increasing for the early (Bhandawat et al., 2007; Ng et al., 2002; Root et al., 2007; Schmuker, Bruyne, et al., 2007; Schmuker and Schneider, 2007; Seki et al., 2017; Wilson, Turner, and Laurent, 2004) and later stages of olfactory processing (Chakraborty and Sachse, 2021; Chakraborty, Chang, et al., 2022). Experiments related to olfaction have shown that the behaviour of insects can be affected by manipulating the odours in their environment and that this relates to memories stored in the *mushroom bodies* (MBs). This explains why olfactory conditioning is the most popular experimental paradigm for studying the memory mechanisms of the fruit fly *Drosophila melanogaster*. Although the representation of odours seems relatively consistent across the ALs of different flies, it has been noticed that the same experiment sometimes results in varying behaviours, which usually depend on the lab performing the experiment, but also on variations across the individual flies. One explanation could be that the complicated memory dynamics of the MBs—for example, due to the *incentive circuit* (IC)—are responsible for this variable outcome. To test this hypothesis, this chapter challenges the IC to explain experimental findings related to olfactory conditioning in fruit flies.

A variety of behavioural experiments were designed aiming to identify how insects acquire and retrieve memories; much effort has been put into trying to model these behaviours. The most common experimental scenarios in fruit flies involve olfactory *conditioning*. This is when the behaviour of the animal changes in response to one stimulus (for example, an odour) after this stimulus was experienced in association with another (for example, the same odour and electric shock). This section provides the necessary background knowledge on the different conditioning types and discusses their relation to olfactory conditioning experiments done with fruit flies.

4.1.1 *Conditioning types*

The process of pairing two initially independent stimuli in the animal's environment is called conditioning. This refers to the training procedure rather than learning itself. Two types of conditioning have been widely examined in psychology: *classical* (or *Pavlovian*) conditioning (Pavlov, 1949), and *operant* (or *instrumental*) conditioning (Skinner, 1938). Both of them involve a reinforcement signal and are based on the fact that the animal changes its behaviour due to the presence of that signal. More specifically, classical conditioning is considered as a prototypical instance of prediction learning; it assumes that a stimulus consistently paired with a reinforcer comes to predict the reinforcer and hence induces anticipatory action. Operant conditioning assumes that actions (usually in response to a stimulus) that are consistently followed by a reward will become more frequent, as opposed to actions followed by punishment (which will become less frequent).

To explain classical conditioning, let us assume that there are an animal and two stimuli that the animal is exposed to. The animal is assumed to start neutral (it does not respond) to the *conditioned stimulus* (CS). This does not mean that the *conditioned response* (CR) is zero, but instead that it is unrelated to the *unconditioned stimulus* (US), which consistently causes a response to the animal: the *unconditioned response* (UR). Classical conditioning can then be defined as the process of creating an association between the CS and the effect of the US when they are presented simultaneously, and usually multiple times (Rescorla and Wagner, 1972). Therefore, the animal will start responding to the CS, but this does not mean that the CR becomes the same as the UR (Rescorla, 1988). For example, the scent of home cooking (CS) usually predicts a tasty meal (US) and results in salivation (CR after pairing); thus, CR is the result of pairing the CS with the US and it is different from the UR (which would probably be the chewing response).

Skinner (1938) argued that classical conditioning is far too simplistic to explain most of the behaviours observed in animals. Therefore, he came up with another theory (operant conditioning) that is based on the *law of effects* (Thorndike, 1927). In principle, his theory and experiments suggested that a reinforced behaviour is more likely to be repeated, while this is not the case for the less reinforced behaviours. Skinner (1938) identified three types of operants: *neutral operants*, *reinforcers*, and *punishers*. Neutral operants do not affect the probability that a behaviour will be repeated. Reinforcers can be positive or negative, and they increase the probability that this behaviour will be repeated. Note that negative reinforcers also increase the probability of repeating the behaviour. They were defined by Skinner (1938) to predict the termination of an unpleasant stimulus, which differs from their definition in common *reinforcement learning* (RL)—where a negative reinforcer is equivalent to what Skinner (1938) called a punisher. Punishers are the opposite of reinforcers, weakening the probability of repetition for the behaviour by either applying an unpleasant stimulus or removing a pleasant one.

Most of the behavioural experiments done in fruit flies are based on classical conditioning paradigms. Skinner (1938) might have found classical conditioning less useful, but after the latest technological improvements that allowed scientists to read the neural responses from the brains of flies, classical conditioning became very popular in neuroscience. This usually requires the animals to live in a controlled environment and undergo experiments fixed-headed under a microscope. However, although in theory there is a clear difference between classical and operant conditioning, in practice they are not always easy to distinguish. In more natural conditions, the behaviour of the animals might affect their experience, which makes it harder to control for the pairing between the CS and US as required for classical conditioning. In this case, operant conditioning might provide better insights into the behaviour. This suggests that both conditioning types might exist in the background and that these are often hard to separate.

4.1.2 *Olfactory conditioning experimental paradigms*

Most of the olfactory conditioning experiments that have been done so far are based on elementary conditioning, where an animal has to choose between two odours (usually one is paired and one is not paired with the US), or is observed to learn to approach (or avoid) an odour that is positively or negatively paired with the US. Such experimental paradigms are usually accompanied by behavioural, electrophysiological, or imaging recordings aiming at the investigation of the brain regions, underlying circuitry, and genetic control involved in such associations (J. S. d. Belle and

Heisenberg, 1994; Dubnau, Grady, et al., 2001; Krashes, Keene, et al., 2007). These studies are nicely reviewed by Fiala (2007), Keene and Waddell (2007), and Vosshall (2007).

Studies in olfactory conditioning usually focus on lab experiments. This is because it is easier to isolate specific components (for example, odour components) or modalities in the lab, and consequently study them separately. However, the behaviours produced in the controlled environment of a laboratory may not represent fully the natural behaviour of the animal. In nature, animals are usually exposed to multiple odour components or modalities (for example, odour and vision) that affect their behaviour. There is a common assumption that learning two odours is equivalent to learning their joint components. This is called *configural learning* in vertebrate studies (Pearce, 1994) and it is not popular in insects (Giurfa, 2007).

J. Young et al. (2011) summarised a set of behavioural experiments aiming to identify more complex forms of olfactory conditioning in *D. melanogaster* including *elemental*, *two-element*, *mixture*, and *overlap* of odour components in olfactory conditioning. Table 4.1 summarises these types, while the following sections describe them in more detail and provide some results from the literature.

Elemental learning (*A+ B-*)

Elemental is an odour composed of only one component. There are two scenarios of *elemental learning*: with or without the involvement of a second odour. In both scenarios, the animal initially receives a reinforcement (US) while exposed to an odour A (CS), and this is usually noted as 'A+'. Then it is exposed to a different odour B without receiving any reinforcement, which is usually noted as 'B-'. This pair of exposures

Learning type	Training	Test
Elemental	A+ B-	A vs B
Multi-element	A+ B+ C-	A vs C, B vs C
Mixture	AB+ CD-	AB vs CD
Overlap	AB+ BC-	AB vs BC
Positive Patterning	AB+ A- B-	A vs AB, B vs AB
Negative Patterning	A+ B+ AB-	A vs AB, B vs AB
Biconditional discrimination	AB+ CD+ AC- BD-	AB vs AC, AB vs BD, CD vs AC, CD vs BD
Blocking	A+ AB+	B vs C

Table 4.1: The different types of olfactory conditioning in insects. '+' denotes reinforcement; '-' denotes no reinforcement; letters denote the different odour identities.

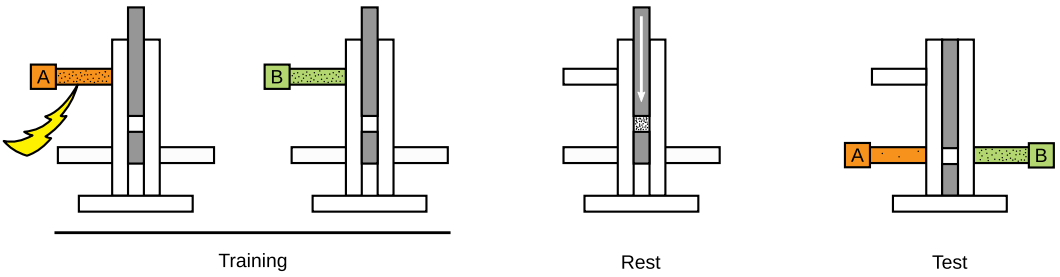

Figure 4.1: Elemental olfactory conditioning paradigm. T-maze is the most standard olfactory classical conditioning experiment in fruit flies (Tully and Quinn, 1985). During training, flies are placed in a tube with an odour (A, orange colour) and electric shock (thunder); then the odour is replaced with another odour (B, green colour) and the electric shock is omitted. This process is usually repeated multiple times (including reciprocal training). The flies are tested by providing both odours without electric shock, letting them choose the most attractive odour. Small dots in the T-maze represent the individual flies. This figure was not part of a simulation; adapted by permission from Springer Nature Customer Service Centre GmbH: Copyright © Vosshall (2007).

is called the *training phase* (seed Fig. 4.1–left) and it is usually repeated a number of times. Sometimes the second part of the training phase is replaced by *resting*, where the animal rests for an equal amount of time. This case is also considered elemental learning as the animal learns to associate only one elemental odour (CS) to the US.

After the training, the animals are moved to a neutral environment to rest before the *test phase* starts. During the test phase, the animals are free to choose between the two odours in the absence of any reinforcement (see Fig. 4.1–right). If a significant majority of the animals approach odour A, it means that this odour is preferred over odour B. In this case, odour A appears to have become more attractive than the neutral odour B, and thus the reinforcement (US) that was paired with it has served as a positive reinforcement. On the other hand, if the significant majority avoids the odour A, then it appears to have become aversive, and the US is concluded to act as a punishment for the animal.

Most of the time, the *reciprocal* condition is also applied, which aims to show that there is no innate bias in the odour preference. In this condition, the odours A and B swap, and the experiment is repeated for B+ and A-. This condition is not necessary in the case where only one odour is used, which makes this case preferable in recent studies. However, when only one odour is used, the experiment is usually accompanied by the *unpaired* condition. This is when the odour and the reinforcement are received alternately instead of together, to control for simple effects of exposure to odours and reinforcers. Quinn, W. A. Harris, and Benzer (1974) first demonstrated that fruit flies can be conditioned to avoid an odour selectively after being shocked

in its presence. J. Young et al. (2011) confirmed this result for six different pairs of odours.

Multi-element learning (*A+ B+ C-*)

An animal that can learn to recognise multiple elemental odours associated with reinforcement, is an animal with the ability of *multi-element learning*. In order to examine whether an animal has this ability, a similar paradigm to the process above is usually used. The difference is that the animal is exposed to a number of reinforced odours in a row (A+ then B+ and so on), followed by another odour with no reinforcement (C-). During the test phase, they are challenged to choose between the odours (choose between A and C, or between B and C). Multi-element learning paradigms are usually named after the number of reinforced elemental odours that the animal is exposed to. For example, in the above case, there are two reinforced odours, so this would be called a *two-element learning*.

In the demonstrations of Dudai (1977), fruit flies were able to learn to avoid selectively several odours at a time, and discriminated between different concentrations of the same odour. In addition, they seemed to remember these associations for hours but forgot them when sedated in the first $20\,\mathrm{min}$ after training. The concept of multi-element learning was also approached by Yin et al. (2009), who additionally showed that the intensity of the reinforcement matters for the choice of the flies between the odours. When the flies were exposed to two reinforced odours with different shock intensities and asked to choose between them afterwards, they approached the one paired with lower shock intensity. Finally, Yin et al. (2009) explored the effect of time on the decisions of flies, turning their focus on the sequence that the animal experienced each odour and reinforcement. They concluded that in multi-element learning more recent pairings of the CS with the US have stronger behavioural effects compared to older pairings (with a different CS and the same US), suggesting that associative memories might fade with time.

Mixture learning (*AB+ CD-*)

A more complicated condition occurs when more than one component is involved in each trial, which results in non-elemental odours (for example, presenting A and B simultaneously will result in the odour AB). The question here is whether the animal is able to distinguish between two non-elemental odours. If it can, then it is capable of *mixture learning*. The testing process of this learning type is very similar to the one of elemental learning. The animal is first trained in a reinforced non-elemental odour (AB+), followed by a non-reinforced non-elemental odour (CD-). This can be

repeated many times and separated by resting phases. In the test phase, the animal is exposed to the two different mixtures and its preference between them is observed.

Borst (1983) showed that flies can learn to discriminate a non-elemental odour from an odour composed of the same components but in different proportions. Flies could also discriminate a non-elemental odour from an elemental odour with just one of its components, which is a special case and is later treated as a separate condition named *positive patterning*. Interestingly, DasGupta and Waddell (2008) came to a similar conclusion, supporting that fruit flies can discriminate individual components (for example, A vs C) after training with non-elemental odours (AB+ and CD-).

Overlap learning (*AB+ BC-*)

A special case of mixture learning is the *overlap learning*. In this type of learning, the animal tries to distinguish non-elemental odours that share at least one component and have at least one component that is not common. The experimental setup is similar to the mixture learning, but the components B and D are identical (B = D). More specifically, the two non-elemental odours contain two components (AB and BC respectively), one of which exists in both mixtures (component B in this case). One of the odours is presented with a reinforcement (AB+) and the other one without it (BC-). In the test phase, the animal is asked to choose between the two combinations.

This problem is more challenging for the animal than simple mixture learning; testing both of them could reveal some very interesting characteristics of the animal's learning mechanism. For example, one could ask whether they treat a compound odour as a separate odour, or if the learned behaviour biases other compound odours that contain one of the learned components. J. Young et al. (2011) demonstrated that flies are able to discriminate overlapping odour mixtures. Despite the fact that overlapping odour mixtures are theoretically harder to discriminate than non-overlapping ones (as one of the components sometimes is reinforced and sometimes it is not), in practice, no significant difference in the ability to learn was observed in experiments with flies. This suggests that flies might ignore components that are involved in both reinforced and non-reinforced odours or assign a new code for each non-elemental odour instead of combining the codes based on their components.

Positive patterning (*AB+ A- B-*)

Animals that are able to associate a non-elemental odour with reinforcement and at the same time dissociate the individual components from it exhibit *positive patterning*. This condition is complementary to overlap learning and it is used in order to

show whether the animal can handle non-elemental odours as a whole or as separate elemental odours. In the training phase, the animal is exposed to a reinforced non-elemental odour (AB+), followed by exposure to the separate elemental odours of its components without reinforcement (A- B-). In the test phase, the animal is challenged to choose between the different pairs of non-elemental and elemental odours (involving a component of odour). If the animal can successfully discriminate between the non-elemental and elemental odours, it means that it can do positive patterning. This would suggest that either the addition of the two odours creates a stronger response than the individuals, or that it creates a new representation for the compound odour (which is a configural association).

It has been demonstrated that honeybees can process configural associations by differentiating reinforced and non-reinforced CS in positive patterning (Deisig et al., 2001; Komischke et al., 2003; Giurfa, 2007). Olfactory conditioning of the *proboscis extension response* (PER) has been used to study this effect. The bees were exposed to a non-elemental odour followed by sucrose and then to its separate elemental odours without sucrose; they seemed to learn not to extend their proboscis in the presence of the non-reinforced odours. Couvillon and Bitterman (1988) tried to challenge the configural association abilities of the honeybees by testing it in freely flying bees. In this case, the animals were trained in positive patterning and conditional problems (choosing between colours on the basis of a common odour, or between odours on the basis of a common colour). This study concluded that the non-elemental odours are represented by a novel neural code.

Negative patterning (*A+ B+ AB-*)

Negative patterning is a complementary condition to positive patterning and it tests the opposite effect (which is whether the animal can dissociate the reinforcement from the combination of two elemental odours after learning their individual association). Compared to positive patterning, this condition can more accurately demonstrate whether an animal can process configural association. The animal is trained to associate one odour with a reinforcement (A+), then another odour with the same reinforcement (B+), and then their mixture without a reinforcement. During the test, the animal is asked to choose between the individual odours and their combination. Note that this condition tests a very different property of the odour encoding, which could not be explained by the simple addition of the two odours (like in the positive patterning). The ability of the animals to solve this task would suggest a more complicated relationship between the odours should be in place (configural association). As it was mentioned before, honeybees can process configural associations (Deisig et al., 2001; Komischke et al., 2003; Giurfa et al., 2009) and as a result both positive

and negative patterning. Most interestingly, this is the only tested insect that could actually do any of them consistently (J. Young et al., 2011).

Biconditional discrimination (*AB+ CD+ AC- BD-*)

The above conditioning types involve a simple mixture of odours and comparison between their separate components. The *biconditional discrimination* is probably the most complicated condition found in the literature, which tries to examine what is the role of the individual components of the compound stimuli to other compound stimuli. These stimuli are usually pairs of odours and light, and they involve multiple modalities (multimodal). The animal is sequentially exposed to two compound stimuli with a reinforcement (AB+ CD+), followed by exposure in another two combinations of the same components but in a different order and without a reinforcement (AC- BD-). During the test phase, the animal is asked to choose between any reinforced or non-reinforced non-elemental odours (AB vs AC, AB vs BD, CD vs AC, and CD vs BD).

Brembs and Wiener (2006) used a flight simulator to train adult fruit flies to discriminate specific heat-reinforced pairs of colours and visual patterns in a variety of conditions. Their experiments showed that flies are able to discriminate against biconditional conditions. They also suggested that the function of MBs includes the maintenance of visual memories over 'context' changes and it is not important for cognitive-like higher-order learning. By 'context' (or 'occasion setting') they referred to the background colour of the pattern, which they suggested determines whether the pattern is paired with heat. Schubert et al. (2002) demonstrated that honeybees can also solve biconditional discrimination problems (as well as positive and negative patterning) in an appetitive context of visual compound stimuli. The bees were trained to discriminate pairs of colours and line orientations; their results suggested that bees choose the reinforced stimulus no matter its complexity. Crickets (Matsumoto and Mizunami, 2004) and cockroaches (Sato et al., 2006) could solve the same problem but in a different paradigm, where they were tested with reinforced odours given antagonistic light and dark conditions. On the other hand, the fruit fly larvae (Yarali, Hendel, and Gerber, 2006) and adults (Yarali, Mayerle, et al., 2008) showed no difference in the behaviour when exposed to shock-reinforced odours in light or dark; this suggested that fruit flies do not have the ability for biconditional discrimination. This was in line with findings of J. Young et al. (2011) (who did not observe biconditional discrimination in adult fruit flies when all the stimuli were odours) and in contrast to the results of Brembs and Wiener (2006) (who suggested that flies might express biconditional discrimination conditioned on the type of the reinforcement signal).

Blocking (*A+ AB+*)

In a blocking paradigm, the animal is trained on a reinforced elemental odour (A+), and later with a reinforced non-elemental odour that includes the component of the first odour (AB+). In the test phase, the animal is challenged to choose between the elemental odour B (a component of the second odour but not of the first) and a novel elemental odour C, which was not experienced before. If the animal shows a preference between odours B and C (B when rewarded or C when punished), then learning of the association between the first elemental odour and the reinforcement did not block the learning of the second (non-elemental odour). If the response to odour B appears unchanged, the animal is said to exhibit *blocking*. Blocking implies that it is redundant to learn about the non-elemental odour as learning involving the elemental odour (for example only A) is sufficient to produce the appropriate response to the non-elemental odour. This also relates to the predicted reinforcement hypothesis, which assumes that the animal tries to store the least information possible that predicts the US.

The blocking condition has been tested in both fruit flies (Brembs and Heisenberg, 2001) and honeybees (B. H. Smith and Cobey, 1994; Gerber and Ullrich, 1999; Couvillon, Campos, et al., 2001; Guerrieri et al., 2005); both failed to report successful results. In general, all evidence for blocking in insects has been debatable, suggesting that it is not a robust occurrence. More specifically, Hosler and B. H. Smith (2000) suggested that blocking in honeybees depends on odour similarity, but the provided evidence was weak. Later, Guerrieri et al. (2005) disputed them by proving the opposite (it is independent of odour similarity). They also showed that four out of twenty-four sets of odours provoked the blocking effect in honeybees, but none of them provoked it when exposed multiple times to the elemental odour. In alignment with this, Blaser, Couvillon, and Bitterman (2008) found no evidence of blocking in free-flying honeybees, but they found clear evidence of facilitation. More specifically, training in a blocking paradigm not only failed to reduce response to the second elemental odour but actually increased it.

4.1.3 *Neural activity interventions*

Recent technological advances in the genetic manipulation of fruit flies allowed the intervention in the responses of targeted neurons. This manipulation is usually independent of their sensory input and of any input known to work as a reinforcement (for example, electric shock). Genetic tools that can be controlled by temperature can either silence (for example, by expressing the *shibire* gene) or activate specific neu-

rons in the fruit-fly brain—for example, by activating the *transient receptor potential A1* (dTrpA1) ion channel. This allows for all the olfactory conditioning paradigms described in the previous sections to be manipulated, which could improve our understanding of the function of targeted *mushroom body output neurons* (MBONs) and *dopaminergic neurons* (DANs).

For example, when flies were exposed to an odour, activating avoidance-driving MBONs (Ichinose et al., 2015) or reward-encoding DANs (Burke et al., 2012; C. Liu et al., 2012; Lin et al., 2014; Huetteroth et al., 2015; Aso and Rubin, 2016) made the odour more attractive, which suggested that forcing avoidance may rebound attraction. This was further supported by a similar effect when silencing punishment-encoding DANs (Yamagata et al., 2016). However, activating punishment-encoding DAN created ambiguous memory dynamics. Although usually the odour became less attractive (Aso, Siwanowicz, et al., 2010; C. Liu et al., 2012; Yamagata et al., 2016), sometimes it became more attractive (Aso and Rubin, 2016), and sometimes there was no clear effect (Claridge-Chang et al., 2009; Aso, Sitaraman, et al., 2014).

The results became even more interesting when genetic intervention was mixed with natural reinforcements (like sugar or electric shock) in elemental learning scenarios. Collectively, these results suggested that silencing the MB extrinsic neurons resulted in relatively stable memory dynamics while activating them caused more unpredicted learning effects (Plaçais et al., 2013; Perisse et al., 2016). More specifically, when any MBON or DAN was silenced during the training phase or throughout the whole experiment, punishing odours became more attractive and rewarding odours less attractive (Aso, Siwanowicz, et al., 2010; C. Liu et al., 2012; Burke et al., 2012; Plaçais et al., 2013; Aso, Sitaraman, et al., 2014; Owald et al., 2015; Huetteroth et al., 2015; Ichinose et al., 2015; Yamagata et al., 2016). Similar results were observed when silencing any MBON during the test phase (Plaçais et al., 2013; Owald et al., 2015; Ichinose et al., 2015; Perisse et al., 2016; Felsenberg, Barnstedt, et al., 2017). These suggested that silencing any of the extrinsic MB neurons negates all the learning effects. However, there were some results suggesting the opposite for the training phase (the learning effects were unaltered when MBONs or DANs were silenced; Aso, Siwanowicz, et al., 2010; Lin et al., 2014; Yamagata et al., 2016) and for the test phase when DANs were silenced (C. Liu et al., 2012; Lin et al., 2014; Ichinose et al., 2015; Perisse et al., 2016; Felsenberg, Barnstedt, et al., 2017).

4.2 RESULTS

In the IC, three types of MBONs drive attraction and three avoidance. This results in six driving forces, for each available odour (see Fig. 4.2). A simple behavioural

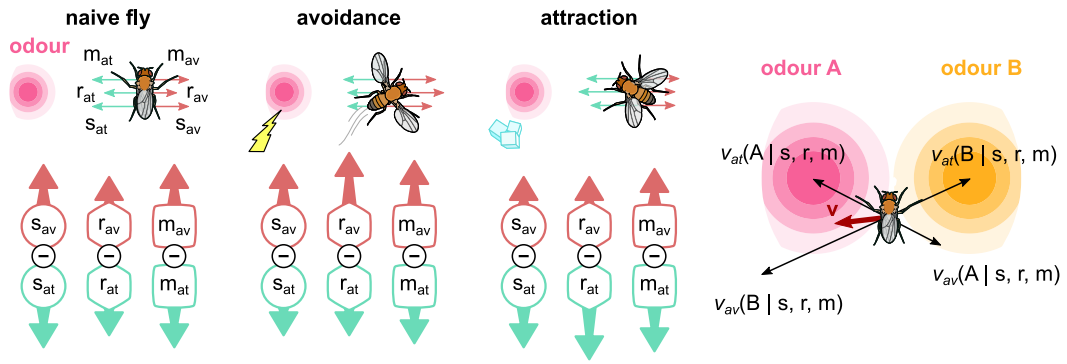

Figure 4.2: The activity of the six *mushroom body output neurons* (MBONs) was translated into forces that drive a simulated fly towards or away from odour sources. For naive flies, the forces are balanced. When an electric shock is paired with an odour, the balance changes towards the avoidance-driving MBONs, which drive the fly directly away from that odour. When sugar is paired with an odour, the balance changes to attraction, driving the fly towards that odour. Combining all attractive and repulsive forces for each odour source currently experienced by the fly produces an overall driving force, $\mathbf{v}$, which determines the behaviour of the fly. Adapted from Gkanias, McCurdy, et al. (2022) under CC BY 4.0.

readout (used in many previous models) would be to take the sum of all attractive and aversive forces at some time point as a measure of the probability of animals choosing odour A or B. This can then be compared to the standard two-arm maze choice assay used in many *D. melanogaster* studies (and it is illustrated in Fig. 4.1). As mentioned in the previous section, classical conditioning might not allow for the exploration of the full dynamics of the circuit. This is because animals live in an uncontrollable world, where they simultaneously explore, learn, express learning, and forget. Therefore, this section challenges the IC to explain the observed behaviour of fruit flies. For consistency and in order to allow for direct comparison to the results of Chapter 3, the same structure and parameters were used for the model.

4.2.1 *T-maze experiments*

As introduced in Section 4.1.2, in order to describe the cognitive capabilities of fruit flies, there are a number of olfactory conditioning paradigms by which they are usually tested. These are the elemental, multi-element, mixture, and overlap learning, positive and negative patterning, biconditional discrimination, and blocking. The behaviour extracted by the IC for the same paradigms was tested here, showing similarities to the summarised behaviours of fruit flies. The results suggested that the odour identity—as encoded by the *Kenyon cells* (KCs)—and the order that the odours were presented played an important role in the behaviour. Fig. 4.3 summarises the results of these experiments, while Table 4.2 reports the mean *preference indices* (PIs) for each

experiment and their individual setups after running the training phase for one, five, or ten cycles.

For all the T-maze experiments, a similar experimental setup to the one used in Chapter 3 was also used here, slightly modified to match the setup of J. Young et al. (2011). All the experiments had a training and test phase. The training phase was split into up to ten cycles. In each cycle, two to four odours were presented in a sequence and paired with reinforcement where appropriate (see Table 4.2—train). The testing phase followed, where two odours (without reinforcement) were presented in a sequence and in a single cycle (see Table 4.2—test). The input to the model was also slightly different (compared to Chapter 3). Each of the four available odours was associated with a single *projection neuron* (PN) and each PN was connected to six (out of twenty) KCs, enforcing one or two KCs overlap between two odours. Thus by default, odour A overlaps with odour B, odour B with odours A and C, odour C with odours B and D and odour D with odour C. Random noise was also added in the PN responses to enforce diversity among the simulated flies, and the five KCs with the highest response were allowed to be active.

First, the model was tested in the elemental and two-element paradigms. These showed a strong preference for non-punished odours from the first training cycle, which is in line with experimental data (J. Young et al., 2011). The preference was increased to its maximum after three training cycles (PI $= 0.59$, SD < 0.01; and PI $= 0.46$, SD < 0.1; for elemental and two-element learning respectively). These results verified that the model could learn useful information about odours, demonstrating classical conditioning. In elemental learning, the identity of the punished and non-punished odours did not affect the results, while in multi-element learning it did. This is consistent with experimental results showing that more recent pairings of an odour with the US show stronger behavioural effects compared to older ones (Yin et al., 2009). Mixture and overlapping odours suggested similar learning curves, with smaller but significantly positive PIs. This is also in line with fly data, suggesting that flies can discriminate mixture and overlapping odours, but less effectively than they do in the single and multi-element paradigms (J. Young et al., 2011).

Opposed to data suggesting that flies can occasionally do positive but they cannot do negative patterning (J. Young et al., 2011), the IC predicted against both of them (see Fig. 4.3). However, in positive patterning, the IC predicted that the order that the elemental odours (not the non-elemental odour) were presented affected the sign of the PI consistently. This is apparent in Table 4.2, by comparing the PIs of the different tests for the positive patterning when odour A was presented before odour B and for ten cycles (AB+ A- B-, A vs AB, PI $= -0.19$, SD < 0.01; AB+ A- B-, B vs AB, PI $= 0.12$, SD < 0.01); this was also consistent in the reciprocal condition. This sug-

Experiment	Train	Test	n	PI		
				#cycles = 1	5	10
Elemental			200	0.11**	0.59**	0.59**
	A+ B-	A vs B	100	0.11**	0.59**	0.59**
	B+ A-	A vs B	100	0.11**	0.59**	0.59**
Two-element			400	0.11*	0.46	0.46
	A+ B+ C-	A vs C	100	0.01**	0.40**	0.40**
	A+ B+ C-	B vs C	100	0.15**	0.55**	0.55**
	B+ A+ C-	B vs C	100	0.07**	0.31*	0.31*
	B+ A+ C-	A vs C	100	0.23**	0.59**	0.59**
Mixture			400	0.01**	0.32**	0.44
	AB+ CD-	AB vs CD	100	0.01**	0.34**	0.34**
	CD+ AB-	AB vs CD	100	0.02**	0.33**	0.33**
	AC+ DB-	AC vs DB	100	0.00**	0.31**	0.52*
	DB+ AC-	AC vs DB	100	0.00**	0.31**	0.55*
Overlap			600	-0.03**	0.26*	0.30*
	AB+ BC-	AB vs BC	100	-0.05**	0.30**	0.30**
	BC+ AB-	AB vs BC	100	-0.05**	0.30**	0.30**
	AC+ BC-	AC vs BC	100	-0.02**	0.24**	0.36*
	BC+ AC-	BC vs AC	100	-0.02**	0.24**	0.23**
	AB+ AC-	AB vs AC	100	-0.02**	0.24**	0.22**
	AC+ AB-	AB vs AC	100	-0.02**	0.26**	0.35*
Positive patterning			400	0.00*	-0.01	-0.04
	AB+ A- B-	A vs AB	100	-0.01**	-0.18**	-0.19**
	AB+ A- B-	B vs AB	100	0.02**	0.15**	0.13**
	AB+ B- A-	A vs AB	100	0.02**	0.14**	0.12**
	AB+ B- A-	B vs AB	100	-0.01**	-0.18**	-0.20**
Negative patterning			400	0.04*	0.07*	0.08*
	A+ B+ AB-	A vs AB	100	0.00**	0.16**	0.14*
	A+ B+ AB-	B vs AB	100	0.08**	-0.03**	0.01**
	B+ A+ AB-	A vs AB	100	0.08**	-0.02**	0.02**
	B+ A+ AB-	B vs AB	100	-0.00**	0.17**	0.12*
Biconditional discrimination			400	0.02*	-0.14*	-0.14*
	AB+ CD+ AC- BD-	CD vs AC	100	0.04**	-0.10**	-0.10**
	AB+ CD+ AC- BD-	CD vs BD	100	0.04**	-0.07**	-0.07**
	AC+ BD+ AB- CD-	AB vs AC	100	-0.01**	-0.18**	-0.21*
	AC+ BD+ AB- CD-	AB vs BD	100	0.00**	-0.22**	-0.19*
Blocking			200	0.14**	0.26*	0.23**
	A+ AB+	B vs C	100	0.14**	0.29*	0.23**
	B+ AB+	A vs C	100	0.15**	0.24**	0.23**
Control blocking			200	0.06**	0.21*	0.16*
	A- AB+	B vs C	100	0.05**	0.17**	0.08*
	B- AB+	A vs C	100	0.07**	0.24**	0.23*

Table 4.2: *Preference indices* (PIs) of the individual T-maze experiments run for simulated flies. A positive PI shows that the tested fly did not prefer the punished odour. n: the sample size. #cycles: the number of times that the training phase was repeated before testing. *: SD < 0.1, **: SD < 0.01. SD: *standard deviation*.

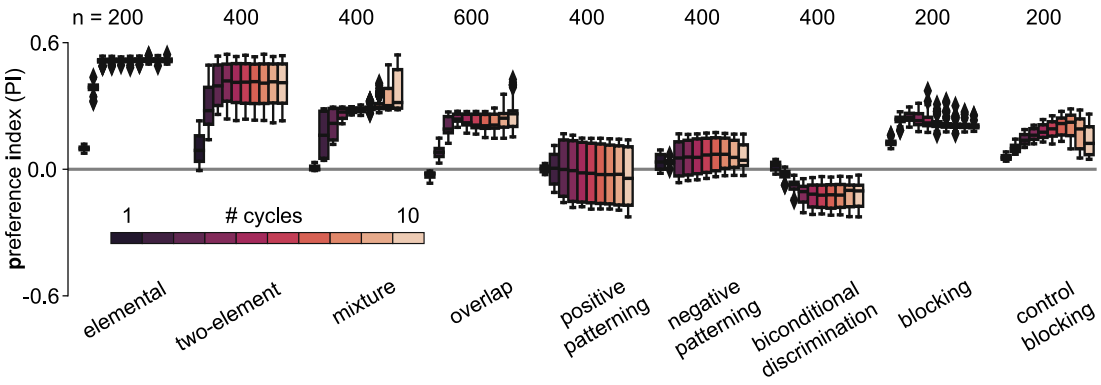

Figure 4.3: Testing the performance of the *incentive circuit* (IC) on the T-maze experiment (Quinn, W. A. Harris, and Benzer, 1974). The circuit was run for all the olfactory conditioning experimental paradigms described in Section 4.1.2 and its *preference index* (PI) was reported after repeating the training for 1-10 cycles. A positive PI indicates that the fly prefers the non-punished odour, while a negative PI shows the opposite. A zero PI indicates no preference.

gested that although the pooled results predicted against positive patterning for the IC, the case-by-case results predicted a bimodal distribution of the PIs; explaining the ambiguous results in flies. Similarly to the multi-element paradigm, the order that the odours were presented affected the result by making the more recently experienced (non-reinforced) odour more attractive compared to the previous odour. Therefore, their combination was more attractive than the oldest presented odour alone but less attractive than the more recent one. Note that Table 4.2 also suggested a similar effect for negative patterning, but it was relatively weak compared to positive patterning. In addition, the IC predicted against biconditional discrimination (see Fig. 4.3 and Table 4.2), which is also consistent with data from fruit flies (J. Young et al., 2011). During the first training cycle, the model showed no preference between the punished and non-punished odour codes. After the second training cycle, the preference (slightly but consistently) shifted towards the punished combinations of odours. This effect was probably due to the way odours were encoded in the KCs and the order that they were presented, as this effect was apparently stronger in some cases than others (see Table 4.2). J. Young et al. (2011) also (weakly) observed this effect.

Finally, in line with fruit-fly data, the IC did not predict blocking. From the first training cycle, the model showed a preference for the novel odour over the one paired with punishment (see Fig. 4.3). A similar effect was observed in the control blocking paradigm (see Table 4.2), where the single-component odour was not paired with punishment, and it should not cause blocking. This is consistent with observations in flies (J. Young et al., 2011). Interestingly, the gradual increase in the preference for the novel odour shown in Fig. 4.3 (both in blocking and control) was also observed in

flies (Guerrieri et al., 2005) and honeybees (Blaser, Couvillon, and Bitterman, 2008), where it was mentioned as 'facilitation'.

4.2.2 *Neural activity intervention experiments*

Bennett, Philippides, and Nowotny (2021) summarised the data of 92 olfactory conditioning intervention experiments from 14 studies (Felsenberg, Barnstedt, et al., 2017; Perisse et al., 2016; Aso and Rubin, 2016; Yamagata et al., 2016; Ichinose et al., 2015; Huetteroth et al., 2015; Owald et al., 2015; Aso, Sitaraman, et al., 2014; Lin et al., 2014; Plaçais et al., 2013; Burke et al., 2012; C. Liu et al., 2012; Aso, Siwanowicz, et al., 2010; Claridge-Chang et al., 2009). From these experiments (summarised in Section 4.1.3 and Fig. 4.4A), they extracted the observed learning effects of silencing or activating specific neurons (Δ_f). They showed that the observed behaviour extracted by the fly data strongly correlated with the simulated behaviour of their proposed models of the MB—*valence-specific* (VSλ) ($r = 0.68$, $p < 10^{-4}$) and *mixed-valence* (MV) ($r = 0.65$, $p < 10^{-4}$).

Following the same approach, the behaviour predicted by the IC correlated with the data slightly better ($r = 0.76$, $p = 2.2 \times 10^{-18}$) than the one reported by Bennett, Philippides, and Nowotny (2021). To obtain these results, the IC model was unaltered from what was described in Section 3.5, while the experiment was modified to match the description of Fig. 4.4A (in a similar way to what was done in the previous section). In the experimental data, the description of the targeted neurons for intervention was not always clear, which introduced some ambiguity on which of the three types of neurons to intervene in the IC. To deal with this ambiguity, the *selected neuron types* method was used, which found the best candidate for intervention in the model by cross-matching the details of the intervened neurons in the experiment and the ones in the model. For reference, the best possible performance of the IC was also calculated by using the *best fit* method. This selects the candidate neurons by comparing the correlation among all the possible combinations, constrained by the assumed valence of the intervened neuron (punishing or rewarding for DANs and avoidance- or attraction-driving for MBONs). The highest correlation coefficient that the IC could achieve was $r = 0.77$, $p = 1.65 \times 10^{-19}$. The results using both methods are illustrated in Fig. 4.4B and they confirm the overall ability of the model to explain the behaviour of flies. In this figure, almost no points fall in the top-left or bottom-right quadrants, which suggests that the model's predictions showed good consistency with the data. This means that the model did not predict positive behavioural changes when negative changes were observed in flies and vice versa.

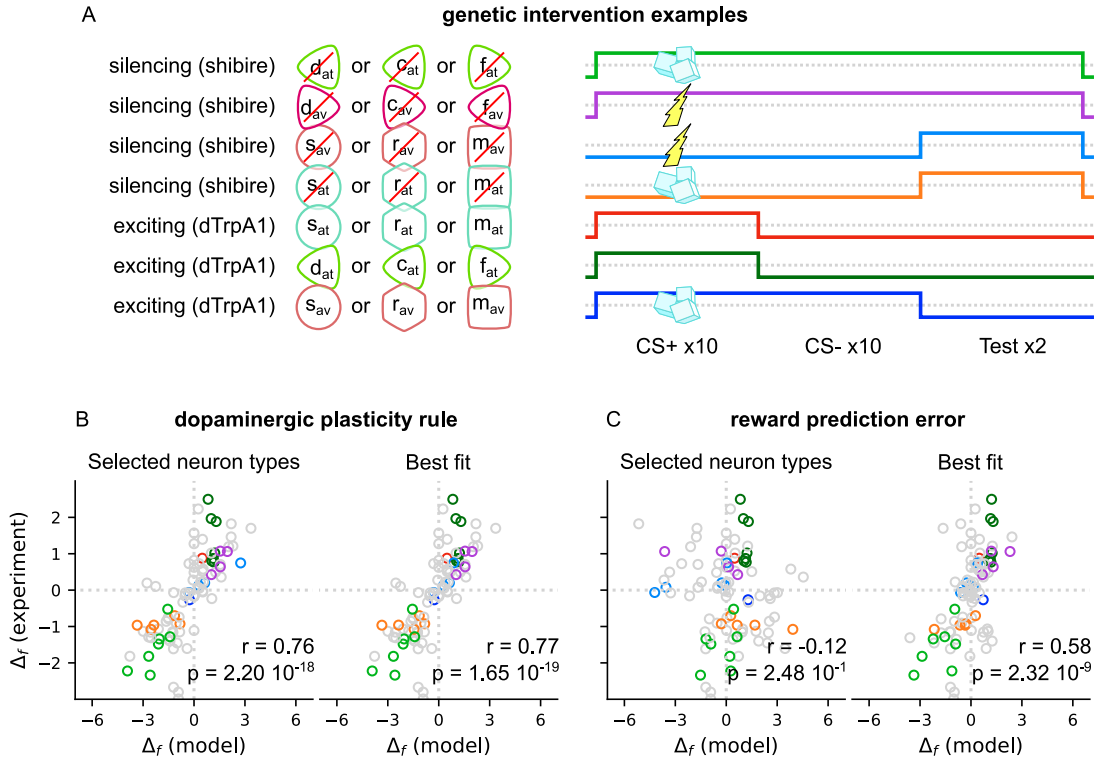

Figure 4.4: Testing the performance of the *incentive circuit* (IC) in the experiment of Bennett, Philippides, and Nowotny (2021)—Figure 5. Bennett, Philippides, and Nowotny (2021) collected behavioural data (Δ_f measure) from 92 experiments, and calculated their correlation coefficient to the behaviour produced by their model (VSλ model: $r = 0.68$, $p < 10^{-4}$; MV model: $r = 0.65$, $p < 10^{-4}$). The behavioural data involved intervention (activation or silencing) in different *mushroom body output neurons* (MBONs) or *dopaminergic neurons* (DANs); the various experiments are grouped by colour codes. For convenience, the same colour codes as in the original paper are used here. (A) Examples of the neural activity intervention: 10 trials of odour A + shock or sugar (indicated with thunder or cubes respectively) or without reinforcement (absence of thunder and sugar cubes), followed by 10 trials of odour B without reinforcement (acquisition phase). Then there are 2 trials of testing odour A vs odour B (extinction phase). In addition, neural activity intervention was modelled by targeting selected neurons of the IC and silencing them via the *shibire* blockage or activating them through the dTrpA1 channel (the timing of the intervention is shown by the coloured lines). The behaviour of the IC was tested using two plasticity rules: (B) the *dopaminergic plasticity rule* (DPR) and (C) the reward prediction error (RPE). As the types of the intervened MBONs (susceptible, restrained, or LTM) and DANs (discharging, charging, or forgetting) were unknown, all the combinations of types were tested. The results based on a guess of the intervened neurons' type are plotted under the 'Selected neuron types'; the results that showed the highest correlation with the data are plotted under the 'Best fit'. The colours of the samples correspond to the different lines of (A). Adapted from Gkanias, McCurdy, et al. (2022) under CC BY 4.0.

By replacing the *dopaminergic plasticity rule* (DPR) with the *reward prediction error* (RPE) plasticity rule in the IC, the correlation coefficient dropped for both the selected neuron types ($r = -0.12$, $p = 2.48 \times 10^{-1}$) and the best-fit methods ($r = 0.58$,

$p = 2.32 \times 10^{-9}$; see also Fig. 4.4C). This demonstrated that the DPR played an important role in the success of the IC in explaining the data.

4.2.3 *Arena experiments*

The behaviour produced by the IC was further tested by simulated flies placed in a virtual arena. These flies were exposed to two odour gradients of different strengths and variously paired with reinforcements. Full access to the neural responses, synaptic weights, and positions of the simulated flies throughout the experiment enabled the identification of different aspects of the produced behaviour and motivation. These included the effect of the *long-term memory* (LTM) on the behaviour, and whether choice occurred because the flies were attracted by one odour or repulsed by the other. Each simulated fly was equipped with the IC model (unaltered from Chapter 3) and placed in a round arena filled with two odours. Its behaviour was driven by four forces computed by its distance from the two sources and the activity of the model's MBONs (as illustrated in Fig. 4.2). First, the fly was left to explore the arena for $20\,\mathrm{sec}$ before the reinforcement (electric shock or sugar) was on-set (for another $30\,\mathrm{sec}$). After the reinforcement off-set, the fly was left in the arena for another $50\,\mathrm{sec}$ to express its learnt behaviour. A behavioural PI was derived based on the time that the simulated flies spent exposed to each odour during relevant time periods. Fig. 4.5 summarises this experimental set-up and results, while the details of the implementation are in Section 4.4.3.

Fig. E.22 shows that most simulated flies did not visit any of the regions where an odour can be detected in the first repeats; therefore, an effect in the averaged statistics started appearing after the second repeat of the experiment (see Fig. 4.5B). In the first couple of repeats, the individual paths already showed a small tendency to the expected behaviour of the flies: they avoided the punished region and approached the rewarded one. Due to the unpredictable behaviour of the individual flies, Fig. 4.5B summarises only times from simulated flies that have visited both odours for at least one second. In later repeats of the experiment, the PI shows that (on average) flies preferred the non-punished and rewarded odours. When both of them were punished or rewarded they equally preferred none or both respectively. Note that the above result does not mean that each fly spent equal time in both odours; instead, some flies chose to spend more time with the one and some with the other odour (as it is shown from the individual cumulative durations in Fig. 4.5B), while their population was (on average) equal. It is interesting that almost in every repeat the flies were neutral about the odours during pre-training (time-step before the reinforced one—shown with red or green dots), they showed a relatively small effect during training and a

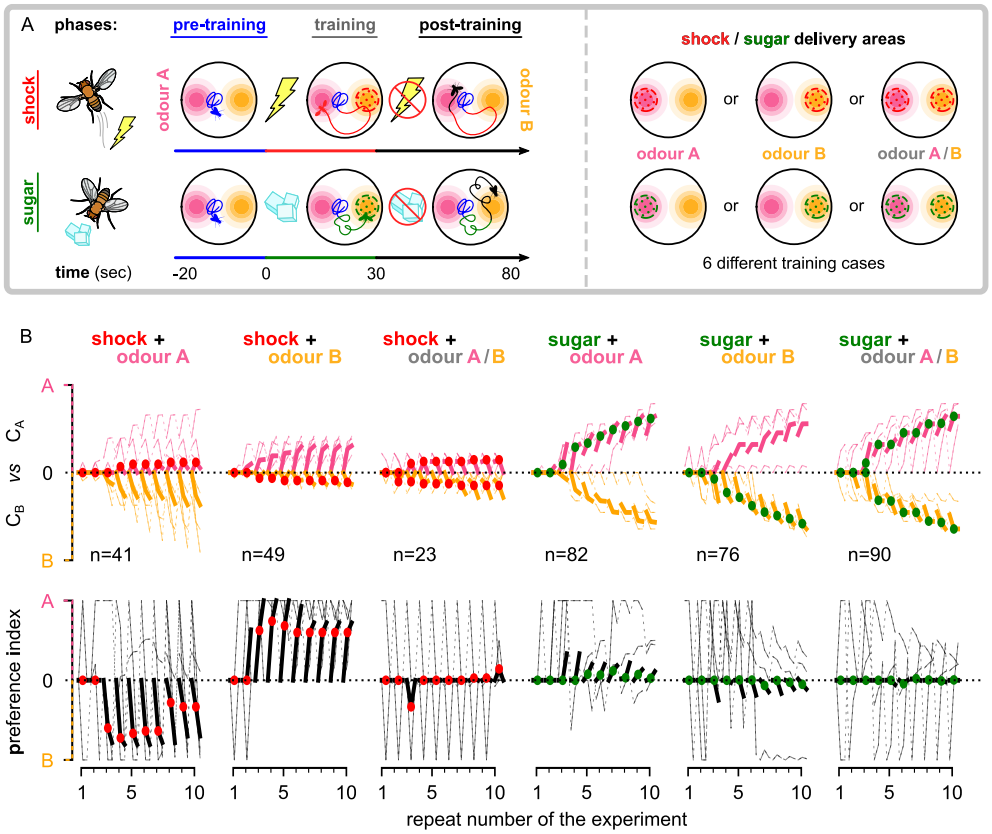

Figure 4.5: The behaviour of the animals controlled by their neurons during the simulation (where they can freely move). The $n = 100$ simulated flies were exposed to a mixture of two odours, whose relative intensity depended on the position of the simulated flies in space. (A) Each experiment lasted for $100\,\text{sec}$ where: the flies were placed at the centre of the arena in time-step $t = -20\,\text{sec}$. During the first $20\,\text{sec}$ (pre-training phase, $t \in [-20, 0)$) the flies explored the arena without any reinforcement (blue tracks). In the next $30\,\text{sec}$ (training phase, $t \in [0, 30)$) they conditionally received reinforcement under one of the six training cases shown on the right: using sugar (green) or shock (red); and reinforcing around odour A (shock + odour A), odour B (shock + odour B), or both odours (shock + odour A/B). During the last $50\,\text{sec}$ (post-training phase, $t \in [30, 80)$) they continued being exposed to the odours without receiving a reinforcement (black tracks). This experiment was repeated (including all its phases) 10 times in order to show the effects of the *long-term memory* (LTM) in the behaviour. (B) Behavioural summary of a subset of simulated flies, that visited both odours at any time during the 10 repeats. Columns show the different conditions and the population that was recorded visiting both odours. Top row: the normalised cumulative time spent exposed in odour A (pink lines) or odour B (yellow lines—note this line is reversed). For each repeat, three values were presented (averaged over all the pre-training, training, and post-training time-steps respectively) where the values associated with the training phase are marked with red or green dots when punishment or reward was delivered to that odour respectively. Thin lines show 3 representative samples of individual flies. Thick lines show the median over the simulated flies that visited both odours. Bottom row: the *preference index* (PI) to each odour extracted by the above cumulative times. Adapted from Gkanias, McCurdy, et al. (2022) under CC BY 4.0.

bigger effect during post-training. This might be because in every repeat of the experiment they were initialised in the centre, so they spent some time randomly exploring before they detected an odour.

The PIs of Fig. 4.5B show a strong effect when the electric shock was paired with odour A or B, but not very strong otherwise. A smaller PI was observed for simulated flies experiencing sugar compared to flies that experienced electric shock, which was also observed in experiments with real fruit flies (Krashes and Waddell, 2011a; Krashes and Waddell, 2011b). When the shock was paired with both odours, the simulated flies were expected to minimise the time spent exposed to either of the odours, which is precisely what the coloured lines show. In contrast, simulated flies seemed to increase the time spent in both odours when paired with sugar, with a slight preference towards the reinforced odour. In general, the results of Fig. 4.5B and Fig. E.20A show that (in time) the simulated flies developed some prior knowledge about both odours when experienced at least one of them with reinforcement. This was because some KCs responded to both odours (overlapping, as shown in Section 3.5.1) creating shared memories for the two odours. Self-reinforcement (through the LTM microcircuit) further enhanced these memories and also transferred them to the other KC $\rightarrow$ MBON synapses of the non-reinforced odour, which is effectively a form of second-order conditioning.

The summarised synaptic weights of Fig. E.19 show that the susceptible MBONs immediately blocked the simulated flies from approaching the punishing odours, but they allowed them to approach the rewarding ones. This explains the smaller PI shown in sugar-related experiments compared to the shock-related ones, which could be verified through the appetitive experiments. The simulated flies preferred the rewarding odour site, allowing the pairing between the odour and the reward so the synaptic weights can change. Susceptible MBONs convulsively broke the balance between attraction and avoidance (created by the restrained and LTM MBONs, also affecting their responses), allowing *short-term memory* (STM) and consequently LTM formation even in the absence of external reinforcement. Fig. E.19 also shows that the restrained MBONs seemed to play an important role during the first repeats (up to 5). Then they seemed to reduce their influence, giving up the control to the LTM MBONs, which increased their influence with time. This was partially an effect of the *memory assimilation mechanism* (MAM) microcircuit, which verified its function and the role of the restrained MBONs in storing STMs. Fig. E.21 shows that the different types of MBONs could separately control the behaviour. However, they seemed to better work when combined, as they complement one another in different stages (for example, during early or late repeats of the experiment).

The above results suggested that the IC was able to explain quite accurately the behaviour of *D. melanogaster* fruit flies in the tested olfactory conditioning tasks. They suggested potential explanations of why the model could sometimes predict positive patterning, while negative patterning was harder to predict. This effect might have been a result of the odour encoding by the KCs, but also of the plasticity effects of the DPR. The IC was able to predict the behaviour of fruit flies even after introducing neural activity interventions to the experiments, and also provided insights for the role of its different microcircuits to a more natural scenario, where actions of the flies were affecting their feedback from the environment. This section discusses the contribution of the DPR and the IC in the results presented in this chapter in relation to the results from Chapter 2 and 3.

4.3.1 *Advantages of the dopaminergic plasticity rule*

The olfactory conditioning experimental paradigms tested in this chapter, allowed for the DPR to demonstrate some advantages in the memory dynamics of the IC, and in the behaviour of the flies. Although it remains very simple, the DPR (proposed in Section 2.2) allowed the animal to express a variety of behaviours depending on their experience. The combination of a positive or negative dopaminergic factor with active or inactive KCs leaded to four possible effects on the synapse: depression, potentiation, recovery, and saturation. These allowed for substantial flexibility in the dynamics of learning in different MB compartments, which were also pictured in the behavioural results presented in Section 4.2.

More specifically, the saturation effect allowed for LTM MBONs to consolidate their memories and made them hard to forget. This only occurred for consistently experienced associations, which then became strongly embedded. Only a persistent change in the valence of reinforcement experienced with a given stimulus could reset the activity of LTM MBONs through the *reciprocal long-term memory* (RLM) microcircuit, which equipped the circuit with flexibility even in the LTMs. Further, the fact that the DPR allowed STMs (restrained) and LTMs to interact through the MAM, increased the capacity of the circuit. Whatever the restrained MBONs learned was eventually assimilated by the LTM MBONs, opening up space for the formation of new memories in the restrained MBONs. When combined with sparse coding of odours in a large number of KCs, the LTM MBONs can store multiple memories, for different odours. Short-term experience might occasionally affect the behaviour when the susceptible and restrained MBONs learn something new. When this happens, the STM masks the

LTM output, but eventually, this is smoothly integrated with the previous experience in the LTM MBONs. As it has been demonstrated, the DPR plays an important role in this mechanism, and the connectivity alone is not enough for this mechanism to work properly.

Fig. E.23, Fig. E.24, and Fig. E.25 demonstrated that the RPE plasticity rule lacks the above flexibility, and failed to maintain useful LTMs when applied to the same circuit architecture in the arena experiment of Section 4.2.3. Note that, RPE could be implemented by circuits (Bennett, Philippides, and Nowotny, 2021; Springer and Nawrot, 2021; Eschbach et al., 2020) in which DANs transmit an error signal computed by their input reinforcement plus the opposing feedback from MBONs (MBONs inhibit DANs that increase the KC $\rightarrow$ MBON synaptic weights or they excite those that suppress the synaptic weights). Such circuits were not explored in this work, but they could implement RPE even when the DPR is used. Although the proposed IC (Section 3.3.1) did not include such connections, it is possible that they exist in the MB. The evidence for MBON $\rightarrow$ DAN feedback connections is well-grounded, but whether they are consistently opposing is less clear. In the microcircuits of the IC, some DANs that depress synaptic weights receive inhibitory feedback from MBONs (Pavlowsky et al., 2018) and some other DANs that potentiate synaptic weights receive excitatory feedback from MBONs (Ichinose et al., 2015). This is based on neurophysiological evidence and it has been demonstrated that the DPR is able to operate with this variety of MBON $\rightarrow$ DAN connections.

4.3.2 *The effects of olfactory conditioning*

During the past decades a variety of learning effects have been investigated in flies (including forward and backward conditioning, first- and second-order conditioning, and blocking), which were used (in this and previous chapters) to challenge the IC. Section 2.2.2 demonstrated that the IC supports the backward (or relief) conditioning results presented in Handler et al. (2019). Backward conditioning is when the reinforcement is delivered just before the odour presentation and it is based on the time dependency between the two stimuli. Handler et al. (2019) suggested that the backward conditioning is a mechanism driven by ER$-Ca^{2+}$ and cAMP in a KC $\rightarrow$ MBON synapse when a single DAN releases *dopamine* (DA) on it. In Section 2.2.2, it was assumed that different time-courses in the response of DopR1 and DopR2 receptors cause the different patterns of ER$-Ca^{2+}$ and cAMP, resulting in the formation of opposite associations for forward and backward conditioning. In the above result, this effect also required that the target MBON inhibits the respective DAN—as in the *susceptible memory* (SM) microcircuit of the IC—altering the time course of neuro-

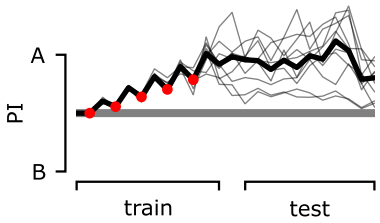

Figure 4.6: The *preference index* (PI) of the agent during the classic unpaired conditioning paradigm. During the training phase, electric shock and odour A were delivered alternately. During the test phase, odour A and odour B were delivered alternately. The PI was calculated by using the *mushroom body output neuron* (MBON) responses for each odour. Adapted from Gkanias, McCurdy, et al. (2022) under CC BY 4.0.

transmitter release. This may suggest that backward conditioning does not occur in all MB compartments (which was also suggested previously for the larval flies; Weiglein et al., 2021). An alternative mechanism for backward conditioning is the post-inhibitory rebound in opposing valence DANs (Adel and Griffith, 2021). Although some role for both mechanisms is possible, the mechanism proposed in this work is better supported.

Backward conditioning can be distinguished from the unpaired conditioning effect. Unpaired conditioning involves the presentation of reinforcement and an odour in alternation with less temporal proximity. It has been observed that this procedure produces a change in response to the odour that is opposite in valence to the reinforcement (for example, approach to an odour that is *unpaired* from shock; Jacob and Waddell, 2020; Schleyer, Fendt, et al., 2018). Note that this effect can be observed both in standard two odour CS+/CS- training paradigms (where an altered response to CS- in the opposite direction to CS+ is often observed) but also in single odour unpaired paradigms. Not surprisingly, the IC also produced unpaired conditioning, notably through a different mechanism than backward conditioning. When DANs are activated by a reinforcement without KC activation, the weights of all KCs are potentially altered (restored towards their resting weight or slightly potentiated). This alteration means that the subsequent presentation of odour alone can be accompanied by MBON-driven activation of DANs, resulting in specific alteration of the weights for the presented odour. In the example of Fig. 4.6, odour A started to self-reinforce its attractive LTM when presented in alternation with shock and will be preferred to an alternative odour B in subsequent testing. However, repeated presentation of other odours during testing (without further shock) might lead to generalisation (equal preference to all experienced odours).

The self-reinforcing property of the positive feedback in the LTM microcircuit could also account for second-order conditioning. If a motivation has been associated with

an odour, MBONs related to that motivation will have increased activity when the odour is delivered (even in the absence of reinforcement). In the LTM microcircuit, the excitatory MBON → DAN connection will consequently activate the charging DAN. So any additional cue (or KC activity) presented alongside the learned odour will also experience an increase in the respective KC → MBON weights; this would create a similar charging momentum for that cue and resulting in a second-order association. Perhaps surprisingly, this predicts that second-order conditioning might happen directly in the LTM microcircuit without being filtered by the susceptible and restrained memories first. This would be consistent with the observation that second-order conditioning in flies requires strong induction of the first-order memory and that first-order memory does not appear to be extinguished by the absence of reinforcement during second-order training (Tabone and J. S. d. Belle, 2011).

Finally (unlike RPE) the DPR would not produce blocking. The blocking effect, as described by Kamin (1967), is when the conditioning to one stimulus subsequently blocks any conditioning to other components of a mixture including that stimulus. Under RPE learning, this is explained by the first stimulus already correctly predicting the reinforcer, so there is no error to drive a change in the weights. Using the DPR, the updates are local to the synapse and do not depend on a calculation of errors summarised across different odour identities, so blocking does not happen (which is consistent with the observed behaviour of fruit flies; J. Young et al., 2011; Brembs and Heisenberg, 2001). Although the presentation of a learned odour along with a novel odour might alter the DAN responses to the reinforcement (through feedback from the MBONs), in the IC this is not generally an opponent feedback and will not cancel the reinforcing effects for the novel odour. This also highlights the difference of the susceptible and LTM microcircuits from the RPE circuits described by Bennett, Philippides, and Nowotny (2021), Springer and Nawrot (2021), Eschbach et al. (2020), and C. Zhao et al. (2021). Nevertheless, the fact that blocking has not been observed in fruit flies could also be explained by the way that the mixture of odours is represented by the KCs (it might not be simply the superposition of the activity patterns of the individual odours; Wessnitzer et al., 2012; Bennett, Philippides, and Nowotny, 2021).

4.4 METHODS

For consistency, the same structure and parameters introduced in Section 3.5 were also used in this chapter. The only parameters of the model that may vary are the number of PNs and KCs, and the way they encode the odours. These are explained in detail in the following sections, and they are specific to the different experiments.

4.4.1 *T-maze experiments*

In these experiments, $n_{\text{fly}} = 100$ flies were simulated in the T-maze set-up (Quinn, W. A. Harris, and Benzer, 1974). Following the methods from Section 3.5.3, a range from 1 to 10 cycles was used for the training (acquisition) phase and only 1 cycle for the test (extinction) phase. Each cycle was run for 100 iterations per participating odour: 33 iterations of fresh air, followed by 33 iterations exposed to odour only, followed by another 34 iterations exposed to odour and electric shock (where appropriate, otherwise to odour only; see Fig. 3.15—A and C). Noise drawn from $\eta \sim \mathcal{N}(\mu = 0, \sigma = 0.2)$ was added to the PN responses to enforce diversity in the behaviour of the individual simulated flies. In this experiment, there was a maximum of 4 possible odours per condition, and therefore, 4 PNs and 20 KCs were used for the model. Each PN was connected to 6 KCs, which enforced some overlap among the odour representations in the KC layer (1-2 KCs overlap). The PN $\rightarrow$ KC synaptic weight followed the same pattern to Eq. (3.2), but for 4 PNs and 20 KCs,

$$\mathbf{W}^{p2k} = \begin{bmatrix} \zeta & \zeta & \zeta & \zeta & \zeta & \zeta & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \zeta & \zeta & \zeta & \zeta & \zeta & \zeta & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \zeta & \zeta & \zeta & \zeta & \zeta & \zeta & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \zeta & \zeta & \zeta & \zeta & \zeta & \zeta \end{bmatrix}, \quad (4.1)$$

where $\zeta = 0.8$. The US (reinforcement) was represented similarly to the one described in Section 3.5.1, but in this case, only the punishing reinforcement was used.

The implementation of all the olfactory conditioning paradigms was similar. The only difference was the number of cycles (in the training phase), the odour identities used for the experiment (both in the training and test phases), the order that these odours were presented, and which odours were paired with reinforcement (in the training phase). Table 4.2 summarises the experiments run for all the different conditions with their parameters (see train and test columns). The individual mixtures of odours are separated with spaces, '+' denotes the presence of the reinforcement paired with this mixture of odours, and '-' denotes its absence.

Preference index

Based on the responses of the MBONs, the *attractiveness* of each odour was calculated using the responses of the MBONs,

$$v_{\text{at}} = \frac{(s_{\text{at}} + r_{\text{at}} + m_{\text{at}}) - (s_{\text{av}} + r_{\text{av}} + m_{\text{av}})}{3}, \quad (4.2)$$

where s_*, r_*, and $m*$ are the responses of the susceptible, restrained, and LTM MBONs respectively during the test; 'at' and 'av' define if they drive attraction or avoidance.

The odours were grouped in reinforced (CS+) or not (CS-), depending on whether they were paired with reinforcement or not respectively. As the reinforcement was always negative (punishment), the PI was calculated always as the normalised difference between them, with the positive part being for the non-reinforced odour. This is written formally as

$$\mathrm{PI}_{\text{T-maze}} = \frac{v_{\text{at}}^{\text{CS-}} - v_{\text{at}}^{\text{CS+}}}{2}. \tag{4.3}$$

Note that the denominator is fixed and equal to 2 because the measured values can be negative as well as positive.

4.4.2 *Neural activity intervention experiments*

In these experiments, the methods presented by Bennett, Philippides, and Nowotny (2021) were copied as accurately as possible, in order to make the comparison meaningful. Therefore, during the training (acquisition) phase, odour A was presented for 10 cycles with positive, negative, or no reinforcement, depending on the condition. Another 10 cycles followed, where odour B was presented without reinforcement. During the test (extinction) phase, 1 cycle of odour A and 1 cycle of odour B was applied. Each cycle was run for 10 iterations. Each experiment was repeated for $n_{\text{fly}} = 50$ simulated flies, with added noise to the PN responses drawn from $\eta \sim \mathcal{N}(\mu = 0, \sigma = 0.1)$. Note that, in this experiment, the PN and KC encoding was unaltered to the one used in Chapter 3, as there are never more than two odours per condition. The interventions were done by adding 5 (activation) or -5 activity units (silencing) to the response of the targeted neuron (MBONs or DANs), before passing it through the activation function. This ensured that the neuron was active or inactive respectively.

Preference index

Bennett, Philippides, and Nowotny (2021) computed the PI based on the population of flies that chose odour A or odour B. This can be written formally as,

$$\mathrm{PI}_{\text{gi}} = \frac{n_A - n_B}{n_A + n_B}, \tag{4.4}$$

where n_A is the number of flies that chose odour A (CS+), and n_B is the number of flies that chose odour B.

Calculating the effect strength of interventions

The fraction of simulated flies that chose odour A was calculated as, $f = (\mathrm{PI} + 1)/2$. The effect strength of the intervention (for each experiment) was calculated by comparing these fractions for the intervention experiment, f_i, to its equivalent without the intervention (control), f_c. Given the assumption that the underlying data are drawn from a binomial distribution, the effect strength of the intervention is,

$$\Delta_f = \frac{f_i - f_c}{\sqrt{\frac{1}{n_{\mathrm{fly}}}(f_i + f_c)[1 - \frac{1}{2}(f_i + f_c)]}}. \tag{4.5}$$

In a similar way, the strength was also measured for the data provided by Bennett, Philippides, and Nowotny (2021). Following their approach, it was assumed that $n_{\mathrm{fly}} = 50$ in the above equation, which was also used for the simulations.

Pearson correlation coefficient

Finally, the correlation between the effect strengths of the model and experimental data was calculated using the Pearson correlation coefficient,

$$r = \frac{\mathrm{cov}\left(\mathbf{w}_r \Delta_f^{\mathrm{mod}}, \mathbf{w}_r \Delta_f^{\mathrm{exp}}\right)}{\sigma_{\mathrm{mod}}\, \sigma_{\mathrm{exp}}}. \tag{4.6}$$

Before calculating the correlation, the effect strengths of the data and model were weighted by $\mathbf{w}_r$, which was calculated based on the magnitude of all the data. This is also similar to the way Bennett, Philippides, and Nowotny (2021) computed the Pearson correlation coefficient, and the same approach was used here to allow a direct comparison of the results.

4.4.3 *Arena experiment*

The experiments lasted for $100\,\mathrm{sec}$ each (with 1 iteration sec^{-1}) and they were split into 3 phases as shown in Fig. 4.5A. In *pre-training*, the flies were placed in the centre of the arena and explored freely for $20\,\mathrm{sec}$. In *training*, either shock or sugar was delivered within a region of $30\,\mathrm{cm}$ around odour A, odour B, or both odours for $30\,\mathrm{sec}$. In *post-training*, the reinforcement was removed, and the flies were let express their learnt behaviour for another $50\,\mathrm{sec}$ creating a forgetting (extinction) condition. The behaviour of the simulated flies was controlled by a combination of the attractive and repulsive forces on the two odours. Fig. 4.5B shows the normalised cumulative time spent experiencing each odour (over 10 repeats of the experiment) and the odour

preference of the flies during the different phases for each of the six training conditions. The actual paths of the flies for all the 10 repeats are illustrated in Fig. E.22.

In practice, in order to create the experiences of $n_{\text{fly}} = 100$ flies, a process was created (similar to the ones in Section 3.5.3) that embedded the simulation of the motion of flies and the environment. The position of each fly, $a(t) \in \mathbb{C}$, and the sources of the odours in the arena, $\mu_A \in \mathbb{C}$ and $\mu_B \in \mathbb{C}$ for odours A and B respectively, were represented in the 2D space as complex numbers in the form $x + iy$. The flies were initialised in $a(t = 0) = 0$ and the sources of the odours were placed in $\mu_A = -0.6$ and $\mu_B = 0.6$. The spreads of the odour distributions were $\sigma_A = \sigma_B = 0.3$.

The odour intensity at the fly's location in each iteration was calculated from the Gaussian function of the respective odour as,

$$\iota_{A/B}(t) = \exp\left\{-\frac{[a(t) - \mu_{A/B}]^2}{2\,\sigma_{A/B}^2}\right\}. \tag{4.7}$$

The PN activity was then calculated as,

$$\mathbf{p}(t) = \begin{cases} \mathbf{p_{AB}}, & \text{if} \quad \iota_A(t) > \theta_{\text{CS}} \text{ and } \iota_A(t) > \theta_{\text{CS}}, \\ \mathbf{p_A}, & \text{if} \quad \iota_A(t) > \theta_{\text{CS}}, \\ \mathbf{p_B}, & \text{if} \quad \iota_B(t) > \theta_{\text{CS}}, \\ \mathbf{p}_\varnothing, & \text{otherwise}, \end{cases} \tag{4.8}$$

where $\mathbf{p}_A$, $\mathbf{p}_B$, $\mathbf{p}_{AB}$, and $\mathbf{p}_\varnothing$ are the identities of odours A, B, "A and B", and none of them respectively in the PNs as described in Section 3.5.1, and $\theta_{\text{CS}} = 0.2$ is the detection threshold for the odours. Note that PN responses depended only on the fact that an odour has been detected or not and it was not proportional to the detected intensity. The KC activity was calculated as in Chapter 3. The reinforcement was applied to the simulated fly when the position of the agent was inside a predefined area around the odour, calculated as $||\mu_{\text{CS}} - a(t)|| < \rho_{\text{US}}$, where $\rho_{\text{US}} = 0.3$ was the radius of the reinforced area. Note that the radius of the area where the odour is detectable is roughly $\rho_{\text{CS}} \simeq 0.54$, which is larger than the reinforced area. Subsequently, the above inputs were used to run a *forward propagation* (see Section 3.5.2).

From the updated responses of the MBONs, an *attraction force*, $\mathbf{v}(t)$, was calculated for the mixture of odours, which modulated the velocity of the fly. This force was calculated based on the odour intensity at the current location of the fly (which is in-

verse proportional to the distance from the odour source) and the difference between the responses of the MBONs that drive the behaviour,

$$v_{\text{at}}(\text{CS}|t) = \frac{1}{3}[s_{\text{at}}(t) + r_{\text{at}}(t) + m_{\text{at}}(t)] \cdot \frac{\mu_{\text{CS}} - a(t)}{||\mu_{\text{CS}} - a(t)||}, \tag{4.9}$$

$$v_{\text{av}}(\text{CS}|t) = \frac{1}{3}[s_{\text{av}}(t) + r_{\text{av}}(t) + m_{\text{av}}(t)] \cdot \frac{\mu_{\text{CS}} - a(t)}{||\mu_{\text{CS}} - a(t)||}, \tag{4.10}$$

$$\mathbf{v}(t) = \sum_{\text{CS}}^{\{A,B\}} \iota_{\text{CS}}(t) \cdot v_{\text{at}}(\text{CS}|t) - \sum_{\text{CS}}^{\{A,B\}} \iota_{\text{CS}}(t) \cdot v_{\text{av}}(\text{CS}|t), \tag{4.11}$$

where ι_{CS} is the odour intensity calculated using Eq. (4.7) and $\text{CS} \in \{\text{A}, \text{B}\}$ is the odour identity. The velocity of the simulated fly was updated as follows,

$$v(t) = v(t - \text{d}t) + \mathbf{v}(t) + \varepsilon_x + i\varepsilon_y, \quad \text{where } \varepsilon_x, \varepsilon_y \sim \mathcal{N}(\mu = 0, \sigma = 0.1), \tag{4.12}$$

$$\hat{v}(t) = 0.05 \cdot \frac{v(t)}{||v(t)||}. \tag{4.13}$$

The velocity was normalised in order to keep the direction information, but the step size was replaced with $0.05\,\text{m sec}^{-1}$. The noise added to the velocity was introduced in order to enable the flies to move in two dimensions and not just between the two odour sources. Also, when the attraction force was $\mathbf{v}(t) = 0$, then the noise and the previous velocity were the ones that drove the behaviour of the flies.

The above process was repeated for $T = 100$ iterations with $\text{d}t = 1\,\text{iteration sec}^{-1}$, and shock or sugar was provided (when appropriate) between iterations 20 and 50; otherwise, a zero-vector was used as the US input to DANs.

Calculating the normalised cumulative exposure and the preference Index

In Fig. 4.5B, for each phase (pre-training, training, and post-training) the normalised cumulative exposure of the flies in each odour and the preference index between them were reported. The normalised cumulative exposure was calculated by,

$$C^R_{\text{CS, phase}} = \sum_{i=1}^{R} \frac{c^i_{\text{CS, phase}}}{T_{\text{phase}}}, \tag{4.14}$$

where R is the repeat of the experiment, i is the iterative repeat, T_{phase} is the number of time-steps for the specific phase, $c^i_{\text{CS,phase}}$ is the number of time-steps spent exposed in the specific $\text{CS} \in \{\text{A}, \text{B}\}$, phase, and repeat.

The preferences index for every repeat was calculated using the above quantities,

$$\text{PI}^R_{\text{phase}}(t) = \frac{C^R_{\text{A, phase}} - C^R_{\text{B, phase}}}{C^R_{\text{A, phase}} + C^R_{\text{B, phase}}}. \tag{4.15}$$

5

VISUAL PLACE RECOGNITION

> *"No man ever steps in the same river twice, for it's not the same river and he's not the same man."*
>
> Heraclitus

For most insects, vision plays an important role in recognising places, particularly in navigation. For example, it is crucial for desert ants and honeybees to be able to return home (as their colony depends on them to feed and survive) and these insects mostly use their visual surroundings to complete this task. A robust navigational system has been developed in their brain that enables *path integration* (Srivastava, Mansimov, and Salakhutdinov, 2015) and *visual place recognition* (VPR) (Wehner and Raber, 1979). Path integration is the process whereby an agent combines distance and direction travelled to compute its displacement from a point of reference. In insects, path integration is processed by the *central complex* (CX); distance is computed by integrating speed over time (through odometry or optical flow), and direction is given by allothetic orientation cues (like the sky). The process generates a *home vector* that they can follow to return home in a straight line (Srinivasan, 2015). In parallel, their VPR mechanism provides information about the familiarity of the current visual input, which can be used to determine their subsequent turning direction (Zeil, 2012; T. S. Collett, Graham, and R. A. Harris, 2007). Buehlmann et al. (2020) and Kamhi, Barron, and Narendra (2020) provided evidence that intact *mushroom bodies* (MBs) are needed for the use of visual memories related to this task.

This chapter explores the role of the MBs in VPR, by hypothesising methods for visual processing in the *projection neurons* (PNs) and *Kenyon cells* (KCs). It also questions the potential role of the *incentive circuit* (IC) in VPR, suggesting that familiarity also depends on the recent experiences of the animal. The output of the MBs could be used to predict the consistency of the familiar views, and it is less likely that it can provide the actual location of the animal.

Interaction between the CX and the MBs has been repeatedly suggested and verified in the context of navigation (Cruse and Wehner, 2011; Wehner, 2003; T. S. Collett and M. Collett, 2002; Menzel, Greggers, et al., 2005). Webb (2019) provided a computational perspective on the topic, and summarised the different hypotheses and methods around the complicated navigation system of insects. Recent work showed that the *fan-shaped bodies* (FBs) of the CX receive (direct or indirect) input from the MBs (Li et al., 2020; Hulse et al., 2021) and this input might be related to the VPR task. This section focuses on what is known about the role of the MBs in VPR and how the visual surroundings are encoded before they enter the MBs.

5.1.1 *Insect vision*

The eyes of insects differ substantially from the mammalian eyes. Mammalian eyes have a single lens that collects light and drives it into an interior array of photoreceptors. The eyes of insects consist of an array of lenses that provide visual information to different sets of photoreceptors (see Fig. 5.1A). For this reason, they are called *compound eyes* and their separate visual units (consisting of one facet—a type of lens—and a group of eight to nine photoreceptors) are called *ommatidia* (from Greek *ommatidion*, which means "a type of eye"). The number of ommatidia also varies between insect species, with 750-950 (per eye) in the fruit fly *Drosophila melanogaster* (Posnien et al., 2012), and 743-1254 (per eye) in the desert ant *Cataglyphis bicolour* (Menzel and Wehner, 1970; Zollikofer, Wehner, and Fukushi, 1995). Both in fruit flies (Borst, 2009) and in desert ants (Zollikofer, Wehner, and Fukushi, 1995), each eye was measured to have 180° field of view in azimuth and elevation. This is essentially a half-sphere, but with 30° overlap (the binocular zone) towards the frontal and dorsal axes, suggesting a 330° panoramic field of view (Zollikofer, Wehner, and Fukushi, 1995). Interestingly, different areas of the eye might have different properties. For example, desert ants are thought to have higher resolution in the horizon rather than the dorsal or ventral areas of their eyes (see Fig. 5.1B). Note that the resolution in the compound eyes is usually defined by the number of ommatidia per square degree. The acceptance angle of the individual ommatidia defines the sharpness (a higher acceptance angle results in blurrier views). In addition, photoreceptors in different areas might respond to different wavelengths or properties of light (Labhart, 1986). For example, ommatidia located at the *dorsal rim area* (DRA) are usually sensitive to *ultra violet* (UV) or blue light, while the ones closer to the horizon are more sensitive to green light. Ommatidia located at the DRA also respond to polarised light, a property of

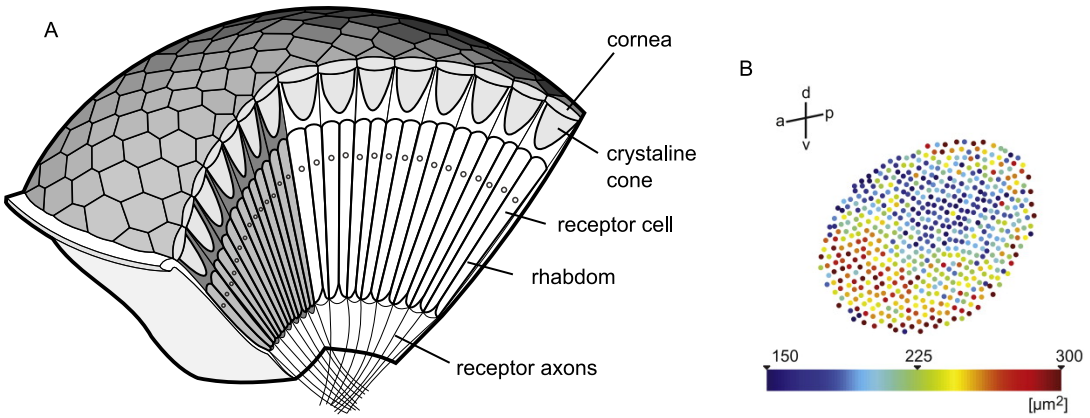

Figure 5.1: Basic structure of the compound eye. (A) The ommatidia compose the different visual-sensor units of the eye. The facet of each ommatidium allows light into the cornea and crystaline cone (in practice a lens); these drive visual information into the rhabdom and the photoreceptor cells. The information travels in the *optic lobe* (OL) in the form of neural activity through the axon of the photoreceptor. Adapted and modified by permission from Elsevier: Copyright © D.-E. Stewart (1958). (B) The distribution of ommatidia in the compound eye of the desert ant *Melophorus bagoti* covers 180° panorama in both anterior-posterior and ventral-dorsal axes. a: anterior, p: posterior, v: ventral, and d: dorsal. Adapted by permission from Elsevier: Copyright © Schwarz, Narendra, and Zeil (2011).

light that is specific to the e-vector of light oscillating mostly in one direction (see Appendix D).

Insects might use the different areas of their compound eyes in various ways, which might also depend on their ability to fly or predate. Nevertheless, in most insect eyes the responses from the photoreceptors in their DRAs contribute to the celestial compass of the CX (Wehner and Labhart, 2006), while the responses from photoreceptors in their lateral areas contribute in *optical flow* calculation (Wittlinger, Wehner, and Wolf, 2006). The photoreceptors connect to the mid-brain of the insect through the *optic lobes* (OLs). The OLs are usually organised in four layers (or neuropils): the *lamina* (LA), *medulla* (ME), *lobula* (LO), and *lobula plate* (LOP), which are all divided into an approximately equal number of columns to the number of ommatidia. Although the exact function of the above layers is not clear yet, there is some evidence that partially describes their function. It has been suggested that LA introduces global and local inhibition in a retinotopic fashion (Stöckl, O'Carroll, and Warrant, 2020), ME and LO detect spatial and temporal changes (Nordström, Barnett, and O'Carroll, 2006), and LOP is responsible for motion vision (like optic flow; Hausen, 1976; Hausen, 1984; Joesch et al., 2008). In addition, motor feedback in the ME, LO, and LOP introduce compensation for the self-motion of the animal that can isolate optic flow information of moving objects (Cruz et al., 2019). However, this is not the complete picture. For example, while the LOP is usually required for processing optic flow, *Cataglyphis* ants (and other hymenopteran species like the praying mantis, locust, and bee) do

not have LOPs whilst still being able to process optic flow (Pfeffer and Wittlinger, 2016; Habenstein et al., 2020).

5.1.2 *Visual projection neurons*

The *visual projection neurons* (vPNs) transfer information from the OLs (specifically from the MEs and LOs) to the mid-brain of insects. According to Habenstein et al. (2020), there are six main types of vPNs: the *anterior optic tract* (AOT), the *serpentine optic commissure* (SOC), the *posterior optic commissure* (POC), the *inferior optic commissure* (IOC), the *anterior superior optic tract* (ASOT), and the *optical calycal tract* (OCT). The AOT neurons convey visual motion (T. Collett, 1972; DeVoe et al., 1982; Paulk, Phillips-Portillo, et al., 2008), chromatic, and polarisation information (Kinoshita, Pfeiffer, and Homberg, 2007; Mota et al., 2011; Pfeiffer, Kinoshita, and Homberg, 2005) to the *anterior optic tubercle* (AOTUs), a structure upstream of the CX. The axons of SOC neurons connect the two OLs, and their role is probably related to detecting and tracking moving objects (Hertel, Schäfer, and Maronde, 1987; Loesel and Homberg, 2001). The POC and IOC neurons are achromatic (Hertel, Schäfer, and Maronde, 1987; Paulk, Dacks, et al., 2009) and project to many regions in the brain including regions in the opposite OL, the CX, and the *convergence zones* (CZs). POCs are assumed to transfer information of stationary targets while IOCs transfer the direction of moving objects (Hertel, Schäfer, and Maronde, 1987; Maronde, 1991).

The other two types of vPNs (the OCT and ASOT) are known to project visual information from the OLs to the calyces of the MBs. In *Cataglyphis* ants, the OCT neurons project into the calyx of the MB of their same hemisphere, while a small number of ASOT neurons projects into the MBs of both hemispheres (Habenstein et al., 2020). Although the OCT neurons were described for the first time in *Cataglyphis* ants (Habenstein et al., 2020), it is possible that they also exist in other insect species. While it is known that OCT and ASOT neurons project to the calyces of the MBs, there is limited information about their actual function. Ehmer and Gronenberg (2002) found that ASOT input to the MBs of honeybees was segregated between the dorsal or ventral halves of the eyes. Although a retinotopic organisation of this input might be in place, they found it quite implausible. On the contrary, they supported the idea of the visual input being organised in retinotopic regions that are related to different tasks (Wehner, Herrling, et al., 1972; Lehrer, 1998), and they suggested that visual input entering the MBs is organised in a similar way. While they did not provide any concrete description of the visual information that these neurons might convey to the MBs, Ehmer and Gronenberg (2002) speculated that this might be related to the colour, form, texture, or movement of the visual surroundings. OCT neurons were

described for the first time by Habenstein et al. (2020) and there is no further information about their function.

5.1.3 *Sparse coding in the Kenyon cells*

A common assumption for the function of the KCs is that they are a sparse representation of the information encoded by the PNs, which might encode input from different modalities (like visual and olfactory). Laurent (2002) suggested that the odour representation in the olfactory PNs is a compact and easy-to-recall memory, and that the responses of the individual PNs should have minimum correlation. The KCs then decode this signal and create a sparse representation in space and time that relies on spatiotemporal coincidences—possibly with help from the *anterior paired lateral* (APL) neurons (Papadopoulou et al., 2011). This means that a tiny number of KCs should fire at a time and this should be brief (for example, two spikes per second). Laurent (2002) also suggested that the goal of this encoding-decoding mechanism is not to increase the storage size of the circuit, but rather to ease the handling of a few and more useful memories for the animal. Subsequent calcium imaging studies suggested that the connections between PNs and KCs are random (Caron et al., 2013).

Combinatorial activity patterns have been detected throughout the different layers of processing—*antennal lobes* (ALs) → PNs → KCs—resulting in representations that become sparser in each layer (Szyszka, Ditzen, et al., 2005). It has been suggested that AL population activity is not very specific to the odour experienced; however, population activity becomes more specific in the PNs and more so in the KCs—specificity increases with each layer. The specificity arises by capturing combinations of elements which construct an odour (or mixture of odours). In the brain, this is accomplished by capturing spatial and temporal dynamics of neuron activity (which neurons fire and in what order). The approach of creating responses (at the population level) that capture such combinations of events is called *combinatorial optimisation*. Litwin-Kumar et al. (2017) tested a PN → KC connectivity pattern based on a theoretically optimal combinatorial approach, showing that they could increase the effectiveness of the circuit (which is its ability to store and recall memories of combined elements). However, the computational cost of finding such optimal solutions is exponential with respect to the size of the circuit (Bjorndal et al., 1995), suggesting that nature might have found a heuristic way to optimise the connections of these circuits.

5.1.4 *The behavioural readout*

Trying to understand how insects use visual information for navigation is a popular topic in neuroethology. Most findings around VPR come from ethological experiments, which provided a great conceptual background of the visual processing mechanisms and the types of memories they use. For example, Wehner, Michel, and Antonsen (1996) suggested that desert ants use the different regions of their eyes to recognise places and landmarks or follow routes. However, it is arguable whether ants actually identify individual landmarks and make calculations among them in order to estimate their precise position. It is more widely accepted that they simply compare their stored and current views (in the form of two-dimensional panoramic images) while rotating systematically to find the most familiar direction (Zeil, Hofmann, and J. S. Chahl, 2003). Sommer, Beeren, and Wehner (2008) further demonstrated that desert ants can remember visual memories from multiple routes. Also, their ability to identify familiar views could be used to drive them towards a previously experienced place, like their nest, a food source, or a previously followed route to the nest (Zeil, 2012; T. S. Collett, Graham, and R. A. Harris, 2007). Another hypothesis is that they associate the views to a direction with respect to the sky (Wehner, 2003) or motor routines (for example, turn right, turn left, or continue straight; T. S. Collett and M. Collett, 2002; Wystrach, Moël, et al., 2020). Menzel, Greggers, et al. (2005) and Wang et al. (2021) suggested that (in honeybees) these directions could actually be vectors (containing both direction and distance information), but there is no evidence for vector memories in ants so far. Which of those strategies are actually used by the different insect species is still an open question.

5.2 RELATED WORK

The anatomical and functional description of the visual processing in the insect brain (as summarised in Section 5.1) provided inspiration for a number of computational models. This section summarises computational work related to visual processing from the level of ommatidia processing to the formation of memories in the MBs.

5.2.1 *Rendering the visual surroundings*

Most image processing methods developed up to this point work with conventional cameras that have a constrained field of view, and capture only a small part of the visual scene (or surroundings). However, insects have almost 360° panoramic vision

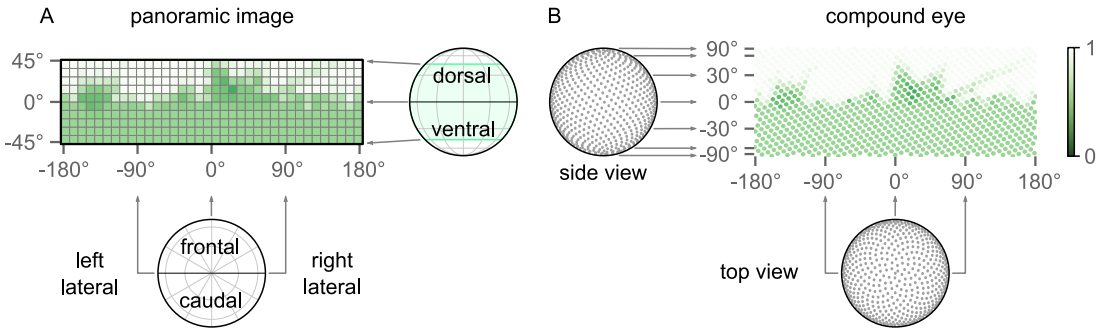

Figure 5.2: Comparison of the visual rendering by panoramic camera-like processes and by the compound eye. (A) Example of rendering using a panoramic camera-like process. They usually keep only information from the $[-45°, +45°]$ elevation (see the green shaded area in the spherical representation), neglecting information from the ground and sky. If they use information from $[-90°, +90°]$, too many pixels are assigned to the poles, where there is not much visual information related to the vegetation. The spherical images are separated by the horizontal axis to the dorsal and ventral areas, and by the lateral axis to the frontal and caudal areas. Grey lines in the spherical representations denote intervals of $30°$. (B) Simulated responses of the photoreceptors from the ommatidia, projected onto the Cartesian space with azimuth on the horizontal and elevation on the vertical axis. The compound eye allows different distributions of ommatidia for different species that are adjusted to their needs. Here, a homogeneous distribution of ommatidia is used for demonstration.

with relatively low resolution and unconventional optical properties. This is usually approximated by panoramic cameras with reduced resolution through downsampling (Baddeley et al., 2012; Ardin, Peng, et al., 2016; Stone, Mangan, et al., 2018; Sun, Yue, and Mangan, 2020; Möel and Wystrach, 2020; Wystrach, Moël, et al., 2020; Goulard et al., 2021). Approaches that rendered the environment by using this representation had to deal with a major problem: the distortion caused by the dewarpping of the spherical into a two-dimensional image. This resulted in a relatively high number of pixels associated with the two poles of the sphere, meaning higher resolution at the poles than at the horizon. To address this problem, most approaches excluded these pixels, keeping only the ones close to the horizon (Fig. 5.2A). Islam et al. (2022) supported this idea, suggesting that *Myrmecia midas* ants do not make a retinotopic comparison of their surrounding, and they seem to focus more on the horizon and the frontal area of their eyes. An exception to the horizon-focused methods was the study from Stankiewicz and Webb (2021), who built a flying robot equipped with a conventional camera pointing at the ground (mimicking the honeybee homing behaviour). Another exception was the work of Zhu, Mangan, and Webb (2020), who used inputs from a neuromorphic event-based camera to a spiking model. This camera was not panoramic, but it inherited some interesting spatiotemporal properties for the representation of the environment. There have only been a few studies that used rendering techniques constrained by the structure and properties of the com-

pound eyes of insects (Stürzl, Boeddeker, et al., 2010; Stouraitis et al., 2017; Polster et al., 2018; Millward, Maddock, and Mangan, 2021), but they did not use it for the purpose of VPR. For VPR, the closest approximations to the insect eye were downsampled panoramic images (in greyscale), and event-based images which attempted to capture spatiotemporal information.

Visual projection neurons

As discussed in the previous section, visual input goes through multiple processing stages before reaching the vPNs. However, Baddeley et al. (2012), Möel and Wystrach (2020), Wystrach, Moël, et al. (2020), and Stankiewicz and Webb (2021) suggested that the raw pixel intensities were sufficient for the purpose of VPR. A similar approach was followed by Ardin, Peng, et al. (2016), with the difference that they transformed the intensities to a binary, spiking code before feeding them to their MB model and normalised them so that half of the vPNs are active at all times. Zhu, Mangan, and Webb (2020) used the same approach, but with spikes from their event-based cameras instead of raw pixel intensities. Goulard et al. (2021) emulated the spatial component of these events by using edge-detection units instead of the raw intensities; this is plausible as the OLs (Section 5.1.1) could implement edge-detection (Stöckl, O'Carroll, and Warrant, 2020). On the other hand, Stone, Mangan, et al. (2018) and Sun, Yue, and Mangan (2020) exploited the rotational invariant properties of *Zernike moments* (ZMs) (Teague, 1980; Khotanzad and Hong, 1990) to construct encoding for the vPNs; Stone, Differt, et al. (2016) also exploited rotational invariant properties but though spherical harmonics (which is similar but not the same to ZM). Like a *Fourier transform*, these approaches transformed the visual input into the frequency domain, but they also exploited the panoramic nature of the input. Given the panoramic eyes of insects, rotational invariance is an attractive feature; however, its plausibility is arguable. Lee and D. Kim (2018), S. Meyer et al. (2020), and Stankiewicz and Webb (2021) used a different transformation, the Haar-like wavelet encoding, to reduce the size of captured images without penalising the performance. Inspired by the visual responses of the ring neurons (upstream the CX of fruit flies), Dewar et al. (2015) implemented filtering neurons (named Rχ) that represented the visual input using a small ring structure; this proved sufficient for small-scale navigation. Overall, the methods implemented to capture the responses of the vPNs capture a small part of their characteristics described in Section 5.1.1. Despite the spatial correlations captured by edge detection (Goulard et al., 2021) and ZMs (Sun, Yue, and Mangan, 2020), none of the above approaches considered temporal correlations in the OLs (Nordström, Barnett, and O'Carroll, 2006) or the de-correlation of the visual input (Laurent, 2002).

Sparse coding in the Kenyon cells

Not all VPR approaches included the KC responses in their model. Baddeley et al. (2012), Möel and Wystrach (2020), and Stankiewicz and Webb (2021) used models that were not constrained by the structure of the MBs, and therefore they did not explore mechanisms for the sparse coding observed in the KCs. Sun, Yue, and Mangan (2020), Wystrach, Moël, et al. (2020), and Goulard et al. (2021) followed the pattern introduced by Wessnitzer et al. (2012) and Ardin, Peng, et al. (2016), which was based on the findings that the connections between vPNs and KCs are random (Caron et al., 2013). Zhu, Mangan, and Webb (2020) followed a similar principle, but they also introduced plastic KC to KC connections that additionally captured the temporal dynamics of events and naturally built a sparse code. Akin to Peng and Chittka (2017), the KC sparsity was enhanced by adding the APL neuron (also suggested by Papadopoulou et al., 2011, and modelled by Nowotny et al., 2005, for olfaction), which ensured that a fixed number of KCs was firing. However, none of the proposed solutions actively exploited the combinatorial approach suggested by Litwin-Kumar et al. (2017)—note that this work did not describe how to build the PN $\rightarrow$ KC connections; instead, they theoretically explored what would be the effects of this approach—which was also backed by data (Laurent, 2002).

5.2.2 *Visual place recognition*

Most of the models mentioned so far tried to solve the VPR task. Only three computational methods have been explored in relation to this task: *perfect memory* (PM), *infomax* (IM), and *Willshaw network* (WN).

Perfect memory

The simplest and most successful model for VPR is the *perfect memory* (PM). As its name indicates, this model represents an unlimited (perfect) memory that can be used to evaluate the familiarity of the current view. This memory is essentially a database of views (for example, in the form of vPNs activities) that is initially empty. During a training phase, views of the environment (in the form of vPN activity) are copied into the database. Any subsequent view is then compared to those stored in the database by using an objective function, such as *root square error* (RSE). The com-

parison that gives the least error indicates the novelty of that current view. This is formally written as,

$$v^{\text{fam}} = 1 - v_{\text{nov}}, \tag{5.1}$$

$$v^{\text{nov}} = \min_{i \in \{0,\ldots,\text{DB}-1\}} ||\mathbf{x}^{\text{pn}} - \mathbf{x}_i^{\text{db}}||, \tag{5.2}$$

where $\mathbf{x}_i^{\text{db}}$ is a view stored in the database as a vector, DB is the maximum number of views that can be stored in the database, $\mathbf{x}^{\text{pn}}$ is the current vPN activity as a vector (which is, the current view), v^{nov} is the novelty, and v^{fam} is the estimated familiarity value of the current view.

This model was introduced by Baddeley et al. (2012) as a ground truth model for VPR; it was later used by Ardin, Peng, et al. (2016), Möel and Wystrach (2020), Wystrach, Moël, et al. (2020), and Stankiewicz and Webb (2021) to explore different applications of VPR. Möel and Wystrach (2020) used this model to store separately attractive (when looking towards the nest) and repulsive views (when looking away from the nest), aiming for robust homing behaviour. The same model was also used by Wystrach, Moël, et al. (2020), who propagated the familiarity signal to FBs of the CX, which resulted in left or right steering commands. Finally, Stankiewicz and Webb (2021) used PM on an autonomous quadcopter (equipped with a downwards-facing camera) and successfully followed a familiar route by performing transverse oscillations above it.

Infomax

PM (as described above) has unlimited storage and can perfectly recall each individual view at any time; infinite capacity is obviously unrealistic. A model closer to the structure of a brain—using *artificial neural networks* (ANNs)—is *infomax* (IM) (Bell and Sejnowski, 1995). This was first used by Baddeley et al. (2012) to predict the familiarity of the views by calculating the novelty vector, $\mathbf{v}^{\text{nov}} \in [-1, 1]^X$, from the raw visual input, $\mathbf{x}^{\text{pn}} \in [0, 1]^X$. The novelty vector was calculated using the equation below,

$$y_j^{\text{nov}} = \tanh a_j, \qquad a_j = \sum_{i=0}^{X-1} x_i^{\text{pn}} w_{ij}. \tag{5.3}$$

The plasticity rule used to update the weights-matrix was given by,

$$\Delta w_{ij} = \frac{\eta}{X}\{w_{ij} - [y_j^{\text{nov}} + a_j] \sum_{k=0}^{X-1} w_{ik} a_k\}, \tag{5.4}$$

where η is the learning rate and X is the number of visual units of the views (which are the number of pixels or number of vPNs). The novelty of a view was then given by the equation below,

$$v^{\text{nov}} = \sum_{i=0}^{X-1} |a_i|, \tag{5.5}$$

while the familiarity was calculated similarly to PM, by using Eq. (5.1).

Willshaw network

The above models performed well overall in VPR but they were not constrained by the structure of the MBs. This is in contrast to the model proposed by Wessnitzer et al. (2012), which was a much better structural approximation of the MBs. This spiking neuronal model was used by Ardin, Peng, et al. (2016) and Zhu, Mangan, and Webb (2020) for VPR and achieved decent performance (though not as good as PM or IM). Wessnitzer et al. (2012) suggested that a firing-rate neural alternative to this model could be the one proposed by (Willshaw, Buneman, and Longuet-Higgins, 1969; as it shares some structural and functional properties) and therefore it was named the *Willshaw network* (WN). This three-layer neural network included vPNs (visual input), KCs (sparse encoding), and a single *mushroom body output neuron* (MBON) (novelty output). Sun, Yue, and Mangan (2020) used the WN to calculate the novelty (and as a result the familiarity) of views on a route. They demonstrated that the familiarity value could build a gradient around the trained route; a simulated ant (equipped with their model) could locate the route by following (ascending) this gradient, which was implemented by driving left-wards steering through the CX. Goulard et al. (2021) also used the WN, where the MBON response took a reinforcing role for the FBs of the CX. They specifically suggested that the familiar views in association with self-motion feedback (through neurons in the *noduli* neuropils) build up an overall preference for a direction with respect to the learnt views.

5.3 RESULTS

In this section, the IC is challenged with the VPR task. There are three significant advancements in visual pre-processing (upstream of the IC): (1) a more realistic insect eye rendering model, (2) a principal component encoding in the vPNs, and (3) a combinatorial sparse coding in the KCs. As it is demonstrated below, these visual processing steps had an indirect effect on the function of the IC in this task.

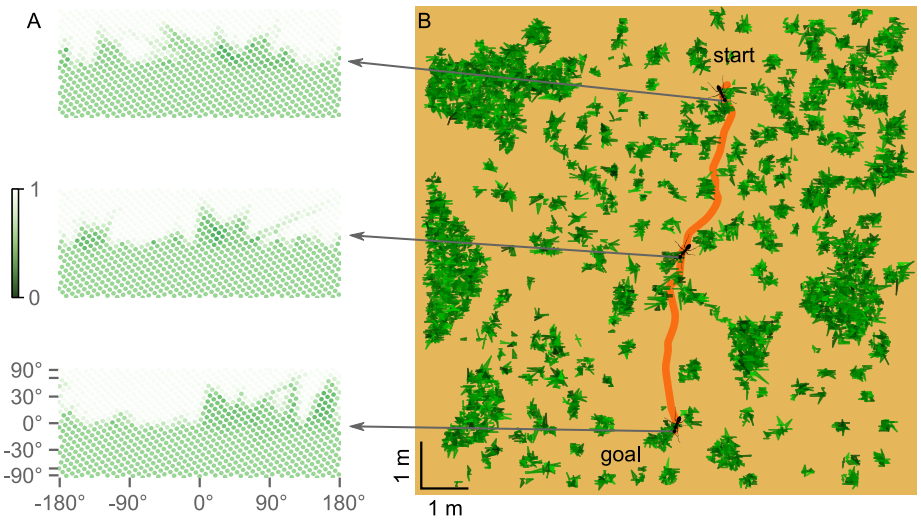

Figure 5.3: Overview of the *visual place recognition* (VPR) task. (A) Responses of the photoreceptors from the ommatidia projected onto the Cartesian space with azimuth on the horizontal and elevation on the vertical axis. (B) The ant-world that was used to test the VPR task. The green triangles represent the vegetation, the orange line is the route used for training, and the small ants represent the position and orientation where each of the sample views was captured.

Due to the unwrapping process of a spherical to a two-dimensional representation, panoramic image-like rendering techniques allocate more pixels to the poles of the sphere compared to the horizon. Subsequently, the visual processing algorithms process more data from the dorsal and ventral areas of the eye, which contain mostly information regarding the sky and ground rather than the vegetation and visual surroundings. A common fix to this problem is to keep only the pixels closer to the horizon (Baddeley et al., 2012; Ardin, Peng, et al., 2016; Möel and Wystrach, 2020; Sun, Yue, and Mangan, 2020), but this results in completely neglecting information coming from the poles. On the other hand, a more realistic compound eye renders the environment with a focus on regions of interest (depending on the needs of each animal), which better approximates the visual perception of the insects. A good example of why this might be useful is the polarised-light sensor described in Appendix D (also in Gkanias, Risse, et al., 2019), which demonstrates that considering the eye structure could provide significant insights for the underlying mechanism. Here, the simplest form of this eye was used, which is a homogeneous sampling of the environment (see Fig. 5.2B for the eye structure and Fig. 5.3A for examples of the rendering), whose detailed description and implementation is available in Section 5.5.2. The homogeneous distribution of ommatidia was chosen to avoid any correlation between the results presented in this section with more sophisticated distributions. These include both distributions with a higher concentration of ommatidia on the poles and on the horizon. The disadvantage of this rendering method compared to image-based

ones is that most common feature extraction methods—such as *lateral inhibition* (LI) or ZMs—cannot be applied to arbitrary distributions of ommatidia without modification (see Appendix B and C).

Using this ommatidial rendering method, the efficiency of the decorrelation in the vPNs and the sparse code in the KCs was analysed based on views captured during 16 different routes in a virtual world (illustrated in Fig. 5.3B). Scene familiarity was then calculated and compared using PM, the WN, and the IC.

5.3.1 *Decorrelation of the visual input*

Following the suggestion that the PNs represent an uncorrelated signal of the sensory input (Laurent, 2002), their function was approximated by applying *principle component analysis* (PCA) whitening. For a vector made up of random variables (for example, a vector of ommatidial responses), *whitening* or *sphering* decorrelates its elements by multiplying the vector by a known covariance matrix. The covariance matrix of the resultant vector becomes the identity matrix; that is, the resultant vector has properties similar to white noise. The vPNs are typically represented by a vector of ommatidial responses; by whitening this vector, a new population emerges and is used as the vPN responses—named the *principle component projection neurons* (PCPNs). These neurons play a dual role in visual pre-processing: to whiten (or decorrelate) the visual input and to reduce the number of neurons (or the *population size*) needed to encode the signal. As the population grows, so does the computational complexity of later processing stages. Therefore, the best solutions usually are based on the trade-off between the smallest population size and the lowest correlation.

Low correlation among vPN response patterns means reduced overlap between similar views, which reduces the likelihood that a novel image will appear familiar (and vice versa). Compared to the raw responses from the photoreceptors in the ommatidia, PCPNs significantly dropped the overall correlation among views that were collected in the same route. Fig. 5.4 shows that the Pearson correlation coefficient dropped significantly when at least 50 PCPNs were used (yellow line) compared to when the raw ommatidial responses were used (blue line). Each PCPN encodes a linear combination of ommatidial responses; this combination acts as a feature that can be used for VPR. The smaller the PCPN population, the greater the correlation between PCPN responses. This means that smaller populations of PCPNs will share more features across different views, while larger population sizes will improve view separability. Concisely, 50 PCPNs provided a sufficiently low correlation between different visual patterns without making the problem computationally intractable (in later processing stages).

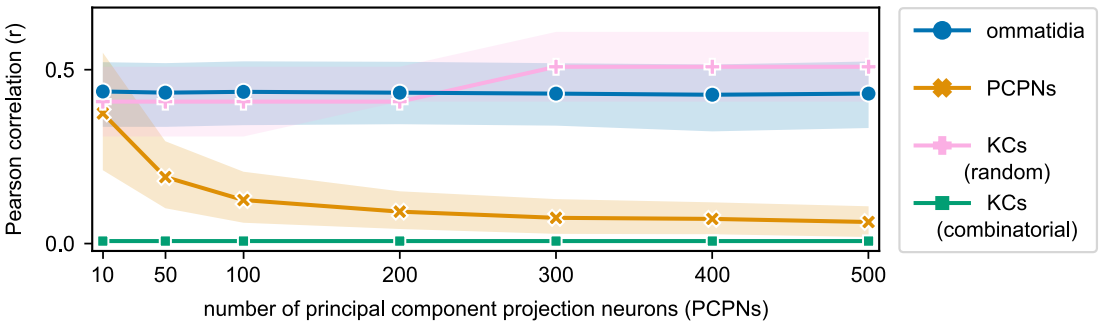

Figure 5.4: Pearson correlation between 50 pairs of views randomly selected from each of the 16 routes tested. The vertical axis shows the (positive) correlation coefficient (r) between views as encoded by the different layers of visual pre-processing. The horizontal axis shows a range of *principal component projection neuron* (PCPN) populations used to represent the views. The lines represent the correlation between response patterns to the views in the different pre-processing layers: ommatidial responses (blue), PCPN responses (yellow), *Kenyon cell* (KC) responses when randomly connected to PCPNs (pink), and KC responses when connected to PCPNs with synaptic weights constructed using the proposed combinatoric approach (green). The shaded areas mark the $[0.25, 0.75]$ quantiles, while the solid lines show the median (0.50 quantile).

5.3.2 *Efficient sparse coding and combinatorial optimisation*

The random sparse connectivity (suggested by Caron et al., 2013; and used by Ardin, Peng, et al., 2016; Sun, Yue, and Mangan, 2020) resulted in poor separability of visual patterns observed along a route (Fig. 5.4—pink line). Therefore, a different approach was followed, based on the suggestion that PN $\rightarrow$ KC connections are set up in a combinatorial fashion (Litwin-Kumar et al., 2017); a heuristic model was developed to create this structure. In mathematics, a *combination* refers to a distinct unordered subset of elements from a set of finite elements. In the context of synaptic connections, the method should use the knowledge that each KC is connected to 1-6 PNs and the objective is to find $n_{\text{kc}} = 4,000$ unique subsets of vPN patterns that share the least information (minimum correlation). Although this problem could trivially be solved by exhaustive search, usually it is impossible to find a solution in polynomial time (Bjorndal et al., 1995). For this reason, a heuristic method that finds an acceptable solution was developed. First, this method produces $6,000$ (which is, $1.5 \cdot n_{\text{kc}}$) unique patterns of one to six active PNs; these represent different patterns of PN $\rightarrow$ KC synaptic connections. Then, it computes the correlation among all these patterns and keeps the $4,000$ (n_{kc}) with the least cross-correlation.

The combinatorial sparse coding of the KCs reduced the correlation among the visual inputs along a route to almost zero for any number of PCPNs (Fig. 5.4, green line; mean Pearson correlation coefficient, $r = 0.0025$). This is reasonable as only ten KCs (out of four thousand) were active for each view and it suggested that for

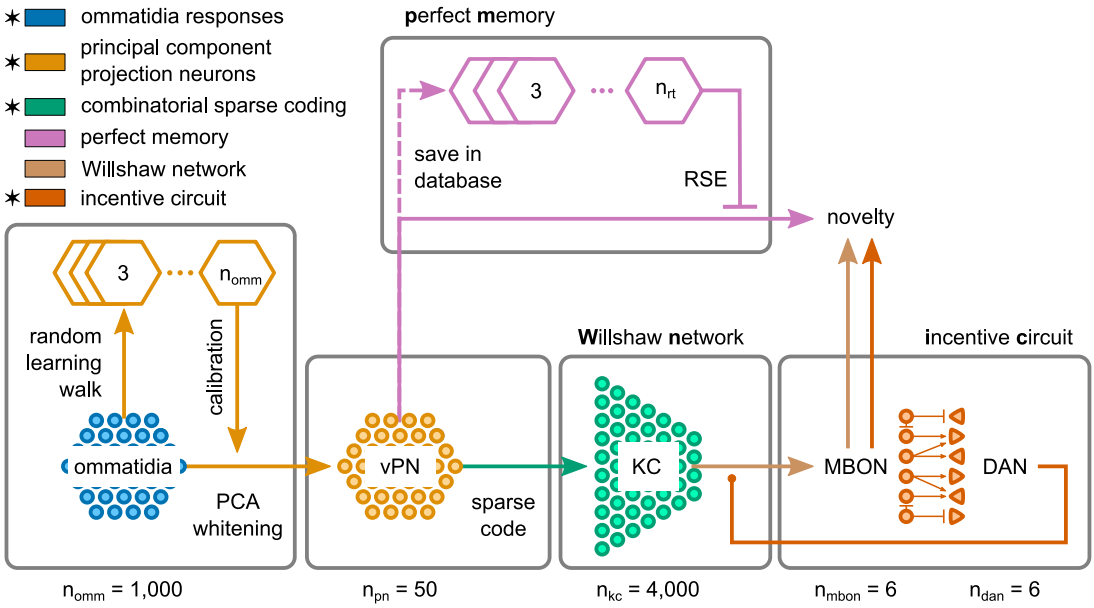

Figure 5.5: Summary of the *visual place recognition* (VPR) contributions of this work (with star, '*'). First, there is the rendering of the environment by using the ommatidia of the compound eye instead of the camera-like input (in blue). During an initial learning walk, the *visual projection neurons* (vPNs) were calibrated into the *principal component projection neurons* (PCPNs), which built a transformation that remained static after the initialisation (in yellow). The connections between the vPNs and *Kenyon cells* (KCs) were calculated based on a heuristic combinatorial solution and they remain static throughout all the experiments (in green). The *perfect memory* (PM), *Willshaw network* (WN), and *incentive circuit* (IC) were 'trained' with a sequence of images along a route. The PM computes the novelty by comparing the vPN responses to previously experienced vPN responses (in pink). The WN calculated the novelty based on the *mushroom body output neuron* (MBON) responses (in brown). The IC uses MBON → DAN (*dopaminergic neuron*) feedback connections to increase the response when the familiarity is consistent (in red).

each pattern only a minority of active KCs overlapped if any. In contrast, when using the random PN → KC connections and ten active KCs, many of the KCs were apparently overlapping (mean Pearson correlation coefficient, $r = 0.3985$). This further supported the argument that sparse connections constructed using a combinatoric approach could improve the separability of patterns where a random approach could not (Litwin-Kumar et al., 2017). Section 5.4.2 discusses some of the advantages and disadvantages of highly uncorrelated patterns in this context.

5.3.3 *Predicting the familiarity*

In the context of the VPR, the role of the MBONs is to predict how familiar (or conversely, novel) a view is (see Fig. 5.5). In order to test this, the IC was trained and tested on recognising simulated 'views' of desert ants (see Fig. 5.3A for sample views). These views were captured while following a route to the nest from a distant

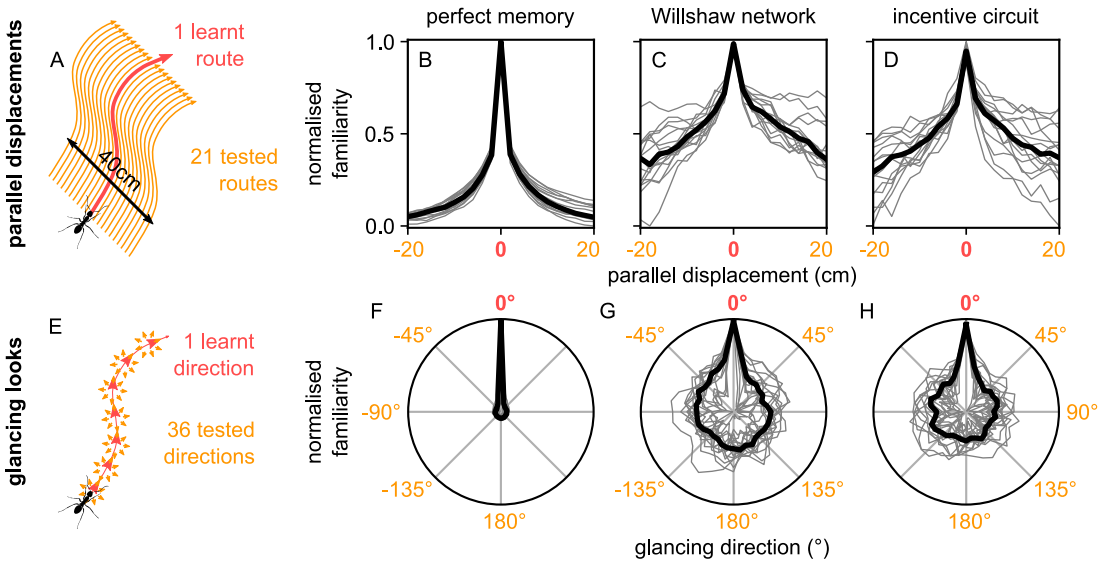

Figure 5.6: Comparison of the familiarity predictions among the models, for the *parallel displacements* and *glancing looks* tasks. (A) Schematic description of the parallel displacements task. Each model is trained and tested on the same route, and additionally tested on 20 shifted copies of this route for up to 20 cm to the left or right. This experiment was repeated for 16 routes, which were recorded from different desert ants; the average distribution of familiarity across the displacements is illustrated when using (B) the *perfect memory* (PM), (C) *Willshaw network* (WN), or (D) the *incentive circuit* (IC). (E) Schematic description of the glancing looks task. Each model is trained on a route (like before) and tested on the same route but changing its heading direction by 36 regular intervals, starting from the original direction. The average distribution throughout each route was then plotted for (F) PM, (G) WN, and (H) IC.

feeder (around 8 m away from the nest) as shown in Fig. 5.3B. The routes used for these results were captured by Mangan and Webb (2012), who tracked desert ants (of the species *C. velox*) in their natural environment (not part of this thesis). After training the model on a single route, two different experiments were performed: one that looked at parallel displacements and another involving glancing looks. These were repeated for sixteen routes performed by fifteen different ants. For comparison, the outputs of PM (Fig. 5.6B and F) and WN (Fig. 5.6C and G)—adapted from Ardin, Peng, et al. (2016)—were presented alongside those of the IC (Fig. 5.6D and H). Fig. 5.5 summarises the information processing used by the three methods, highlighting both their shared and unique characteristics.

In the *parallel displacement* experiment, the models were trained on a route and tested on 21 copies of the same route displaced by up to 20 cm to the left or right with respect to the overall direction of the route (see Fig. 5.6A). As expected, for all three models the familiarity dropped proportionally to the distance from the original route. The familiarity estimated by PM (Fig. 5.6B) dropped off quickly (displacement < 10 cm), while the WN (Fig. 5.6C) and the IC (Fig. 5.6D) was more robust to

displacements (see also grey lines for the individual routes). This suggested that the tiny correlations among the sparse codes of the KCs were sufficient to capture similar features in less proximal views, allowing for a smoother familiarity distribution around the learnt route. However, note that the familiarity distributions over the displacements as predicted by the WN and IC were very similar. This suggested that the MBON → *dopaminergic neuron* (DAN) connections included in the IC did not significantly impact familiarity estimation (compared to the computationally simpler WN), implying that the memory dynamics they introduce may be redundant for this task.

In the *glancing looks* experiment, the models were trained on a route (Fig. 5.6E, red line) and tested on the same route, but looking at 36 different deviations from the original direction. The outcome was similar to the parallel displacement experiments, though the familiarity distributions were narrower. This was because directional deviations usually result in larger visual changes compared to parallel displacements (see Fig. 5.6F-H). Again, PM had a very narrow familiarity distribution while the WN and IC had wider ones, meaning they were also more robust to deviation. However, this was because of the noisy distributions produced by the WN and IC for the individual routes, not because of rotational invariant features. Therefore it is expected that, in a novel place, the familiarity distribution across the viewing directions would be (on average) uniform. As before, the distributions estimated by the WN and IC looked very similar, which suggested that the added memory dynamics of the IC do not play an important role in resolving rotational ambiguity.

5.3.4 *Route following by differential familiarity*

Ardin, Peng, et al. (2016) suggested that using the estimated visual familiarity for navigation would lead the animal to navigate in a direction parallel to the one indicated by the trained route, but not towards approaching the route. This was in contrast to the behaviour observed by T. S. Collett, Graham, and R. A. Harris (2007), where desert ants approached the familiar route when familiar views were detected and followed it to find the nest. A mechanistic explanation for this behaviour was presented by Sun, Yue, and Mangan (2020); they used rotation-invariant views to drive the animal towards a familiar route (using the MB), which then switched to a rotation-specific route following mechanism (using the CX). This section tries to examine whether the rotational-specific outputs of the MB could be used for route following, overcoming the problem faced by Ardin, Peng, et al. (2016). Based on the results of Chapter 3 and 4, there are two major differences between the IC and the other two models used for comparison in the previous section (PM and the WN): (1) the IC introduces some memory dynamics, and (2) subsequent experiences of

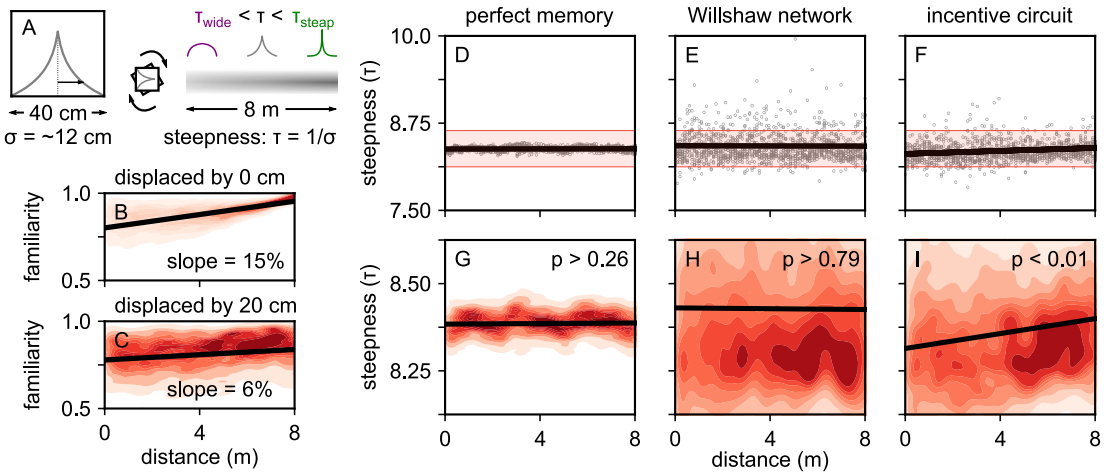

Figure 5.7: Comparison of the estimated steepness across the models. (A) Schematic description of how the steepness was calculated for these results and how to interpret it—increasing steepness means that the distribution becomes narrower, and decreasing means that it becomes flatter. (B, C) The familiarity calculated along 16 recorded routes (from 15 ants; ~ 830 data points per route) by using the *incentive circuit* (IC) when tested on (B) the training route or (C) on a 20 cm displaced route. (D, E, F) Measured steepness along 16 recorded routes from 15 ants (~ 830 data points per route, $n = 13,268$ data points in total) by using (D) the *perfect memory* (PM), (E) *Willshaw network* (WN), or (F) IC. Steepness is plotted against the distance walked, and a line is fitted (using linear regression) showing the correlation between the steepness and the walked distance. (G, H, I) Zoom in on the regions marked with red boxes in (D), (E), and (F) respectively, aiming to make the effect more obvious. The individual points are replaced by their equivalent distribution, which highlights their concentration.

the same views may result in different responses. These could allow dynamic predictions of the familiarity that increase when the views experienced are consistently on the route, but decrease when these are on a parallel route. This would allow the animal to know whether it should continue following the route or try to find a more familiar parallel route. In order to examine if this is plausible by the model, the change of the familiarity distribution along the different routes of the parallel displacements experiment was analysed further.

The familiarity distributions can be characterised by a single value: their *standard deviation* (SD)—also denoted as σ—but for this task, a more intuitive measure is the *steepness*, $\tau = 1/\sigma$, which describes how narrow the distribution is (see Fig. 5.7A). Constant steepness along a specific route (as calculated by the combination of its parallel routes shown in Fig. 5.6A) would suggest that the familiarity distribution stays roughly the same. Changing steepness would suggest a change in the familiarity distribution as illustrated in Fig. 5.7A. An overall increase in steepness along the route would suggest that the familiarity distribution becomes narrower (and vice versa). Note that for these experiments, the model was tested on each parallel route separately. This means that the MBON responses might depend both on the current view

Model	repeats	Slope	SE	χ^2_L	p
Perfect memory	1 - 5	$0.13\,\%$	$5.49 \cdot 10^{-5}$	70.43	0.2638880
Willshaw network	1 - 5	$-0.19\,\%$	$4.22 \cdot 10^{-4}$	92.81	0.7948878
Incentive circuit	1	$2.28\,\%$	$2.96 \cdot 10^{-4}$	48.52	0.0105797
	2	$3.20\,\%$	$2.78 \cdot 10^{-4}$	31.92	$5.94 \cdot 10^{-5}$
	3	$3.54\,\%$	$3.05 \cdot 10^{-4}$	15.78	$3.01 \cdot 10^{-9}$
	4	$3.66\,\%$	$2.85 \cdot 10^{-4}$	23.94	$1.08 \cdot 10^{-6}$
	5	$3.41\,\%$	$2.65 \cdot 10^{-4}$	25.67	$2.90 \cdot 10^{-6}$

Table 5.1: Summary of the steepness changes along the different routes, for the *perfect memory* (PM), *Willshaw network* (WN), and *incentive circuit* (IC). Repeats: the number of times that the training was repeated before testing the models on the parallel routes. Slope: the slope of the fitted linear regression line on the data converted to a percentage, where $100\,\%$ slope would denote an increase of the steepness from 0 to 1 along the route; a positive slope denotes an overall increase of the steepness, and negative slope an overall decrease. SE: the standard error of the slope ($n = 13,268$ samples); in this context, it represents the spread around the fitted line. χ^2_L: Page's L test statistic transformed into a chi-square distribution value with one degree of freedom; this tests whether there is an increasing trend in the steepness over the route. p: the Page's p-value that shows how consistent is the trend.

and on the state of the IC after experiencing the previous views of the same route. Therefore, increasing steepness could have two alternative explanations: either the familiarity increases along parallel routes that are closer to the one used for training, or the familiarity decreases along routes that are more distant from the training route. In both cases, the immediate change in the familiarity estimation of the animal (increase or decrease) could be an indicator of whether its action would keep the animal on route or whether its action would move the animal off route.

Each model was used to estimate a familiarity distribution around the route (PM, Fig. 5.7D and G; WN, Fig. 5.7E and H; and IC, Fig. 5.7F and I); using these distributions, the steepness over the distance travelled could be extracted. Table 5.1 summarises the regression analysis and Page's trend test on these data. PM and WN did not show any consistent changes in steepness along the routes (Page's trend test: $p > 0.26$ for the PM, and $p > 0.79$ for the WN). This suggested that the familiarity distribution was not changing by default along the route and that there was no bias towards increasing or decreasing familiarity. On the other hand, the IC predicted a small and weakly significant increase in the steepness over the route (regression slope = $2.28\,\%$, Page's trend test: $p < 0.02$). To ensure that this effect was consistent, model training was repeated up to five times before conducting parallel displacement experiments. Interestingly, the increasing steepness effect became stronger and more significant from the second repeat of the training (regression slope = $3.05\,\%$, Page's trend test: $p < 6 \cdot 10^{-5}$), but its strength was not linearly increased with fur-

ther training (regression slope = $3.41\,\%$, Page's trend test: $p < 3 \cdot 10^{-6}$, for 5 repeats; see Table 5.1). Note that a similar increase was observed in the behavioural effects of Section 4.2.1, which showed that the *preference index* (PI) saturated after three cycles of training in elemental olfactory learning. However, the small increase observed in the steepness of the familiarity distribution was due to an increase in the familiarity in all the parallel routes (see Fig. E.26), with slopes decreasing from $15\,\%$ (along the training route, Fig. 5.7B) to $6\,\%$ (at $20\,\mathrm{cm}$ distance from the training route, Fig. 5.7C). This suggested that the increasing steepness of the familiarity distribution was due to a higher increase of familiarity across the most familiar route, rather than a decrease in parallel routes. Thus, the IC introduced a novel increase in the familiarity along previously experienced routes, which was also proportional to the proximity of the point where the sample was captured to the trained route.

5.4 DISCUSSION

5.4.1 *Visual projection neurons*

In contrast to other vPN representations (like the ZMs; Stone, Mangan, et al., 2018; Sun, Yue, and Mangan, 2020; or edge detection; Goulard et al., 2021) the processing in the PCPNs was not retinotopic. However, some preliminary results showed that retinotopic methods alone or in combination with the PCPNs were less successful. For example, in the compound eye, edge detection could be approximated by *lateral inhibition* (LI)—that is, each ommatidium inhibits its neighbouring ommatidia. Appendix B showed that LI decreased the correlation of the views further, which resulted in steeper familiarity distributions for all the tested models. In addition, Appendix C showed that customising the ZMs for the homogeneous distribution of ommatidia used in this work, lacked the rotation invariant features described before (Stone, Differt, et al., 2016; Stone, Mangan, et al., 2018; Sun, Yue, and Mangan, 2020). This suggested that the previously observed rotation-invariant features of ZMs probably rely on the array-like structure of panoramic images. Nevertheless, the rotational and translational familiarity distributions estimated by PM when using ZMs were shallower than the ones estimated when using the PCPNs, showing higher tolerance to small rotations and translations. Although the direct comparison of the extracted features might provide a good estimate of the familiarity (as demonstrated by using PM), they were harder to process by the WN and IC, which suggests that the extracted features were highly correlated. Although this was not tested explicitly, it is possible that this problem would not be solved even after applying additional decorrelation to the ZMs (for example, through an extra layer of PCPNs).

The responses of PCPNs were calculated by transforming the raw photoreceptor responses of the ommatidia with a pre-computed transformation matrix. This matrix was computed in an unsupervised fashion by using samples of the photoreceptor responses during the exploration of the environment. This process is called *calibration*, and it might be realised in insects during the early stages of their life (for example, during the transition from inner workers to forgers in desert ants). Alternatively, the calibration could be a continuous process, which constantly updates the representation in vPNs to minimise the correlation among the views of the animal in changing environments. Although at the beginning of their life this might result in relatively big changes in the representations, later this should only happen during big environmental changes. In support of this suggestion, there is evidence that the MB calyx of desert ants substantially increased in size when they were first exposed to light (Kühn-Bühlmann and Wehner, 2006).

This was the first time that the uncorrelated activity of the PNs was explored computationally for the VPR task in MB models. Here, the PCA whitening technique was used to explore this property (see Section 5.5.3), but more sophisticated ways to extract uncorrelated features could be explored in the future. For example, a method that applies non-linear or rotational invariant whitening among the different views could be used (for example, graph-based whitening; Aaron, 2006). Alternatively, a method could search for common features between views experienced in spatial or temporal proximity. This was partially explored by Zhu, Mangan, and Webb (2020) with an event-based camera (exploring temporal proximity), but no methods were found to directly associate spatial proximity to views.

5.4.2 *Sparse coding by the Kenyon cells*

The combination of sparse coding and the increased separability of the representations by the PCPNs resulted in further advances introduced by the KCs. One of those was its advantage over the commonly suggested random connection. Fig. 5.4 suggested that unconstrained random connections do not result in an efficient sparse code. In this work, unconstrained random connections were achieved by drawing a random number of connections for each pattern (in the range $[1, 6]$) and then assigning these connections randomly (without considering the patterns generated earlier). This is different from sampling from a pool of synaptic patterns, where each pattern has an equal probability of occurrence. In fact, in the second case, the result could sometimes come to a similar solution to the one suggested by the proposed heuristic method (Section 5.3.2). Note, however, that in most previous works where 'random connections' were generated, they were also trying to limit the number of overlapping

patterns (although not directly trying to keep their correlation as low as possible), which is different from both the unconstrained random and heuristically combinatorial connections used in this work.

Another, common advance of the sparse coding is the increased memory capacity of the overall circuit (as demonstrated by Ardin, Peng, et al., 2016; Litwin-Kumar et al., 2017). In the context of the MB, memory capacity is defined to be proportional to the maximum number of vPN patterns that can be stored in the KC $\rightarrow$ MBON connections. For a single MBON, this depends both on the total number of KCs and on the percentage of them being active at the same time (spatial sparseness). Given a fixed number of KC $\rightarrow$ MBON synaptic connections, every time that a new KC pattern is learnt, the respective KC $\rightarrow$ MBON synaptic weights are changed (predicting familiarity). The more patterns it learns, the more synaptic weights are changed, consuming the memory. Increased sparseness (a lower proportion of KC being active for each pattern) results in weight change for fewer synapses per pattern, resulting in less memory consumption per learnt view.

However, overly increased sparseness would result in either completely correlated or completely uncorrelated patterns (similar to the effects of one-hot encoding). One hot encoding is a popular encoding technique for categorical values in *machine learning* (ML), and it is a vector of zeroes where only one element is equal to one. In this extreme scenario (where only 1 KC responds in each visual input and each visual input results in the response of a different KC) the KC representation of each visual input is completely uncorrelated to all the other inputs. This would result in $100\,\%$ familiarity for the learnt inputs and a $0\,\%$ otherwise. When all KCs are occupied and a new input needs to be learnt, it would completely overlap with an occupied KC, creating $100\,\%$ correlation with the visual input previously associated with this KC. Associating two KCs per visual input would result in a larger range of combinations of KC identities, but it would also sacrifice the binary predictions of familiarity. This is because there is always some correlation to other visual inputs sharing the response of one of the two KCs. Such a correlation could be particularly useful if it described a feature that is shared between related visual inputs, but not between unrelated ones. This would save some space (lower consumption of the memory) and also allow generalisation among the visual inputs. Therefore, it is reasonable to think that the purpose of the processing between the PNs and KCs is to optimise the balance between the sparseness and correlation of the visual inputs.

Although fixed PN $\rightarrow$ KC synaptic weights were assumed, the plasticity of these connections has been experimentally supported (Hammer, 1993; Szyszka, Galkin, and Menzel, 2008) and computationally explored (Finelli et al., 2008; Z. Wu and Guo, 2011; Arena et al., 2013; Peng and Chittka, 2017; Faghihi et al., 2017), suggesting that

the visual representation in the KCs is crucial for the VPR task. Axo-axonic KC → KC connections (Eichler et al., 2017) could also play a role in this, and it is possible that correlations are first detected by these connections before they are used for the PN → KC synaptic modulation. The function of KC → KC connections has not yet been explored with respect to the above mechanism. However, Zhu, Mangan, and Webb (2020) explored a different hypothesis, suggesting that these connections relate to time-dependent correlations. By using a computational model, they showed that the MBON responses depended on whether the views were experienced in the correct order, which was also supported experimentally (Schwarz, Mangan, et al., 2020).

5.4.3 *The effect of consistent familiarity*

In the context of VPR, the MBON responses are translated to the familiarity or novelty of the incoming views that are represented by the responses of the KCs. The MBON responses predicted by the WN seemed sufficient to produce a gradient of familiarity around the learnt route showing high translational and rotational sensitivity. The MBON → DAN feedback connections (through the IC) seemed to have a minor contribution to the overall performance of the circuit predicting the familiarity. However, these connections introduced dynamic changes in the familiarity that depended on the time spent in familiar (or unfamiliar) ground. When the circuit was repeatedly exposed to familiar views the familiarity tended to increase and this effect was proportional to the proximity of the parallel route to the trained one (see Fig. 5.7B, C and Fig. E.26). This effect became stronger after multiple training cycles on the same views (see Table 5.1) indicating that it was affected by learning.

Fig. 5.7B and C suggested that the familiarity increases for consequent familiar views and that this increase is sharper and more consistent on the learnt route compared to the parallel displacements; this explains the observed increase in steepness of the familiarity distribution (shown in Fig. 5.7I). This effect is related to consistent familiar predictions that imply an integration over time, or knowledge of the relative order between views. It is not very plausible that the IC captures the order between learnt views, as a more sophisticated mechanism should be in place, such as differential dynamics in the visual input (Zhu, Mangan, and Webb, 2020). It is more plausible that repeatedly presented familiar views cause the increase in the steepness, which implies that even shuffled familiar views would have the same effect. However, assuming that the familiarity increases over time due to the MBON → DAN feedback connections, views that are closer to the nest are more reinforced that those away from it. This suggests higher familiarity predictions for views closer to the nest, which build a gradient of familiarity along the route as well as around it.

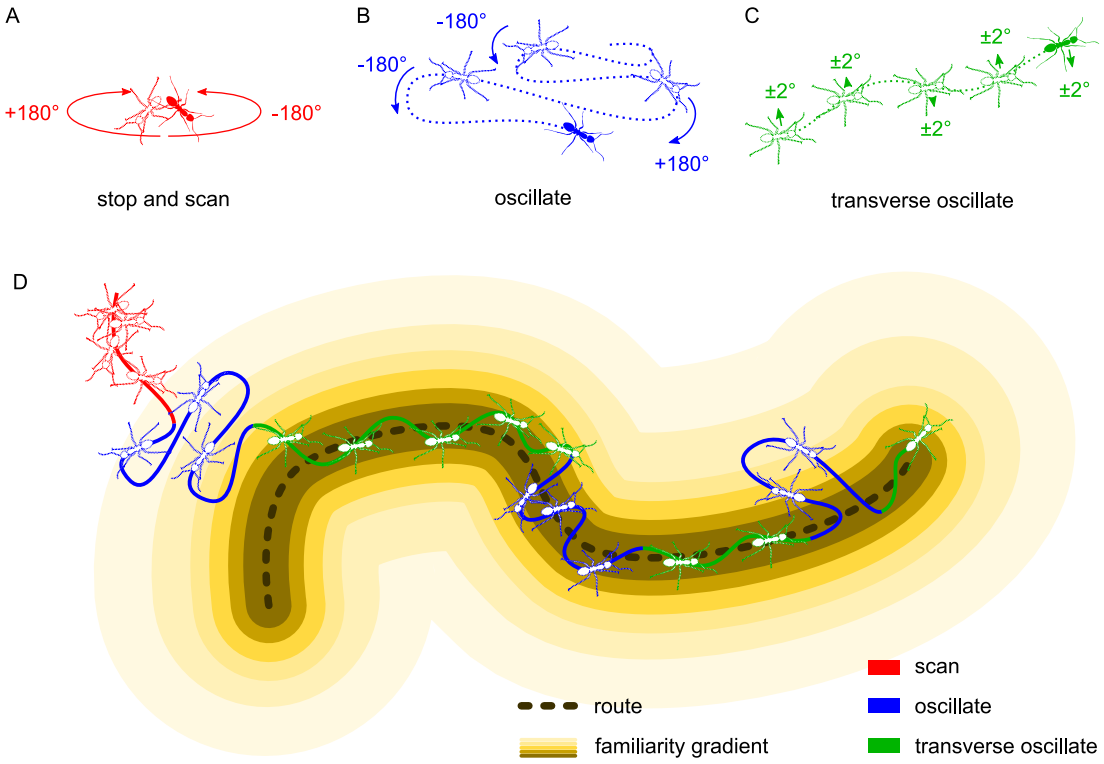

Figure 5.8: Hypothetical model of how the familiarity predicted by the *incentive circuit* (IC) can control three *visual place recognition* (VPR) behaviours for route following. (A) The *stop-and-scan* behaviour of desert ants. When in unfamiliar ground, ants stop and scan for almost 360°, followed by a saccadic motion towards an apparently random direction. (B) The *oscillating* behaviour is performed in partially familiar ground. Wide oscillations towards more familiar ground allow the ants to simultaneously scan and move slowly with small corrections, which keeps them on relatively familiar ground. (C) *Transverse oscillations* are performed on highly familiar ground and allow the ants to track the peak of familiarity without changing their direction. (D) Example of how the different VPR behaviours interplay during route following. Lower familiar grounds enable the stop-and-scan behaviour (red trail). Increasing familiarity enables the oscillating behaviour that leads the ant to cross the familiar route (blue trail). Consistently increasing familiarity notifies the ant that it is on the route, enabling the transverse oscillating behaviour (green trail). Consistent decrease in the familiarity gears down the ant, enabling the oscillating behaviour, and so on.

The different rates of increasing familiarity for the parallel routes (as shown in Fig. E.26) could provide the basis for a navigation strategy of the animal, specifically, route following. Stone, Mangan, et al. (2018) and Sun, Yue, and Mangan (2020) suggested that the familiarity should be rotationally invariant (which means that it should not change based on the direction faced) and that it should create a manifold of familiarity that can be used by insects to cross familiar routes. Once the familiar route is crossed, a different pathway of visual input (external to the MB) provides the directional information that allows the animal to return to the nest. The PCPN, KC, and IC models used in this thesis suggested that the tolerance in orientation per-

turbations with respect to familiarity estimations is limited (see Fig. 5.6F-H), even when using the ZMs (see Appendix C). This is in line with the observation that ants scan their environment (Kohler and Wehner, 2005; Schwarz, Clement, et al., 2020) or oscillate (Murray et al., 2019; Clement, Schwarz, and Wystrach, 2022) systematically when they are lost (illustrated by Fig. 5.8A, B). A reasonable hypothesis would then be that once they find the direction with the highest familiarity (for example, by scanning), they fixate and move towards it with transverse oscillations (which is, side-walking with a minor change in the visual direction, as illustrated by Fig. 5.8C). This technique was successfully employed by Stankiewicz and Webb (2021) using a drone that mimicked honeybee route following (taking inspiration from wasp homing behaviour, Stürzl, Zeil, et al., 2016). Transverse oscillations could allow the animal to sample the familiarity distribution and drift its position to the centre of the distribution without changing its direction. An overall increase in familiarity would reassure the animal that it is moving along a familiar route. A decrease in the overall familiarity would alert the animal that it is moving away from the route, causing its speed to drop and start bigger oscillations (medium drop in speed) followed by scans (zero speed) until the familiarity starts increasing. This hypothetical model assumes that familiarity is a relative term, which is not always the same for a given view, but it depends on previous (short- or long-term) experiences of the animal. The results presented in this chapter suggested that the IC allows for this relativity, suggesting that the familiarity manifold shown in Fig. 5.8D would be in constant flux, along with the way that the animal interprets the environment.

5.4.4 *Integration with the central complex*

The hypothetical model described in the previous section could be realised through the CX, as illustrated in Fig. 5.9A. This suggests that the MBONs of the IC converge in a single neuron of the *convergence zones* (CZs)—most likely in the *crepine* (CRE). The three MBONs that predict familiarity should excite that neuron and the other three (that predict novelty) inhibit it. The familiarity is also propagated to another interneuron of the CZ that introduces a delay in that signal. The two interneurons converge to a single FB tangential neuron, which encodes differential familiarity with respect to time. This neuron should be active when the familiarity is increasing and inactive otherwise. Similarly, another FB tangential neuron could have the opposite effect and encode decreasing familiarity. The FB tangential neurons connect to all the columns of the FB and mainly on the fourth layer, providing motivational information and driving the behaviour of the animal (Hulse et al., 2021).

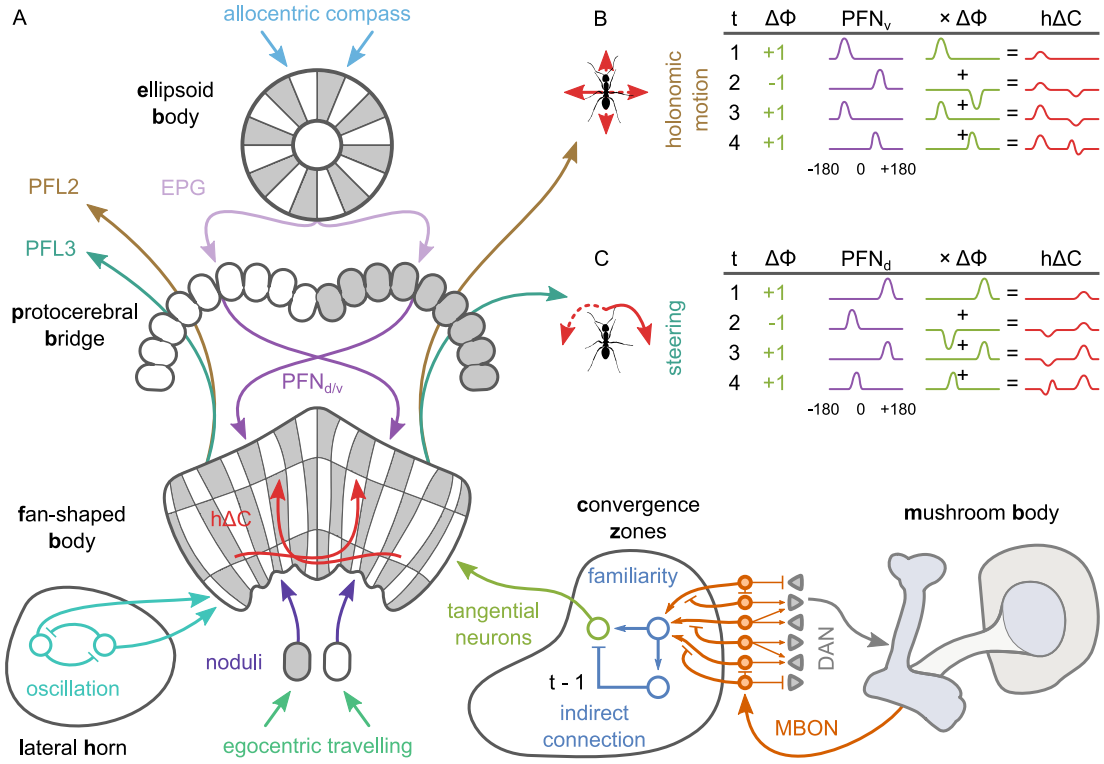

Figure 5.9: Integration of the *central complex* (CX) with the differential familiarity model and the *visual place recognition* (VPR) behaviours. (A) The overall circuit. Allocentric compass information enters the CX from the *ellipsoid body* (EB) and propagates all the way to the *fan-shaped body* (FB) through the *protocerebral bridge* (PB). Egocentric travelling information is provided by the *noduli*. *Lateral horn* (LH) provides the innate VPR behaviours in the FB, while the *mushroom bodies* (MBs) gate these behaviours through the tangential FB neurons ($\Delta\Phi$), which encode the differential familiarity information. The integrated behaviours encoded in the FB produce steering and holonomic motion commands, which are processed by the *lateral accessory lobes* (LALs). (B) Example of how the holonomic motion is integrated into the hΔC neurons of the FB. (C) Example of how the steering is integrated into the hΔC neurons of the FB.

Allocentric directional input enters the CX from the *ellipsoid-body* (EB). Gkanias, Risse, et al. (2019) and Appendix D suggested a model that implements such input through a celestial compass, which transforms the skylight input into responses in the EB. It has been suggested that this compass is integrated with other compasses (related to different sensory modalities like wind, magnetic, and visual landmarks) in the *ring neurons* prior to the EB (Hulse et al., 2021). The *ellipsoid-body protocerebral-bridge gall* (EPG) neurons form a sinusoidal pattern in the EB, whose phase encodes the heading direction and its amplitude encodes the certainty of the compass cue. This is integrated in the *protocerebral bridge* (PB) with egocentric travelling feedback from the *noduli*, which allows keeping track of the heading direction even without allocentric input. This circuit was computationally explored by Pisokas, Heinze, and Webb (2020), who demonstrated that it can effectively track the heading direction

of the animal in darkness. The heading direction is propagated to the FB through the *protocerebral-bridge fan-shape-body noduli* (PFNs) columnar neurons, whose activity further integrates egocentric travelling information from the noduli. These converge to the *protocerebral-bridge fan-shape-body lateral-accessory-lobe 3* (PFL3) neurons and control the steering commands. Stone, Webb, et al. (2017) demonstrated that the PFNs can efficiently integrate the heading direction and egocentric travelling information (from the noduli) into a *path integration* vector that can be used to return home; they further suggested that the FB can perform calculations among vectors (Moël et al., 2019).

Lyu, Abbott, and Maimon (2022) supported experimentally the above hypothesis in fruit flies. They verified that interneurons of the FB (the hΔB neurons) can invert, add, and subtract vectors, but they also allow the encoding of travelling direction independently of the actual heading. This was also picked by Matheson et al. (2021), who specifically tested how the different MBONs affected the responses of the hΔC neurons through the tangential FB neurons. They showed that the MB and *lateral horn* (LH) output neurons can gate how the PFNs and hΔCs affect the activity of PFL3 and *protocerebral-bridge fan-shape-body lateral-accessory-lobe 2* (PFL2) neurons, controlling the steering and holonomic motion respectively. Note that, the LH is usually hypothesised to provide the innate behaviour of the animal, which sets the groundwork for the following hypothesis.

It is reasonable to assume that the three VPR behaviours described in the previous section (which are, the stop-and-scan, oscillate, and transverse oscillate; see Fig. 5.8) could be implemented by the LH (innate behaviour) and the *lateral accessory lobe* (LAL) (innate oscillations; Clement, Schwarz, and Wystrach, 2022). A combination of LH and LAL output neurons could enable parallel navigational strategies in the FB (potentially through different layers) by regularly shifting the activity bump of specific groups of PFN neurons. The stop-and-scan behaviour would shift the bump of the PFN_d neurons (which manipulate the steering commands), the transverse oscillations would affect the PFN_v neurons (which correspond to holonomic motion), and the oscillating behaviour would affect the bumps of both groups of neurons (which would enable both holonomic motion and steering). In addition, the activity of the tangential neurons that receive the differential familiarity input could be multiplied with the different PFN populations, while the hΔC neurons could integrate the resultant activity. Therefore, the tangential neuron that encodes the increasing familiarity would positively gate the PFN_v activities and negatively gate the PFN_d activities before they are integrated by the hΔC neurons; the tangential neuron that encodes the decreasing familiarity would do the opposite (see Fig. 5.9B and C). Note that this could be an alternative solution to the integration in PFL3s previously suggested by

Goulard et al. (2021). The above model would allow for a smooth interplay among the behaviours and it would provide a reasonable hypothesis of how the familiarity computed by the MB could contribute in the complex navigation system of insects through the CX.

5.5 METHODS

5.5.1 *The simulated environment*

The simulated environment was built following the approach of Ardin, Peng, et al. (2016). The vegetation of the environment was represented by the location, size, and colour of 5,000 three-dimensional triangles, and the two-dimensional position and orientation of 15 desert ants (*C. velox*) formed 133 different routes (Mangan and Webb, 2012). This data is made publicly available by the Ant Navigation Challenge[1]. The world and a sample route of one desert-ant are illustrated in Fig. 5.3C. As the recorded routes provided only two-dimensional information, the eye height of the ants was assumed to always be at $1\,\mathrm{mm}$ height from the ground, and its pitch and roll orientation to always be zero (which means that the head was aligned to the horizon).

5.5.2 *Rendering the environment*

The compound eye model had $n_{\mathrm{omm}} = 1,000$ ommatidia, homogeneously distributed on a sphere using the Fibonacci series (Keinert et al., 2015). The elevation, e_ι, and azimuth, a_ι, of each ommatidium, ι, was calculated as follows,

$$e_\iota = 90° - \cos^{-1}(2\,\frac{\iota}{n_{\mathrm{omm}} - 0.5} - 1), \qquad \iota \in \{0, ..., n_{\mathrm{omm}} - 1\}, \tag{5.6}$$

$$a_\iota = (1 + \sqrt{5})\,\iota\,180° \mod 360°, \qquad \iota \in \{0, ..., n_{\mathrm{omm}} - 1\}, \tag{5.7}$$

where the $\mod$ operation returns the remainder of a division (which is also known as the *modulo*), $e_\iota \in [-90°, +90°]$, and $a_\iota \in [0°, 360°)$. Note that $\varphi = \frac{1+\sqrt{5}}{2}$ is the golden ratio, which is connected to the Fibonacci numbers. For ease of computations, the above spherical coordinates were transformed into *quaternions*, which prevent the

1 insectvision.dlr.de/walking-insects/antnavigationchallenge (last visited: July 2022)

gimbal lock problem of the Euler angles and spherical coordinates. This was done by using the equations below,

$$w_\iota = \cos 45^\circ \cos\frac{e_\iota}{2}\cos\frac{a_\iota}{2} + \sin 45^\circ \sin\frac{e_\iota}{2}\sin\frac{a_\iota}{2}, \tag{5.8}$$

$$x_\iota = \sin 45^\circ \cos\frac{e_\iota}{2}\cos\frac{a_\iota}{2} + \cos 45^\circ \sin\frac{e_\iota}{2}\sin\frac{a_\iota}{2}, \tag{5.9}$$

$$y_\iota = \cos 45^\circ \sin\frac{e_\iota}{2}\cos\frac{a_\iota}{2} + \sin 45^\circ \cos\frac{e_\iota}{2}\sin\frac{a_\iota}{2}, \tag{5.10}$$

$$z_\iota = \cos 45^\circ \cos\frac{e_\iota}{2}\sin\frac{a_\iota}{2} + \sin 45^\circ \sin\frac{e_\iota}{2}\cos\frac{a_\iota}{2}, \tag{5.11}$$

where $q_\iota = w + ix_\iota + jy_\iota + kz_\iota$ is the quaternion that represents the orientation of the respective ommatidium, with one real (w_ι) and three complex numbers $(ix_\iota, jy_\iota,$ and $kz_\iota)$. Note that $i^2 = j^2 = k^2 = ijk = -1$ is an important property of the different imaginary numbers in quaternions.

Each ommatidium was assumed to have an acceptance angle of $10°$. This was approximated by sampling from a Gaussian density function with *standard deviation* (SD), $\sigma = 5°$. Seven samples were used for each ommatidium, one in the centre and six homogeneously distributed in $5°$ radius around the centre. The quaternionic distance of each sample from the centre of each ommatidium was given by the equation below,

$$q_\xi^\sigma = \cos\frac{\sigma}{2}\cos(60°\xi) + i\sin\frac{\sigma}{2}\sin(60°\xi) + j\sin\frac{\sigma}{2}\cos(60°\xi) + k\cos\frac{\sigma}{2}\sin(60°\xi), \tag{5.12}$$

where $\xi \in \{1, ..., 6\}$. This exempts the $\xi = 0$ sample (which points in the central direction), whose quaternionic distance from the centre is described as $q_0^\sigma = 1$. The rendering orientation of each sample with respect to the orientation of the whole eye can then be computed as $q_{\iota,\xi} = q_\iota \cdot q_\xi^\sigma$, where $\cdot$ denotes quaternion multiplication.

The global orientations of the compound eyes of the recorded ants can also be transformed into quaternions by using their *yaw* orientation, ϕ_{eye}, and the equation below,

$$q_{\text{eye}} = \cos\frac{\phi_{\text{eye}}}{2} + k\sin\frac{\phi_{\text{eye}}}{2}. \tag{5.13}$$

Thus, the global orientation of the sampling directions is given by $q_{\iota,\xi}^{\text{glob}} = q_{\text{eye}} \cdot q_{\iota,\xi}$.

During rendering, ommatidial samples were initialised as follows,

$$r_{\iota,\xi}^o = \begin{cases} 0.4, & \text{if } e_{\iota,\xi}^{\text{glob}} \leq 0 \quad \text{(the ommatidium points towards the ground)}, \\ 1.0, & \text{otherwise}, \end{cases} \tag{5.14}$$

where $e^{\text{glob}}_{\iota,\xi}$ is the global elevation angle of the ommatidial sample. The above equation suggests that, by default, the ommatidia fully respond when pointing above the horizon (due to the skylight), and partially respond otherwise (for example, when pointing at the ground). Note that all the ommatidia were assumed to neglect any skylight gradient or ground texture information. Additionally, the spectral sensitivity of the eye was assumed to be limited to the green colour channel, inspired by the limited colour vision of desert ants.

The position of the $n_{\text{grass}} = 5,000$ triangles (representing grass units in the environment) was transformed into their relative position to the compound eye. In order to determine whether a light beam reflected from a grass unit could reach the eye, or whether the grass blocked the skylight, their Cartesian coordinates were transformed to spherical coordinates with respect to the position of the eye. Consequently, the azimuth and elevation of their corners were computed as,

$$a_{\upsilon,\zeta} = \tan^{-1} \frac{y^{\text{tri}}_{\upsilon,\zeta} - y^{\text{eye}}}{x^{\text{tri}}_{\upsilon,\zeta} - x^{\text{eye}}}, \tag{5.15}$$

$$e_{\upsilon,\zeta} = \tan^{-1} \frac{\sqrt{(x^{\text{tri}}_{\upsilon,\zeta} - x^{\text{eye}})^2 + (y^{\text{tri}}_{\upsilon,\zeta} - y^{\text{eye}})^2}}{z^{\text{tri}}_{\upsilon,\zeta} - z^{\text{eye}}} - 90°, \tag{5.16}$$

where $\upsilon \in \{0, ..., n_{\text{grass}} - 1\}$ is the identity of the triangle and $\zeta \in \{A, B, C\}$ are its three corners. This transformation projects the triangles on the surface of the compound eye. Based on the *same-side technique*[2], an ommatidial sample sees the triangle when it looks at the same side of AB as C, of BC as A, and of CA as B. Thus, it receives the (green) light intensity associated to this grass-triangle, $g_\upsilon \in [0, 1]$, only when these conditions are satisfied. This is written formally as,

$$r^o_{\iota,\xi} = \begin{cases} & \text{if } (\overrightarrow{\text{BC}} \times \overrightarrow{\text{B}}) \cdot (\overrightarrow{\text{BC}} \times \overrightarrow{\text{BA}}) \geq 0, \text{ and} \\ g_\upsilon, & \text{if } (\overrightarrow{\text{AC}} \times \overrightarrow{\text{A}}) \cdot (\overrightarrow{\text{AC}} \times \overrightarrow{\text{AB}}) \geq 0, \text{ and} \\ & \text{if } (\overrightarrow{\text{AB}} \times \overrightarrow{\text{A}}) \cdot (\overrightarrow{\text{AB}} \times \overrightarrow{\text{AC}}) \geq 0, \\ r^o_{\iota,\xi}, & \text{otherwise,} \end{cases} \tag{5.17}$$

2 blackpawn.com/texts/pointinpoly (last visited: July 2022).

where $\cdot$ is the dot product and $\times$ is the cross product of the vectors. The vectors used here represent the apparent projections of the corners of the triangle on the eye, given by,

$$\begin{aligned}
\overrightarrow{\mathrm{A}} &= (a^{\text{glob}}_{\iota,\xi} - a_{\upsilon,A}, e^{\text{glob}}_{\iota,\xi} - e_{\upsilon,A}), & \overrightarrow{\mathrm{B}} &= (a^{\text{glob}}_{\iota,\xi} - a_{\upsilon,B}, e^{\text{glob}}_{\iota,\xi} - e_{\upsilon,B}),\\
\overrightarrow{\mathrm{AB}} &= (a_{\upsilon,B} - a_{\upsilon,A}, e_{\upsilon,B} - e_{\upsilon,A}), & \overrightarrow{\mathrm{BA}} &= (a_{\upsilon,A} - a_{\upsilon,B}, e_{\upsilon,A} - e_{\upsilon,B}),\\
\overrightarrow{\mathrm{AC}} &= (a_{\upsilon,C} - a_{\upsilon,A}, e_{\upsilon,C} - e_{\upsilon,A}), & \overrightarrow{\mathrm{BC}} &= (a_{\upsilon,C} - a_{\upsilon,B}, e_{\upsilon,C} - e_{\upsilon,B}),
\end{aligned}$$

where $a^{\text{glob}}_{\iota,\xi}$ and $e^{\text{glob}}_{\iota,\xi}$ are the global azimuth and elevation of the ommatidial sample. If the triangle was in the limits of the azimuthal axis (for example, over $180°$ or under $-180°$), two copies of the same triangle were created instead, one with only positive and one with only negative azimuth, and the same method was applied to both. Before rendering, the triangles were sorted based on their distance from the eye location. This allowed the closest triangle (with respect to the eye) to be rendered last and overwrite the information coming from triangles at the back.

Based on the rendered intensities of ommatidial samples, the summarised photoreceptor responses of the ommatidium (as shown in Fig. 5.2B and Fig. 5.3A) were then calculated using the equation below,

$$r^o_\iota = \sum_{\xi=0}^{6} w^{\text{Gauss}}_\xi \, r^o_{\iota,\xi}, \tag{5.18}$$

$$\mathbf{w}^{\text{Gauss}} = \begin{bmatrix} 0.25 & 0.125 & 0.125 & 0.125 & 0.125 & 0.125 & 0.125 \end{bmatrix}, \tag{5.19}$$

where $\mathbf{w}^{\text{Gauss}}$ is a vector that stores the normalised Gaussian density weights of the ommatidial samples—shared among all ommatidia as they all have the same properties. The photoreceptor responses of the ommatidia are finally stored in a vector, $\mathbf{r}^o$, which is used as input to the models described in the following sections.

5.5.3 *Principle component projection neurons*

To build the weight matrix that transforms the photoreceptor responses of the ommatidia to the *principle component projection neuron* (PCPN) responses, a covariance matrix of the input responses is needed. However, at least as many samples as the number of ommatidia are needed to avoid singularity in the covariance matrix, given the assumption that these are independent samples from the same (Gaussian) distribution. Therefore, $n_{\text{sample}} = 1,000$ samples were collected from random directions and positions in a $2\,\text{m}$ radius around the end of the routes (where the nest was lo-

cated), and stored in a two-dimensional matrix, $\mathbf{X} \in [0,1]^{n_{\text{sample}} \times n_{\text{omm}}}$. The elements of the covariance matrix, $\mathbf{C} \in \mathbb{R}^{n_{\text{omm}} \times n_{\text{omm}}}$, were then calculated as follows,

$$c_{\iota,\iota'} = \frac{1}{n_{\text{sample}} - 1} \sum_{s=0}^{n_{\text{sample}}-1} (x_{s,\iota} - \mu_\iota^o)\,(x_{s,\iota'} - \mu_{\iota'}^o), \tag{5.20}$$

$$\mu_\iota^o = \frac{1}{n_{\text{sample}}} \sum_{s=0}^{n_{\text{sample}}} x_{s,\iota}, \tag{5.21}$$

where μ^o is the mean vector across the samples. The eigenvectors, $E \in \mathbb{R}^{n_{\text{omm}} \times n_{\text{omm}}}$, and eigenvalues, $\lambda \in \mathbb{R}^{n_{\text{omm}}}$, were calculated using standard PCA, and they were subsequently sorted with respect to the eigenvalues. This ensured that the eigenvectors were ranked with respect to the variance covered and that the first ones would be the most useful descriptors of the information in the photoreceptor responses.

The elements of the weight-matrix that transforms the photoreceptor to PCPN responses, $\mathbf{W}^{o2p} \in \mathbb{R}^{n_{\text{omm}} \times n_{\text{PN}}}$, were calculated by,

$$w_{\iota,\zeta}^{o2p} = \frac{e_{\iota,\zeta}}{\lambda_\zeta}, \tag{5.22}$$

for $\iota \in \{0, ..., n_{\text{omm}} - 1\}$ and $\zeta \in \{0, ..., n_{\text{PN}} - 1\}$.

Finally, the responses of the PCPNs, $\mathbf{r}^p \in [0,1]^{n_{\text{PN}}}$, were calculated using the responses of the ommatidia,

$$r_\zeta^p = \nu\Big(\sum_{\iota=0}^{n_{\text{omm}}-1} r_\iota^o\, w_{\iota,\zeta}^{o2p}\Big), \qquad \nu(x_\zeta) = \frac{x_\zeta - \min_k x_k}{\max_k x_k - \min_k x_k}, \tag{5.23}$$

where $\nu(x)$ normalises the PN responses in $[0,1]$.

5.5.4 *Kenyon cells combinatorial coding*

The vPN $\rightarrow$ KC synaptic weights were built using a heuristic combinatorial approach. Note that this method does not rely on the actual vPN responses. A set of synaptic patterns were built initially, composed by $n_s = 1.5\, n_{\text{KC}} = 6,000$ vPN patterns. Then the $n_{\text{KC}} = 4,000$ patterns with the lowest correlation among them were selected as the final synaptic weights.

In order to generate unique synaptic patterns (of size n) for a given number of synapses (k), the `combinations` function of the `itertools` Python package was used, which is summarised in Algorithm 1. This function iteratively returns unique combinations of k values from a set of n available values (through the `yield` command).

Algorithm 1 Sequentially generates "n choose k" unique sets of k values in $[0, n-1]$.

```
1:  function Combinations(n, k)
2:      if k > n then return                          ▷ Ensure k is valid
3:      p ← [0, ..., k − 1]                           ▷ Initialise with the simplest pattern
4:      yield p                                       ▷ Return the first (simplest) pattern
5:      ▷ Generate the rest of the patterns by changing the indices ◁
6:      loop
7:          ▷ Terminate when no other patterns are available ◁
8:          terminate ← true
9:          for i = k − 1, ..., 0 do
10:             if p_i ≠ i + n − k then
11:                 terminate ← false
12:                 break
13:         if terminate then return
14:         ▷ Update the pattern ◁
15:         p_i ← p_i + 1
16:         for j = i + 1, ..., k − 1 do
17:             p_j ← p_{j−1} + 1
18:         yield p
```

The property of iteratively returning combinations makes this function a generator that creates patterns by request and avoids memory overflow.

Combinations was used by GenerateUniquePatterns$(n, n_p, k^{\text{min}}, k^{\text{max}})$ in order to create n unique synaptic patterns of size n_p, with k^{min} to k^{max} synapses each (see Algorithm 2). This was done sequentially, by picking a random number of synapses $k \in \{k^{\text{min}}, ..., k^{\text{max}}\}$ and generating a novel synaptic pattern using Combinations(n_p, k). Each number of synapses (k) was associated with a different Combinations generator, which allowed for individually generating patterns with different numbers of synapses per pattern. For each k (number of synapses), the function ensured that all the generated patterns were unique by checking the number of generated patterns against the total number of possible synaptic patterns. This could be estimated by the *binomial coefficient* or the "n choose k" number. Patterns generated with different ks could not be the same by definition, so there was no control for that. Finally, the function ensured that the synaptic weights in each generated pattern sum to 1.

By using the above functions, Algorithm 3 describes how heuristically optimal vPN $\rightarrow$ KC sparse synaptic weights were generated in a combinatorial fashion. First, it uses GenerateUniquePatterns to generate $n_s = 6,000$ unique synaptic patterns of size $n_{\text{PN}} = 50$, each with $k \in \{1, ..., 6\}$ synapses. Then it calculates the cross-correlation between the pairs of patterns. For each pattern, all the patterns are sorted based on

Algorithm 2 Generates n_p unique patterns of size n, where $[k^{\text{min}}, k^{\text{max}}]$ values are non-zero (positive).

function GENERATEUNIQUEPATTERNS(n, n_p, k^{min}, k^{max})

▷ *Initialisation of the parameters* ◁
$\mathbf{S} \leftarrow$ matrix of zeros $\in [0, 1]^{n \times n_p}$
$\kappa \leftarrow$ empty list of combination generators (of size k^{max})

▷ *Generate n_p unique patterns* ◁
for $\xi = 0, ..., n_p - 1$ **do**
  $k_\xi \leftarrow$ random $\in \{k^{\text{min}}, ..., k^{\text{max}}\}$ ▷ *the number of synapses in the pattern*

  ▷ *Calculate the number of times that k_ξ was picked* ◁
  $n_\xi^k \leftarrow \sum_{\zeta=0}^{\xi} \delta_K(k_\xi, k_\zeta)$ ▷ *Kronecker delta:* $\delta_K(i,j) = \begin{cases} 0 & \text{if } i \neq j, \\ 1 & \text{if } i = j \end{cases}$
  ▷ *If there are no available unique synaptic patterns, increase n_ξ^{syn}* ◁
  while $n_\xi^k > \binom{n}{k_\xi}$ **do** ▷ *Binomial coefficient:* $\binom{n}{k_\xi} = \frac{n!}{k_\xi!(n-k_\xi)!}$
    $k_\xi \leftarrow k_\xi + 1$
    $n_\xi^k \leftarrow \sum_{\zeta=0}^{\xi} \delta_K(k_\xi, k_\zeta)$

  ▷ *If k_ξ was picked for the first time, create a new combination generator* ◁
  if $n_\xi^k = 1$ **then**
    $\kappa(k_\xi - 1) \leftarrow$ COMBINATIONS(n, k_ξ)

  ▷ *Generate the next available unique pattern with k_ξ synapses* ◁
  $\mathbf{i}^{\text{syn}} \leftarrow$ **next** $\kappa(k_\xi - 1)$
  for all $i \in \mathbf{i}^{\text{syn}}$ **do**
    $s_{i\xi} \leftarrow \frac{1}{n_\xi^{\text{syn}}}$ ▷ *Ensure that (for each KC) the sum of the weights is one*
return S

their correlation to that pattern (lowest goes first), and the number of times (frequency) that each pattern was in the $n_{\text{KC}} = 4,000$ lowest correlations was calculated. The patterns were then sorted based on that frequency (highest first), as it would mean that they have a relatively lower correlation with all the patterns. The n_{KC} patterns with the highest frequency were used as the vPN $\rightarrow$ KC synaptic weights, $\mathbf{W}^{p2k} \in [0, 1]^{n_{\text{PN}} \times n_{\text{KC}}}$.

Finally, the KC responses, $\mathbf{r}^k \in \{0, 1\}^{n_{\text{KC}}}$, were calculated by the PN responses in a similar way to Eq. (3.1) by using,

$$r_\xi^k = \text{WTA}_{0.0025}\Big(\sum_{\zeta=0}^{n_{\text{PN}}-1} r_\zeta^p \cdot w_{\zeta,\xi}^{p2k} \Big), \tag{5.24}$$

where $\text{WTA}_{0.0025}(x)$ is the *winner-takes-all* (WTA) activation function that keeps only 10 KCs ($0.25\,\%$ of them) active, based on the strength of their activity.

Algorithm 3 The combinatorial sparse coding weights generation.

Require: $n_{\text{PN}} = 50$ ▷ *The total number of PNs*

Require: $n_{\text{KC}} = 4,000$ ▷ *The total number of KCs*

Require: $n_{\text{min}}^{\text{syn}} = 1$ ▷ *The minimum number of PNs connected to a KC*

Require: $n_{\text{max}}^{\text{syn}} = 6$ ▷ *The maximum number of PNs connected to a KC*

Ensure: $n_{\text{PN}} \geq n_{\text{max}}^{\text{syn}} \geq n_{\text{min}}^{\text{syn}} > 0$

$n_s = 1.5\, n_{\text{KC}}$

$\mathbf{S} \leftarrow \textsc{GenerateUniquePatterns}(n_{\text{PN}}, n_s, n_{\text{min}}^{\text{syn}}, n_{\text{max}}^{\text{syn}})$

▷ *For each pattern, sort all the patterns by ascending order of correlations* ◁

$\mathbf{I}^c \leftarrow$ matrix of zeros $\in \mathbb{N}^{n_s \times n_s}$

for $\xi = 0, ..., n_s - 1$ **do**

  ▷ *Calculate the correlation between the current pattern and all the others* ◁

  $\mathbf{c} \leftarrow$ array of zeros $\in \mathbb{R}^{n_s}$

  for $\zeta = 0, ..., n_s - 1$ **do**

    $c_\zeta \leftarrow \frac{\sum_{\iota=0}^{n_{\text{PN}}-1} s_{\iota,\xi} \cdot s_{\iota,\zeta}}{\sqrt{\sum_{\iota=0}^{n_{\text{PN}}-1} s_{\iota,\xi}^2} \sqrt{\sum_{\iota=0}^{n_{\text{PN}}-1} s_{\iota,\zeta}^2}}$

  ▷ *Sort the patterns by their correlation (lowest to highest)* ◁

  $\mathbf{i}_\xi^c \leftarrow \textsc{ArgQuicksort}(\mathbf{c}, \text{'ascending'})$

▷ *Count the number of times that each pattern was in the n_{KC} lowest correlations* ◁

$\mathbf{n}^i \leftarrow$ array of zeros $\in \mathbb{N}^{n_s}$

for $\iota = 0, ..., n_s - 1$ **do**

  $n_\iota^i \leftarrow \sum_{\xi=0}^{n_s-1} \sum_{\zeta=0}^{n_{\text{KC}}-1} \delta_K(i_{\xi,\zeta}^c, \iota)$ ▷ *δ_K is Kronecker delta*

▷ *Sort the patterns by their frequency in the n_{KC} lowest correlations (highest to lowest)* ◁

$\mathbf{i} \leftarrow \textsc{ArgQuicksort}(\mathbf{n}^i, \text{'descending'})$

▷ *The synaptic patterns with the most frequent lower correlations form the weights matrix* ◁

$\mathbf{W}^{p2k} \leftarrow$ matrix of zeros $\in [0, 1]^{n_{\text{PN}} \times n_{\text{KC}}}$

for $\iota = 0, ..., n_{\text{PN}} - 1$ **do**

  for $\xi = 0, ..., n_{\text{KC}} - 1$ **do**

    $w_{\iota,\xi}^{p2k} \leftarrow s_{\iota,i_\xi}$

5.5.5 *Running the experiments*

During training on a route, ξ, with length, n_ξ^{route}, the compound eye model was set to the recorded positions and pose of a desert ant (retrieved from the behavioural data). The eye rendered the ant's view in that pose as described in Section 5.5.2 and generated the responses of the ommatidia. These were used to compute the responses of the PCPNs—Eq. (5.23)—which were then processed by different models in order

to update their memory and estimate the familiarity for this view. The memory could either take the form of a database (PM) or KC $\rightarrow$ MBON synaptic weights.

Perfect memory

Perfect memory (PM) was trained by storing the vPN responses during the sample route in a database matrix, $\text{DB} \in [0,1]^{n_{\text{route}} \times n_{\text{PN}}}$. Then the familiarity of the current PN responses, $\mathbf{r}^p$, was calculated using Eq. (5.1) and Eq. (5.2), which are summarised here for convenience,

$$f^{\text{PM}} = 1 - \min_{i \in \{0,\dots,n_{\text{route}}-1\}} \sqrt{\sum_{j=0}^{n_{\text{PN}}-1} (r_j^p - \text{DB}_{i,j})^2}. \tag{5.25}$$

Willshaw network

The *Willshaw network* (WN) used the KC responses, $\mathbf{r}^k$, for storing and retrieving memories in the KC $\rightarrow$ MBON synaptic weights, $\mathbf{w}^{k2m} \in [0,1]^{n_{\text{KC}}}$. This was done by using the anti-Hebbian plasticity rule, which was simplified for this task as,

$$w_i^{k2m} = \begin{cases} 0, & \text{if} \quad r_i^k \, w_i^{k2m} > 0, \\ w_i^{k2m}, & \text{otherwise.} \end{cases} \tag{5.26}$$

Note that this model has only one MBON and therefore, during testing, the familiarity was computed as

$$f^{\text{WN}} = 1 - \frac{1}{\sum_{j=0}^{n_{\text{KC}}-1} r_j^k} \sum_{i=0}^{n_{\text{KC}}-1} r_i^k \, w_i^{k2m}. \tag{5.27}$$

In this equation, the output of the WN was normalised to ensure that the familiarity is always bounded, $f^{\text{WN}} \in [0,1]$.

Incentive circuit

The *incentive circuit* (IC) also uses the KC responses, $\mathbf{r}^k$, for storing and retrieving memories in the KC $\rightarrow$ MBON synaptic weights, $\mathbf{W}^{k2m} \in [0,50]^{n_{\text{KC}} \times 6}$, but with 6 MBONs as the model indicates. The update of the synaptic weights was done using the Eq. (3.16) including the dynamics of the circuit as described in Section 3.5.2. However, all the processing up until the calculation of the KC responses was calculated as

in the WN. Using the responses of the six MBONs of the model the familiarity was calculated as

$$f^{\mathrm{IC}} = \frac{s_f - s_n + r_f - r_n + m_f - m_n}{6} + 0.5, \tag{5.28}$$

where the *attract* valence was replaced by the *familiar* (f), and the *avoid* by the *novel* (n) respectively.

Testing the models

During testing, the models stopped updating their memory and their extracted familiarity was recorded in specific poses, defined with respect to the test condition. In the *parallel displacements* experiment, the test routes were copies of the training route displaced for 10 intervals of $2\,\mathrm{cm}$ to the right and 10 to the left, which resulted in 21 parallel routes. The displacements were parallel to the central axis of the training route, defined by the line connecting its start and end positions (see Fig. 5.6A). This was calculated as,

$$\mathrm{d}x_\delta = \frac{-\delta\,(y_{\mathrm{end}} - y_{\mathrm{start}})}{\sqrt{(x_{\mathrm{end}} - x_{\mathrm{start}})^2 + (y_{\mathrm{end}} - y_{\mathrm{start}})^2}}, \tag{5.29}$$

$$\mathrm{d}y_\delta = \frac{\delta\,(x_{\mathrm{end}} - x_{\mathrm{start}})}{\sqrt{(x_{\mathrm{end}} - x_{\mathrm{start}})^2 + (y_{\mathrm{end}} - y_{\mathrm{start}})^2}}, \tag{5.30}$$

where $\delta \in \{-0.2, ..., 0, ..., 0.2\}$ defines the 21 displacement magnitudes and directions in meters. The new positions were then calculated by adding the displacement to the recorded positions of the original route,

$$x_\delta = x + \mathrm{d}x_\delta\,, \qquad\qquad y_\delta = y + \mathrm{d}y_\delta\,, \tag{5.31}$$

while the directions of the tested routes were identical to the training routes, $\phi_\delta = \phi$.

In the *glancing looks* experiment, the positions of the test routes were unaltered while the directions were changed (see Fig. 5.6E). The training routes were copied 36 times, altering the heading direction of the ants in $10°$ intervals, which is $\phi_\gamma = \phi + \gamma$, where $\gamma \in \{0°, ..., 350°\}$.

In all the above experiments, for each tested position and direction, the familiarities predicted by the different models were collected and summarised in matrices, $F^{\mathrm{PD},M} \in [0,1]^{16\times 21}$ and $F^{\mathrm{GL},M} \in [0,1]^{16\times 36}$ for the parallel displacements and glancing looks experiments respectively, were $M \in \{\mathrm{PM}, \mathrm{WN}, \mathrm{IC}\}$. These matrices stored

the mean familiarity along each of the 16 routes. These were calculated for each displacement and altered direction as follows,

$$f^{\mathrm{PD},M}_{\xi,\delta} = \frac{1}{n^{\mathrm{route}}_{\xi}} \sum_{j=0}^{n^{\mathrm{route}}_{\xi}} \frac{f^{\mathrm{PD},M}_{\xi,\delta,j}}{\max_{i\in\{-0.2,\ldots,0.2\}} f^{\mathrm{PD},M}_{\xi,i,j}}, \tag{5.32}$$

$$f^{\mathrm{GL},M}_{\xi,\gamma} = \frac{1}{n^{\mathrm{route}}_{\xi}} \sum_{j=0}^{n^{\mathrm{route}}_{\xi}} \frac{f^{\mathrm{GL},M}_{\xi,\gamma,j}}{\max_{i\in\{0^\circ,\ldots,350^\circ\}} f^{\mathrm{GL},M}_{\xi,i,j}}. \tag{5.33}$$

Note that the familiarities were normalised along the displacements and glancing directions. This ensured that the familiarities were comparable across different routes and models. Finally, these matrices were plotted as lines for the different models in Fig. 5.6B-D and Fig. 5.6F-H respectively, along with their mean along all the routes (thick black lines).

5.5.6 *Steepness analysis*

Only the familiarity matrices from the parallel displacements experiment were used for this analysis, so the PD superscript is omitted for simplicity. For the familiarities of each model, the mean displacement distance was computed (weighted by the familiarity) across the different displacements and along the routes,

$$\mu^{M}_{\xi,j} = \frac{1}{\sum_{i\in\{-0.2,\ldots,0.2\}} f^{M}_{\xi,i,j}} \sum_{\delta\in\{-0.2,\ldots,0.2\}} \delta\, f^{M}_{\xi,\delta,j}, \tag{5.34}$$

where $j \in \{0,\ldots,n^{\mathrm{route}}_{\xi}\}$ are the different steps of the route ξ. The mean was used for computing the respective SD,

$$\sigma^{M}_{\xi,j} = \sqrt{\frac{1}{\sum_{i\in\{-0.2,\ldots,0.2\}} f^{M}_{\xi,i,j}} \sum_{\delta\in\{-0.2,\ldots,0.2\}} (\delta-\mu^{M}_{\xi,j})^2\, f^{M}_{\xi,\delta,j}}, \tag{5.35}$$

and this was subsequently converted to the *steepness* shown in Fig. 5.7B-G as,

$$\tau^{M}_{\xi,j} = \frac{1}{\sigma^{M}_{\xi,j}}. \tag{5.36}$$

Linear regression

Linear regression (through the scikit-learn Python package, Pedregosa et al., 2011) was used to fit the lines in Fig. 5.7D-I and compute the slope of the line (as an indicator

of the increasing or decreasing steepness). This uses the *gradient decent* algorithm to minimise the *residual sum of squares* (RSS) objective function formalised below,

$$\mathcal{L}^{M}_{\text{RSS}}(m, b) = \sum_{\xi=0}^{15} \sum_{j=0}^{n^{\text{route}}_{\xi}-1} [\tau^{M}_{\xi,j} - (0.01\, j\, m + b)]^2, \tag{5.37}$$

where m is the slope and b is the intercept (free parameters that are set by the regression process). In the above equation, $0.01\, j$ represents the step size (which is $1\,\text{cm}$) times the step count, as plotted on the horizontal axis of Fig. 5.7B-G (in meters).

This Python package does not use the common pseudo-inverse matrix. This is in order to support a broader range of problems with an arbitrary number of parameters and size of data. The pseudo-inverse requires the calculation of the covariance matrix, which becomes unstable when the number of parameters is larger than the size of the data.

Page's trend test

In order to test whether the steepness is increasing, the Page's trend test was used. This test shows whether the observations have a particular order (Page, 1963). In this work, it was used for testing whether the steepness (observations) is increasing with time (order), by comparing the null,

$$\tau^{M}_{\xi,0} = \tau^{M}_{\xi,1} = \ldots = \tau^{M}_{\xi,n^{\text{route}}_{\xi}-1}, \tag{5.38}$$

against the alternative hypothesis,

$$\tau^{M}_{\xi,0} < \tau^{M}_{\xi,1} < \ldots < \tau^{M}_{\xi,n^{\text{route}}_{\xi}-1}. \tag{5.39}$$

The test shows whether the steepness is increasing with time for the individual routes, which is more useful than comparing the overall slope from all the routes against the null hypothesis (flat slope).

Page's test requires the same number of orders (time-steps) per observation (steepness) for all the samples (routes). However, each recorded route had a different number of time steps. Thus, the steepness in each route was binned in $n_{\text{bin}} = 16$ discrete bins, calculating the median steepness over batches of $50\,\text{cm}$ length. The number of bins was selected in order to support all route lengths. As some routes were shorter than $753\,\text{cm}$, $n_{\text{bin}} > 16$ would result in at least one empty bin.

By using these converted data, Page's L statistic was calculated for all the $n_{\text{samples}} = 16$ routes as,

$$L = n_{\text{bin}} \sum_{j=0}^{n_{\text{bin}}-1} (j+1) \sum_{i=0}^{n_{\text{samples}}-1} \frac{\tau_{i,j}}{\max_k \tau_{i,k}}, \tag{5.40}$$

which can be transformed to a quantity that is comparable to values of the chi-square distribution with one degree of freedom as,

$$\chi^2_L = \frac{[12\,L - 3\,n_{\text{samples}}\,n_{\text{bin}}\,(n_{\text{bin}}+1)^2]^2}{n_{\text{samples}}\,n^2_{\text{bin}}\,(n^2_{\text{bin}}-1)\,(n_{\text{bin}}+1)}. \tag{5.41}$$

It can also be transformed into a p-value that measures the overall correlation of the data with the alternative hypothesis as

$$p = 12\,\frac{L}{n_{\text{samples}}}\,(n^3_{\text{bin}} - n_{\text{bin}}) - 3\frac{n_{\text{bin}}+1}{n_{\text{bin}}-1}. \tag{5.42}$$

Note that the outcome of this statistical test is supposed to be reliable for any number of routes if $n_{\text{bin}} > 9$.

6

DELAYED REINFORCEMENTS

> *"A person who has been punished is not less inclined to behave in a given way; at best, he learns how to avoid punishment."*
>
> B. F. Skinner

Throughout the previous chapters, the reinforcement was considered to be an electric shock, sugar, or homing state (returning home from a food source that was visited for the first time). In that sense, it could be defined as a sensory input or internal state that affects the future behaviour of the animal. However, the reinforcement might occur quite sparsely in time and it might trigger changes in the behaviour that involve previous sensory inputs or internal states that are not temporally paired with the reinforcement. For example, in the route-following task (which was discussed in the previous chapter) the reinforcement was assumed to be the state of returning home, but only when the food source was visited for the first time. An alternative reinforcement could be the state that the animal arrives home, which lasts only for one moment. Another possibility is that the reinforcement is the state in which the animal finds or possesses the food. These different reinforcement functions could lead to different learning pipelines, that might result in distinct behaviours. In this chapter, the possibility of having a sparse and delayed reinforcement was examined, challenging the *incentive circuit* (IC) to solve tasks where the behaviour should change with respect to a stimulus that was never explicitly paired in time with the reinforcement.

Finding solutions for delayed reinforcement tasks is popular in more standard *reinforcement learning* (RL) algorithms, which have to find the optimal sequence of states and actions that leads to the overall maximum reinforcement. In this chapter, the capabilities of the IC are also tested in a subset of these tasks, and its performance is compared to standard RL algorithms that are known to be able to provide solutions. The results show that the IC was incapable of finding an optimal way to receive the delayed reinforcement, and a number of possible reasons are discussed. Finding a solution for these problems requires a policy for taking action over time, which has started to become evident in fruit flies (Dylla et al., 2017; Grover et al., 2022). However (as discussed in the previous chapter), the *mushroom body output neuron* (MBON) re-

sponses might not be parallel to the actual actions of the animal (for example, turn left or right), but they might provide the motivation to repeat an action or not depending on the context. This would enable a different system—like the *central complex* (CX)—to adapt the default behaviour based on the current motivation provided by the *mushroom bodies* (MBs); this suggests that the MBs provide the value of the current state in the context of the behaviour, and they might take the role of the adaptive critic in an RL framework.

6.1 BACKGROUND

Reinforcement learning (RL) is a term used in engineering and *artificial intelligence* (AI) to specify learning algorithms that rely their behaviour on the principle of reinforcement (Kimble, 1961). In experimental psychology, the term 'reinforcement' is used to describe the increased probability of an animal learning how to take a specific action during an event when this action is paired with the satisfaction of a drive, which follows the classic *law of effects* (Thorndike, 1927). In engineering, this is usually mathematically formulated as an optimisation problem, which comprehensibly tries to find the most suitable action (or a policy that generates actions) and form a behaviour. It is important that an RL system tries to produce the optimal behaviour by using an objective function that overall maximises the positive reinforcements. Based on the complexity of the objective function characterising these algorithms and the available information that they have access to, Barto (1997) categorised them into three basic types: *non-associative*, *associative*, and *sequential* RL.

In non-associative RL, the agent is trying to find the action that maximises its satisfaction, which sounds equivalent to 'pure operant' learning, where only actions determine the reinforcement (Brembs, Lorenzetti, et al., 2002). In this type of algorithm, the only external input to the learning system is the reinforcement signal (although there is still internal information about the action taken; see Fig. 6.1A), resulting in a lack of discrimination between different states of the process guided by actions. This differs from the type of RL that Thorndike (1927) described, in which optimal action selection depends on the context. In practice, the description of Thorndike (1927) matches with the associative RL, which aims to create an accurate *associative map* between the sets of states and actions by using reinforcements from the environment (see Fig. 6.1B). In sequential RL, the associative map represents a policy of acting over time and it is used to solve asynchronous problems (where the reinforcement may not be available immediately after the action is taken).

In AI, usually agents take the place of animals and their behaviours are described by systems. An agent (that learns by reinforcements) explores its environment and

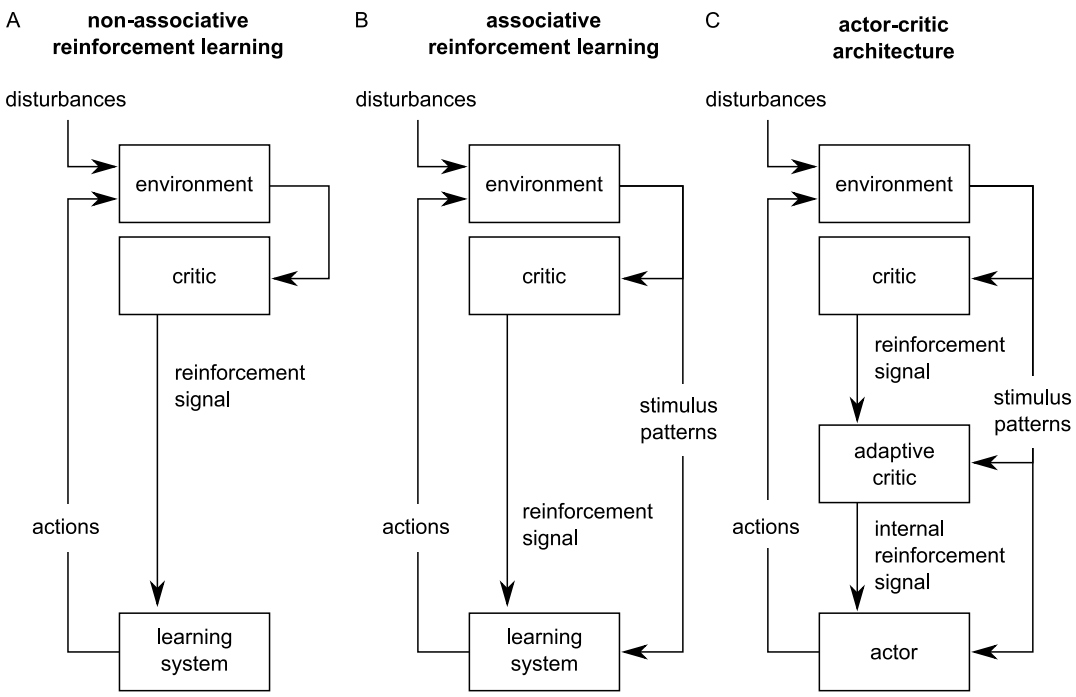

Figure 6.1: *Reinforcement learning* (RL) architectures. (A) In *non-associative RL*, the environment is influenced by the actions of the system and some unknown disturbances. The critic uses observations from the environment to evaluate the actions and provide a reinforcement signal. This is used by the learning system to update its parameters (without observations of the environment). (B) In *associative RL*, the learning system can also observe the environment (through stimulus patterns) in addition to the reinforcement signal from the critic. In this case, the actions of the system also depend on the stimulus patterns creating an associative map between the states and the actions. (C) In an *actor-critic* architecture, an adaptive critic learns to predict the reinforcement signal; this prediction forms an internal reinforcement signal, which is given as input to the actor; this allows for learning even when the reinforcement is omitted and forms a policy that controls the behaviour. Adapted by permission from Elsevier: Copyright © Barto (1997).

learns its behaviour by trial and error. Fig. 6.1 provides a schematic representation of RL architectures for non-associative and associative RL. In this figure, each diagram represents a different agent, but all agents follow similar principles. The behaviour is influenced by the agent's action, $\{a_i \,|\, i = 0, ..., A-1\}$ (and some other unknown facts—noted as 'disturbances'); it affects the state of the agent, $\{s_j \,|\, j = 0, ..., S-1\}$, and its environment, from which the 'critic' gets input and decides whether the agent is satisfied or not. Its decision produces a reinforcement signal, $r_{j,i}$, which is transmitted to the learning system (or the adaptive critic). The agent modifies its action selection mechanism using some criteria and selects an action that influences the behaviour.

The above loop is the principle that most of the RL algorithms are based on. As Fig. 6.1 indicates, Barto (1997) categorised the algorithms with respect to their con-

straints on what high-level information is available or not in their learning system (for example, state, reward, or time), increasing the complexity as more information is available. Other categories of RL algorithms are based on the availability of a world model (model-free and model-based), sampling (episodic RL), and others on different descriptions of the sequential patterns—such as *Markov decision processes* (MDPs) and *long short-term memories* (LSTMs).

6.1.1 *Associative reinforcement learning*

In associative RL, information about the state is available for both the critic and the learning system. So, the selected actions and plasticity rules add the state of the agent in the loop (see Fig. 6.1B), which results in different actions being best in different contexts (Barto, 1997). The objective here is to maximise the immediate reward (or minimise the immediate punishment) by forming an optimal associative map between potential stimuli and available actions. For example, let $x(t)$ be the stimulus sensed on time-step t (equivalent to the state), which is used by the learning system to select the action $a(t) = a_i$ with a probability $p_i[x(t)] = Pr\{a(t) = a_i\}$. Consequently, the critic shows its satisfaction through the reinforcement, $r(t)$, which is drawn from the actual probability of signalling 'success', $p_i^s[x(t)]$; the system tries to optimise the expected probability of success when taking this action, $p_i[x(t)]$, by taking into account the state of the agent and the reinforcement provided by the critic.

In this category of RL, it is common to represent the learning systems using neural networks, as they provide a handy representation of associative maps and can also introduce non-linearity and multiple levels of processing, leading to more complicated plasticity rules. Let $\mathbf{x}(t)$, $\mathbf{w}(t)$, $a(t)$, and $r(t)$ respectively denote the stimulus vector, weight vector, action, and the reinforcement signal for a time-step t. Let also $V(t)$ denote the activation of the neuron-like unit given the stimulus components at time-step t,

$$V(t) = \sum_{i=0}^{S-1} w_i(t)\, x_i(t). \tag{6.1}$$

The learning system selects an action using the activation level values, $V(t)$, and updates the network by changing the weights with a plasticity rule. The activation level is used to estimate the probability of an action to cause a satisfactory signal from the critic, $p_i[\mathbf{x}(t)] = f[V(t)]$, where $f(\cdot)$ is a sigmoid function (for example, the logistic function), which ensures that $0 \leq p_i[\mathbf{x}(t)] \leq 1$.

6.1.2 *Sequential reinforcement learning*

In associative RL, the agent tries to optimise the values of an associative transition map from states to actions. In sequential RL, the agent tries to optimise this associative map by building a strategy (or policy) for acting over time, which creates associations between states and actions depending on reinforcements that come later in time (Barto, 1997). This means that the agent may penalise its performance in the short-term aiming to maximise it in the long term. Such properties are usually desirable in optimal control problems, which are very common in robotics. The panels in Fig. 6.1 can also describe the various sequential RL algorithms, as the difference from the associative RL is mainly on the adaptive critic block.

A sequential RL system tries to maximise the reinforcement value, $V(t)$, that is expected to receive over time. This is usually the *discounted sum* of all the reinforcements that it expects to receive in the future, which is mathematically formulated as,

$$V(t) = E[\sum_{\tau=0}^{\infty} \gamma^{\tau} r(t-\tau)] = E[r(t)] + \gamma V(t+1), \tag{6.2}$$

where $E[\cdot]$ represents the expectation over the possible future experiences, and $0 \leq \gamma \leq 1$ is the discount factor that sets the horizon of the agent; it penalises reinforcements that are expected to be received far in the future (γ^{τ} decreases as τ increases). If $\gamma = 0$, the agent is myopic (it is concerned only for the current time-step, like in the associative RL) and leads to $V(t) = r(t)$; if $\gamma = 1$, it would lead to $V(t) = \infty$. Note that the above is a variation of Bellman's optimality equation (Bellman, 1966).

A special case of sequential RL is when the only reinforcement that the agent receives is at the final (goal) state. In this case, there is no feedback on which exact action was important in order to reach the solution and get credit for it. The *adaptive critic* systems (which will be described later in this section) can solve these problems by learning an internal evaluation function that is more informative than the critic's evaluation. The units that will be described in this section focus on predicting the internal reinforcement value (see Fig. 6.1C) and act as the critic of the system (instead of being used for action selection, like in associative RL). Non-associative or associative RL units can be used for action selection (as actor units) in sequential RL systems. For this reason, these systems are sometimes described as *actor-critic architectures* (Barto, 1997).

Adaptive critic units

These units focus on estimating the internal reinforcement value, $V(t)$, which is the accumulated reward in a time horizon specified by γ. Given Eq. (6.1) and making the assumption that this value is accurately estimated by these units, the expected discounted reward would be,

$$\hat{r}(t) = V(t) - \gamma V(t+1). \tag{6.3}$$

As the reinforcement values can be calculated by using stimuli from the environment, $x(t)$, the system is able to evaluate its estimated reinforcement in the following time-step, $t+1$, using the *temporal difference* (TD) error to update its weights,

$$\begin{aligned} \Delta \mathbf{w}(t) &= \eta[r(t) - \hat{r}(t)]\mathbf{x}(t) \\ &= \eta[r(t) + \gamma V(t+1) - V(t)]\mathbf{x}(t). \end{aligned} \tag{6.4}$$

Note that this update of the weights in time-step t is delayed and actually happens in the next time-step, $t+1$. This means that the error depends on successive estimations of the critic, which give the name of TD to this error. This unit learns to accurately predict the discounted sum of reinforcements. Non-linear predictions are also possible (for example, by back-propagating the TD-error in multi-layer networks).

Action-dependent adaptive critic units

Probably the most successful sequential RL algorithms are the *action-dependent adaptive critics* (ADACs). Instead of updating a value for all the expected future rewards, these units store a value for each state-action pair, creating a look-up table of $Q(\mathbf{x}, a)$ values,

$$Q_t[\mathbf{x}(t), a(t)] = E[r(t)] + \gamma V(t+1). \tag{6.5}$$

This table is updated in a similar way as the values in the simple adaptive critic units, with the difference that only the value associated with the specific state and action changes every time as,

$$\Delta \mathbf{w}_t[\mathbf{x}(t), a(t)] = \eta(t)\{r(t) + \gamma V(t+1) - Q_t[\mathbf{x}(t), a(t)]\}\mathbf{x}(t), \tag{6.6}$$

where $\eta(t) > 0$ is a time-dependent learning rate.

The most popular variants of this algorithm are *Q-learning* (Bozinovski, 1982; Lukes, Thompson, and Werbos, 1990; Watkins and Dayan, 1992), *Sarsa* (Rummery and Ni-

ranjan, 1994; Singh and Sutton, 1996), and their variations. These methods are the representatives of *off-policy* and *on-policy* control respectively, which is related to the way that they compute the value, $V(t+1)$, that is used in the update equation. Specifically, Q-learning uses the maximum value over all the available actions in time-step $t+1$ (off-policy),

$$V(t+1) = \max_{a \in A(t+1)} Q_t[\mathbf{x}(t+1), a], \tag{6.7}$$

while Sarsa uses the value associated with the expected state and action in that time step (on-policy),

$$V(t+1) = Q_t[\mathbf{x}(t+1), a(t+1)]. \tag{6.8}$$

Due to their look-up table requirement, these algorithms can be used only in systems with finite states and actions. However, their variations can approximate the values by interpolating for an infinite number of states and actions. A very popular example of these algorithms for function approximation is the deep neural networks (for example; Mnih et al., 2015).

Episodic reinforcement learning

In the context of an agent learning to perform a specific task, when the agent reaches a goal state or after a fixed amount of time, an episode ends and the process starts over. This process can be done in either a continuous or discontinuous manner and it is widely used in the field in order to efficiently train agents. Note that an episode in non-sequential RL could be a single time-step.

In episodic RL, the agent stores samples of its experience and uses them to improve its estimations of the expected rewards and values of its states. This can be done by either storing multiple episodes or updating the average value of its states at the end of each episode. In particular, this sets the main role of $E(\cdot)$ in Eq. (6.2), where it averages over a number of episodes. RL algorithms that have this property usually belong to the group of *Monte Carlo* (MC) methods (Sutton and Barto, 2018) as it is assumed to use MC sampling over the space of episodes.

Eligibility trace

The idea behind the eligibility trace is quite simple. In every time step, each active state initiates a different *short-term memory* (STM) trace, which decays with time using the decay factor, $0 \leq \lambda \leq 1$ (Klopf, 1972). There are different variants of how to

update such traces, like replacing and conventional ones. The replacing eligibility trace replaces the trace with a new one, for example using the equation below,

$$e_{t+1}(\mathbf{x}) = \begin{cases} \gamma\lambda e_t(\mathbf{x}), & \text{if } \mathbf{x} \neq \mathbf{x}(t), \\ 1, & \text{if } \mathbf{x} = \mathbf{x}(t). \end{cases} \tag{6.9}$$

The conventional eligibility trace adds a new trace at the top of the existing one amplifying its contribution for reinforcements received almost concurrently (Singh and Sutton, 1996). The traces can be interpreted as a vector of backward horizons for individual states which are multiplied by the components in the update rule of the units. Eligibility traces can be used in any sequential RL rule and create a trade-off between episodic and TD learning.

6.1.3 *Standard benchmarks*

The problems that RL algorithms usually try to solve can be split into two categories based on the nature of their action space: discrete and continuous. *Discrete action space* (DAS) problems are more friendly to *artificial neural networks* (ANNs) with binary output units, as their discrete actions (for example, move up, down, right, or left) can be directly represented by the individual units. On the other hand, the continuous actions of a *continuous action space* (CAS) need to be transformed before they are used by ANNs with binary units. An example of continuous actions could be the torques of joints in walking robots or robotic arms. However, both types of problems can be applied to ANNs with real-valued output units or with extra layers of non-RL units. Although the IC has real-valued units (with constraint activity in $[0, 2]$), in this chapter, only DAS problems were considered. Therefore, some standard DAS benchmarks are described in what follows: the *cliff walking*, *frozen lake*, and *taxi* problems. These differ in the way that the reinforcement is delivered, but they are all based on grid-world spaces (as summarised in Fig. 6.2). These benchmarks are used broadly to test models for sequential RL capabilities.

Cliff walking

In this task, the world is represented by a 12×4 grid, where the starting point (S) is always at the bottom left position and the goal (G) is at the bottom right. The cells in between the start and goal positions in the bottom row represent the cliff (C), which should be avoided. The aim is to go from the start to the goal position without stepping over the cliff. The state is the position of the agent, represented by a vector

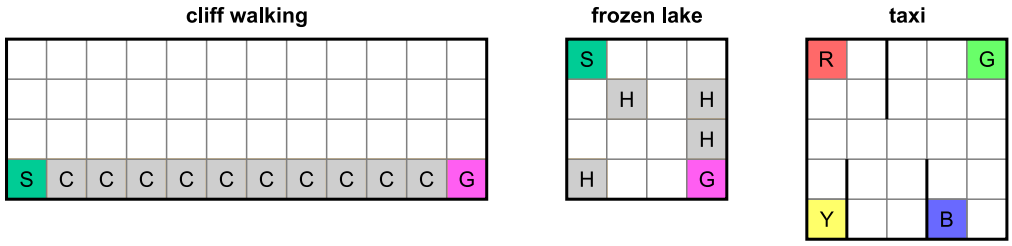

Figure 6.2: *Discrete action space* (DAS) benchmarks based on grid space representation. In the *cliff walking* task, the agent starts at the start (S) position and tries to go to the goal (G) position of the grid, avoiding the cliff (C) cells. Every move is penalised with -1 reinforcement; falling into the cliff comes with a -100. In the *frozen lake* task, the agent starts from the start (S) position and tries to go to the goal (G) position of the grid avoiding the holes (H) of the lake. The white cells represent blocks of ice which means that there is some probability of slipping into an adjacent cell. The agents get a positive +1 reinforcement when reaching the goal, and zero in any other case. In the *taxi* problem, the agent is a taxi driver starting from one of the 4 colour-coded cells: red (R), green (G), yellow (Y) or blue (B). The task is to pick up a passenger from another coloured cell and move it to its destination, which should be in a different coloured cell. There is -1 reinforcement for every step, +20 for dropping off the passenger to the correct cell, and -10 for faulty pick-ups and drop-offs.

with $S_{\text{cliff}} = 37$ dimensions (in the form of an array with size $S_{\text{cliff}} = 37$), where all the dimensions are zero and only one is equal to one (one-hot encoding). Note that the cliff and goal cells are not included in the state representation as they are terminating states (they terminate the episode). This reduces the state space by 11 dimensions. There are $A_{\text{cliff}} = 4$ available actions that allow the agent to move up, right, down, or left. The feedback from the environment comes as a reinforcement of $r(t) = -1$ in every time-step, and $r(t) = -100$ when stepping into a cliff cell.

Frozen lake

In the Frozen lake task, the world is represented by a smaller grid (4×4) compared to the cliff walking task. The grid represents a frozen lake that has holes (H). The agent starts at the top left corner and tries to reach the bottom right corner of the grid without falling into the holes. Similarly to the cliff walking task, the state is a vector with $S_{\text{lake}} = 16$ dimensions, encoding the position of the agent on the grid. $A_{\text{lake}} = 4$ actions represent the 4 distinct moves that the agent can do: move left, down, right, or up. Because of the ice on the lake, every state comes with some probability, emulating slipping on the ice. This means that a step forwards might result in moving to an adjacent cell with some probability. The agent gets a reinforcement of $r(t) = +1$ when successfully stepping on the goal position, and zero otherwise (no immediate punishments for falling into a hole; the 'punishment' is the absence of the reward at the end of the episode).

Taxi

In this task, the map is a 5×5 grid, with 4 predefined colour-coded cells: red (R), green (G), yellow (Y), and blue (B). The agent is a taxi driver that starts at one of the colour-coded cells, and its goal is to pick up a passenger from another colour-coded cell and drop them off at their desired destination (a different colour-coded cell). The state is a vector of $S_{\text{taxi}} = 500$ dimensions, representing the position of the taxi ($5 \times 5 = 25$) for each passenger location (4 coloured-cells + in-taxi = 5) and for each goal location (4 coloured-cells). The taxi driver can take 6 actions: move south, north, east, or west, and pick up or drop off the passenger. For every move, the agent gets $r(t) = -1$ reinforcement, while they get $r(t) = -10$ for faulty pick-ups or drop-offs, and $r(t) = +20$ for a successful drop-off.

6.2 RESULTS

For the RL tasks described before, the RL algorithm must support an arbitrary number of actions. However, the IC supports only two opposite actions ('attract' and 'avoid', or 'do' and 'do not'). Therefore, the simplest extension of the IC was used here (named the m-IC), which is actually m (number of) ICs running in parallel (each one representing a different action), where $m = A^{\text{task}}$ is the number of available actions for each task. The activities of the six MBONs of each IC were used collectively to estimate the Q-value, which was used for the action selection process. The state was represented by the *Kenyon cells* (KCs), which were shared among the different ICs of the model. The parameters for each IC were the same as in Chapter 3, Chapter 4, and Chapter 5, while the KC representation was defined differently for each experiment.

Sarsa, Q-learning, and m-IC were trained to solve the delayed reinforcement tasks described in Section 6.1.3 (cliff walking, frozen lake, and taxi). For each task, the models were trained for $3,000$ episodes and each episode run for a maximum of 200 time steps (or terminated when a termination state was reached). The performance of the models was measured by the average reward they collected during the last episode of training, averaging over 20 runs of the experiments with different random number generators for the models. Two sets of experiments were performed testing different state representations with spatial (KCs that fire together) or temporal correlation (KCs that fire in consecutive time steps, by using eligibility traces).

6.2.1 *Spatially correlated states*

Three alternative representations of the state were tested: the *one-hot*, *spatial*, and *sparse*. All of them represented the state in the form of a vector. The one-hot representation is the one usually used in *machine learning* (ML) and RL, and it is a vector of dimension equal to the size of the state space (S^{task}) with only one non-zero element (usually equal to one). In this representation, there is no correlation between the different states. The spatial representation extends the one-hot adding correlations (shared elements) among states with spatial proximity (in the grid space). For example, in the cliff-walking task, each state has three to eight neighbours and therefore it would share an active element with them creating a correlation based on their spatial proximity. The sparse representation uses the combinatorial encoding approach hypothesised to be computed by the KCs (described in Section 5.5.4). This creates a sparse vector representation, where more than one element is active per state. However, note that the function of the sparse encoding, in this case, is different than the one presented in Chapter 5. The activity of the *projection neurons* (PNs) here is assumed to be the one-hot representation described above, which is already an uncorrelated representation. The sparse representation creates random (with respect to the grid space) correlations among the different states, which are consistent for each task (cliff walking, frozen lake, or taxi). For example, a single PN (which is the only one active in a given state) activates a set of KCs that are connected to it; similarly, another PN (activated in a different state) activates another set of KC that might overlap with the previous set. Thus, a single KC might be active when the agent is in two different states, which creates a correlation between these states.

By using the above state representations, Sarsa, Q-learning, and m-IC were tested in three delayed reinforcement tasks: cliff walking, frozen lake, and taxi (as described in Section 6.1.3). The average rewards they collected during the last episodes of training when using the different state representations are illustrated in Fig. 6.3. The m-IC could not solve any of the tasks (regardless of the state representation used); this was opposed to Sarsa and Q-learning, which managed to solve the cliff-walking and taxi tasks most of the time. However, all three algorithms seemed to perform equally badly in the frozen-lake task. This might be due to the uncertainty of the environment (caused by the random displacements), or because of the very small and sparse reinforcement (only rewarded when the model reaches the goal position during an episode). In any case, the different state representations did not seem to significantly affect the performance of any of the models in any of the tasks. For Sarsa and Q-learning, this might suggest that the spatial and sparse representations are redundant, as the necessary associations among elements of the state vector are formed in

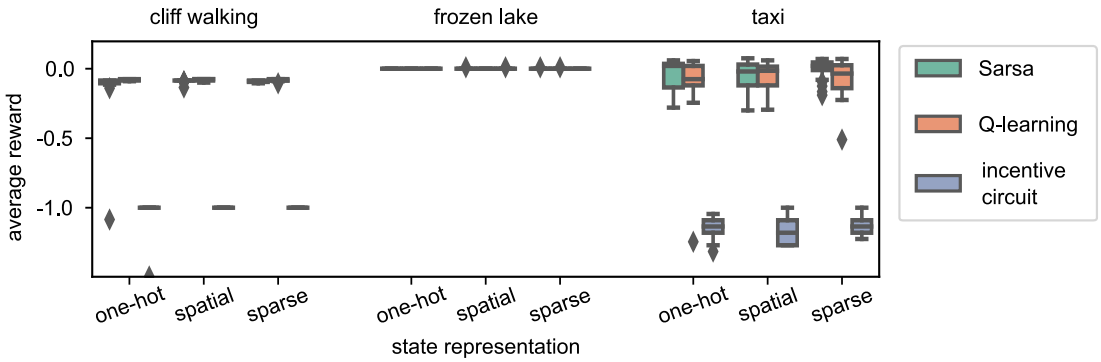

Figure 6.3: The average reward of the last episode (out of 3,000 episodes) for the different models and state representations. Each box plot summarises the average reward of 21 independent runs of the same experiment. The experiments run for different state representations: *one-hot*, *spatial-correlated*, and *sparse*; and for different models: Sarsa (green), Q-learning (orange), and m *incentive circuits* (m-IC, purple).

their Q tables. The same explanation could be valid for the m-IC, which also creates such associations in the KC $\rightarrow$ MBON synaptic weights of the different ICs (representing the different actions). The average reward that the m-IC achieved was around -1 for the cliff-walking and taxi problems, and 0 for the frozen-lake problem. This suggested that it learned to avoid the very punishing states—falling over the cliff (-100) and pick-up or drop-off the passenger in incorrect locations (-10)—but it did not try to maximise the rewards accumulated during each episode.

6.2.2 *Temporally correlated states*

In an episode of the experiments performed in the previous section, Sarsa and Q-learning updated the Q-values of visited states using expectations for reinforcements that would be received in the future (estimated in previous episodes). Averaging the estimated Q-values across these episodes resulted in increasing values along the optimal route (in the state space) towards the goal state. By always choosing a state with increased value, the agent could reach the goal state efficiently. This property came from their plasticity rule (a variation of Bellman's equation) and introduced some temporal correlation. However, this property is omitted from the *dopaminergic plasticity rule* (DPR), which could explain the behaviour observed in the previous section. A reasonable hypothesis is that (for the m-IC) temporal correlations could be introduced in the level of the KC activities (state representation), and this could be realised by eligibility traces. This would allow the current reinforcement to update the parameters related to previously experienced states (proportionally to their temporal proximity to the current state) without the need for this property to be explicitly added to the plasticity rule.

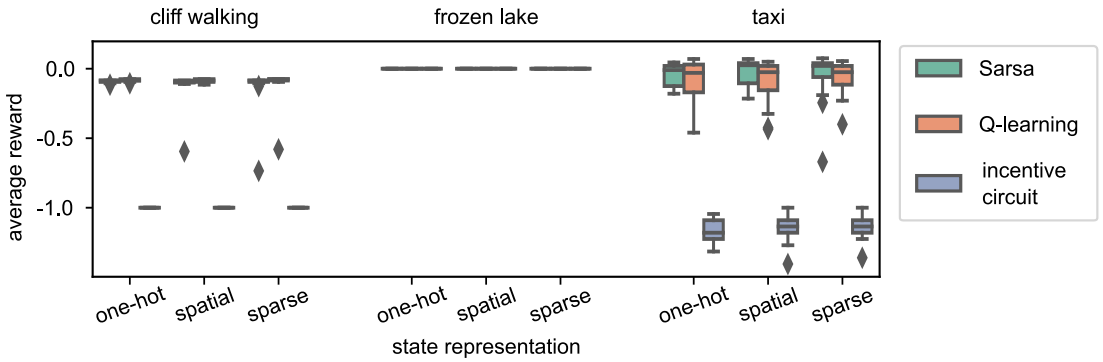

Figure 6.4: The average reward of the last episode (out of 3,000 episodes) for the different models and state representations with added eligibility traces. Each box plot summarises the average reward of 21 independent runs of the same experiment. The experiments run for different state representations: one-hot, spatial-correlated, and sparse; and for different models: Sarsa (green), Q-learning (orange), and m incentive circuits (m-IC, purple).

However, the above hypothesis could not be verified experimentally. By adding eligibility traces to the state representation and repeating the experiments of the previous section, the performance of all three models was unaffected (see Fig. 6.4). The combination of eligibility traces with any of the spatially correlated state representations (introduced in the previous section) had also no apparent effect on the result. This suggested that even when temporal correlations were included in the state representations (eligibility traces), the m-IC model could not account for delayed reinforcement problems.

6.3 DISCUSSION

This is the first time that a computational model of the MBs is challenged in delayed reinforcement tasks; the results of this chapter suggested that the m-IC cannot provide a solution for these tasks. Testing an extension of the IC for multiple actions in different delayed RL benchmarks suggested that the system learns to pair actions with specific states by using the immediate reinforcement, but fails to propagate the delayed reinforcement to states experienced earlier in the episode. This could lead to the conclusion that the function of the MB (as implemented in this work) is closer to an associative RL algorithm than a sequential one. This section discusses these results and suggests potential reasons why the IC could not build a policy of acting over time. The discussion concludes that the MBs might be part of an RL framework and that their role could be related to the one of an adaptive critic.

6.3.1 *The representation of Q-values*

In every time step, ADACs update the cumulative reinforcement value for the experienced state-action pair (also referred to as the Q-value) by using Eq. (6.6). The Q-values are used collectively to extract the policy of the agent, and they could be parallel to the responses of the output neurons in a circuit (like in the IC), where the neuron with the highest response determines the action. Following this hypothesis, the behaviour of the m-IC in the delayed reinforcement experiments was extracted by the MBON activities. For the current KC responses (state), the Q-values for the different actions were calculated by subtracting the activities of MBONs 'against' the action (motivation not to take the action) from the activities of those 'in favour' (motivation to take the action), and collectively calculate one Q-value per IC (action). After training the m-IC for three thousand epochs of two hundred iterations each, all *long-term memories* (LTMs) saturated (as expected), while the restrained and susceptible memories stayed flexible but less important in these long experiments. This explains the binary behaviour extracted from the m-IC and suggests that the combination of the MBON activities (as described above) did not encode Q-values, but just the accumulation of reinforcements (over time and across episodes) per action.

Despite their saturation, LTMs seemed to be the best candidates for encoding the Q-values. This is because of their circuitry, which adds the current reinforcement and a weighted (discounted) MBON response to calculate the responses of the *dopaminergic neuron* (DAN) that modulates the KC synapses onto the same MBON (see Fig. 6.5A); this demonstrates similarities to Eq. (6.5). This equation uses the discounted value of the current time-step—MBON response, $Q[\mathbf{x}(t), a(t)]$—and the reinforcement of the previous time-step—$r(t-1)$—to update the state-action pair of the previous time-step—$Q[\mathbf{x}(t-1), a(t-1)]$ (see Fig. 6.5B). This is equivalent to updating the KC $\rightarrow$ MBON synaptic strengths of the IC that demonstrated the highest Q-value in the previous time-step, based on the Q-value computed by the IC with the highest Q-value in the current time-step. However, in the m-IC all these variables were from the same time-step and the same IC. This suggests that delaying the transmission of the reinforcement and the KC response (internally) could allow the MBs to form LTMs that represent Q-values. However, these values would be calculated for each IC individually and they would still have to deal with the saturating memories.

Note that this is the first time that delayed reinforcements have been explored for the MBs. Previously proposed computational models explored the plasticity and output of the MBs as non-sequential RL models and they considered only immediate reinforcements (similarly to the results presented in Chapter 4). Although other models were not tested explicitly in these tasks, a similar outcome to the one presented

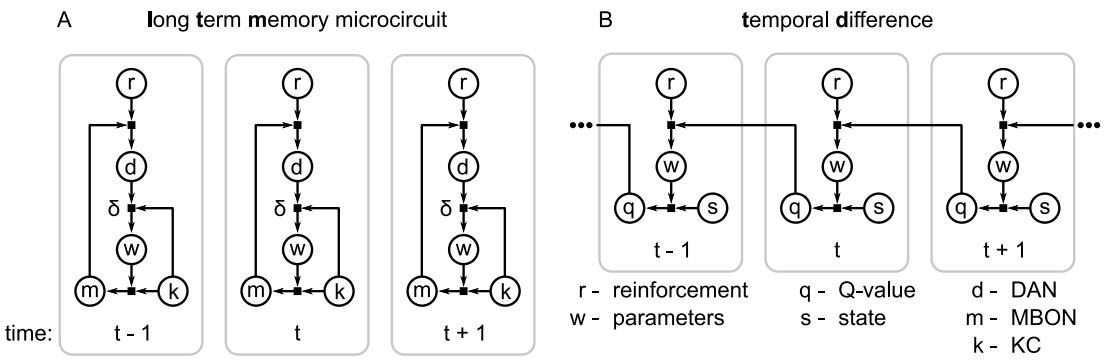

Figure 6.5: Factor graph representation of value-update by the *long-term memory* (LTM) microcircuit and *temporal difference* (TD). The circles denote variables and the squares functions (factors); incoming arrows to factors denote the use of the variable as a parameter of the function. (A) The LTM update of the synaptic weights by using the *dopaminergic plasticity rule* (DPR, δ). The *mushroom body output neuron* (MBON) activity for the current state (*Kenyon cells*, KC) and the reinforcement are used to calculate the response of *dopaminergic neuron* (DAN). The dopaminergic factor is calculated by the DAN response, and together with the KC response, they modulate the synaptic weights (DPR). (B) The TD update of the Q-values. The current reinforcement and the Q-value of the next time-step update the parameters for calculating the Q-value of the current time-step for future use.

here is expected. In fact, the difference between the *reward prediction error* (RPE) and the temporally-different RPE (TD learning) is that for the former the predicted reinforcement is the value itself, while for the latter it is the difference between the predicted values in the current and next time-steps. Similarly to the IC, the output of models that implemented the RPE was the value of two opposing motivations (Bennett, Philippides, and Nowotny, 2021) or a single motivational output (Wessnitzer et al., 2012), which would lead to the same problem when trying to compute Q-values. In order to successfully compute Q-values, the KC $\rightarrow$ MBON synaptic weight of the current state (KC response) and its value with respect to an action (MBON response) has to be calculated based on the response of the respective MBON in the next time-step and the current reinforcement (see Fig. 6.5B). Although it has been demonstrated that fruit flies can solve delayed and trace conditioning problems (Dylla et al., 2017; Grover et al., 2022), this capability is omitted from all the existing models (including the one proposed in this work), and future models should start considering ways to explain how this could be realised computationally.

6.3.2 *Temporally correlated representations*

The plasticity rules play an important role in how the synaptic weights (and Q-values) are affected by the eligibility traces. Bellman's equations update only the Q-value associated with the state-action pair that led to the obtained reinforcement, while the

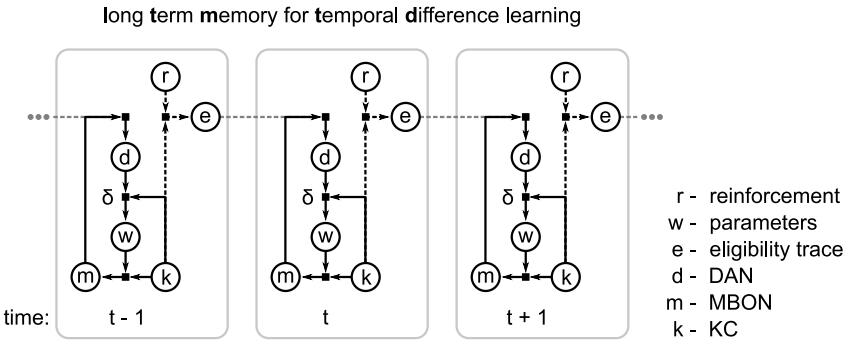

Figure 6.6: Factor graph representation of value-update by the *long-term memory* (LTM) microcircuit with tags from the eligibility traces. The circles denote variables and the squares functions (factors); incoming arrows to factors denote the use of the variable as a parameter of the function. Dashed lines highlight the difference from the LTM microcircuit without eligibility traces. The LTM update of the synaptic weights is done by the *dopaminergic plasticity rule* (DPR, δ). The eligibility traces are calculated by the reinforcement and *Kenyon cell* (KC) activity in the current time-step and affect the response of the *dopaminergic neuron* (DAN) in the next time-step. The *mushroom body output neuron* (MBON) activity for the current state (KC) and the eligibility trace (calculated in the previous time-step) are used to calculate the response of the DAN in the current time-step. The dopaminergic factor is calculated by the DAN response, and together with the KC response of the current time-step, they modulate the synaptic weights.

eligibility traces determine how much the Q-values of recently experienced states will be affected by this change. This means that a limited number of Q-values are updated at a given time step. On the other hand, the DPR affects all the KC $\rightarrow$ MBON synapses of the IC that represents the last action taken. It strengthens the LTM for all the KCs with active eligibility traces in that IC, which might correspond to a different action from the one actually taken when these KCs were active. In addition, due to the LTM charging momentum (see Section 3.3.4), it further strengthens the already formed LTMs in the ICs of previously taken actions, and weakens any contradicting memory to the current eligibility traces. This confuses the system by building up LTMs between states and actions that were never experienced together.

The above suggests that eligibility traces are more useful in the KC $\rightarrow$ MBON synapse level rather than in the level of KC responses. In other plasticity rules, such as the three-factor rule, *spike-timing-dependent plasticity* (STDP) took the role of eligibility traces, and the Hebbian coincidence of KC and MBON activity tagged their synapse for later modulation by the DAN (Wessnitzer et al., 2012). This could allow for the trace to be formed for each state-action pair independently and provide a solution to the problem described above. This solution seems to be in line with data from trace conditioning experiments in fruit flies, which suggested that the reinforcement can be associated with an odour through *adenylyl cyclase* or *protein kinase C* at the level of the KC $\rightarrow$ MBON synapse (Dylla et al., 2017). They also suggested that this (in-

ternal to the synapse) signal is probably involved in the plasticity and it can last for more than $5\,\text{sec}$ (and even $90\,\text{sec}$ in the CX; Grover et al., 2022), which suggests that it could be related to eligibility traces. Using this signal to tag changing synapses for the DPR (most probably through the dopaminergic factor) could lead to more successful use of the eligibility traces. If the tag signal just replaces the KC responses in the plasticity rule, changes in the synaptic strength would still occur through the passive effects of DPR. Thus, tags must come through the dopaminergic factor (gating the synaptic change of the weights), which also provides a potential role for the KC $\rightarrow$ DAN connections (Cervantes-Sandoval et al., 2017; Takemura et al., 2017). Additionally, the long duration of this signal allows for flexibility on when the actual update will happen. Thus, integration of the KC response and the reinforcement signal could initiate the eligibility trace internally to the DAN (see Fig. 6.6). In the next time step the MBON response will be added to the eligibility trace influencing the DAN response, which is used for updating the synaptic weights of the previous state (tagged by the eligibility trace), and essentially implement Bellman's equation.

The above mechanism could work for the IC through its LTM microcircuits. However, in the m-IC model separate ICs were allocated to different actions, where the activity of the output neurons of the one could not affect the synaptic weights of any other circuit. In ADAC models, the Q-value of the current state-action pair was used for the update in the Q-value of the previous state-action pair—see Eq. (6.6). If the readout of the Q-value can be computed by the activity of the neurons representing the action—like in Eq. (6.1)—the exchange of information among the available actions could be the key to building a useful policy. This suggests that potential connections between MBONs and DANs of the different ICs (along with the eligibility-traces extension described in the previous paragraph) could implement Bellman's equations for more than two actions. In the DPR, the DAN activity affects the magnitude of the dopaminergic factor, and as a result, the rate at which the synaptic weight is changing. The activity of an MBON can excite a DAN irrespective of the external reinforcement, which could result in an increasing change rate in the target KC $\rightarrow$ MBON synaptic weight. This property was explored in Chapter 3 and 4, where different types of memories were created in a single IC; however, it could also apply to an extended m-IC (like the one presented in Appendix A, where STMs and LTMs were encoded by the same neurons but for different motivations) and explore whether this can implement a policy for delayed reinforcement tasks. The exact details of the structure and function in such a circuit would need further investigation.

6.3.3 *Reinforcement learning architecture of the insect brain*

In this chapter, multiple ICs were used to compensate for the arbitrary number of actions in the different RL tasks. This approach was based on the assumption that different MBONs affect different actions of the animal, and it resulted in an associative RL architecture (as illustrated in Fig. 6.7A). However, having separate ICs per action seemed unnatural (for the animal) and it made learning in these types of tasks difficult. Therefore, other mechanisms of how the actions might depend on the MBON responses are discussed here as potential extensions of this work in the future.

In the *visual place recognition* (VPR) experiment (discussed in Section 5.4.3) only one IC was used, which could not allow for the calculation of Q-values (as discussed in Section 6.3.1). This is because the LTMs were updated based on the values of the previous time-step (not of the next one as TD learning suggests). However, they also resulted in increasing values over the route (which was the familiarity in the context of VPR). Because the two different motivations ('route following' and 'route finding') were interconnected through the IC, the problem of the disconnected ICs described before did not apply, allowing for antagonistic values to build for the two motivations. As the different groups of MBONs encoded an internal value for their representing motivation (predicting the reinforcement), the IC could have the role of an adaptive critic module. Section 5.4.4 suggested that these motivations could be used as input to the *fan-shaped bodies* (FBs) of the CX (along with input from the *lateral horns* (LHs)) and produce the actions of the animal. This could essentially result in the adaptive critic RL architecture illustrated in Fig. 6.7B, suggesting a comprehensive RL framework where the MBs take the role of the adaptive critic (without necessarily the use of TD learning) and the CX takes the role of the actor. In this case, there is no need for separate ICs per action, as the single IC would provide sufficient information about the value of the motivation, which can be used by the actor to select an action to take. Eligibility traces could extend this model (as discussed in Section 6.3.2) and implement TD learning, resulting in a policy over time for the two motivations.

The architecture in Fig. 6.7B suggests a backwards arrow from the MB (adaptive critic) to the *convergence zone* (CZ) (critic), which does not exist in the architecture of Fig. 6.1C. This represents the MBON $\rightarrow$ DAN feedback connections, which seem to be necessary for the formation of the different types of memories, and (as already discussed in Section 6.3.1) they might also play a role in the formation of the expected cumulative reinforcements. Other than reinforcements coming only from the environment, animals also use internal motivations as reinforcements (represented by this arrow), which might also prove useful for actor-critic architectures. In the CX, the motivational output of the MBs might be integrated based on the action taken (similar to

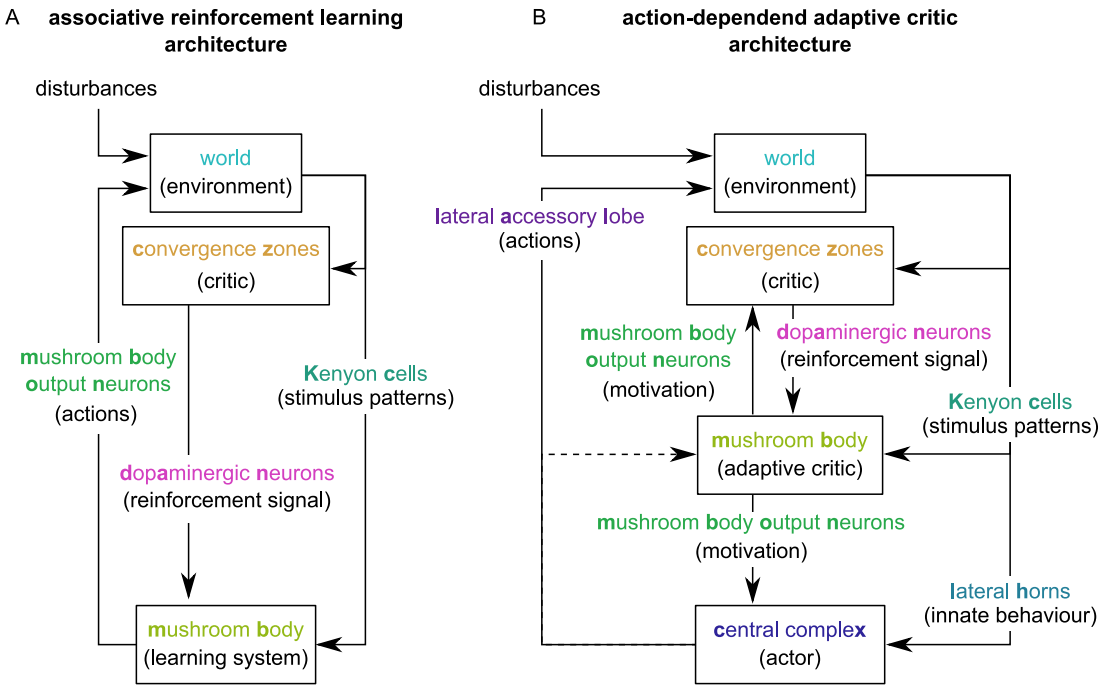

Figure 6.7: *Reinforcement learning* (RL) architectures of the *mushroom body* (MB). (A) In an associative RL architecture, the *Kenyon cells* (KCs) encode information from the world and transmit it to the MB. The *convergence zones* (CZs) also use information from the world to create the reinforcement signal through the *dopaminergic neuron* (DAN) responses. The actions of the agent are then drawn from the *mushroom body output neurons* (MBONs) and influence the behaviour. (B) In an *action-dependent adaptive critic* (ADAC) architecture, the MB takes the role of an adaptive critic, and the MBON responses represent the motivation. This is used from the *central complex* (CX), along with other external input (for example, from the *lateral horn*, LH), which consequently produces the actions in the *lateral accessory lobes* (LALs). If the output of the CX has also terminals in the MB (dashed line), it would allow the actions information to pass to the MB and formulate action-dependent motivations that lead to a policy prior to the CX.

the mechanism hypothesised in Section 5.4.4 for route following), which might allow for the necessary correlations between the states (KC activity) and actions (encoded in the CX) outside of the MBs. For example, in Section 5.4.4 transverse oscillations and scanning behaviours were the actions of the animal, which were also hypothesised to drive integration of the familiarity in different groups of *protocerebral-bridge fan-shape-body noduli* (PFN) neurons in the CX.

The MBs could also take the role of an ADAC in the same RL framework (see dashed line in Fig. 6.7B) providing motivation for different behaviours (sets of actions) handled by downstream actors like the CX. There are more than six MBONs in the MBs, which could represent motivational values for different behaviours. Thus, a model similar to the *incentive wheel* (IW)—which is discussed in Appendix A—could be in place, allowing a motivation for one behaviour to be used as reinforcement for another. Further feedback from MBON → DAN and CX → DAN connections could

set the context for these behaviours and indirectly affect the dopaminergic factor (and therefore the effects of the DPR). The MBONs representing the different motivations could eventually converge in different actors (including the CX) and collectively form the overall behaviour of the animal.

6.4 METHODS

6.4.1 *The OpenAI gym simulations*

The different algorithms were challenged in three different tasks that are benchmarks for RL, and they were selected based on the way that the delayed reinforcement was transmitted. These were the cliff walking, frozen lake, and taxi tasks, which were described in Section 6.1.3. The OpenAI gym Python package (Brockman et al., 2016) was selected to simulate the three problems, as it is a popular package in the community with a big collection of RL simulated environments.

In all the tested tasks, there was a fixed number of possible states, $S^{\text{task}} \in \mathbb{N}$, and actions, $A^{\text{task}} \in \mathbb{N}$, where task $\in \{\text{cliff walking}, \text{frozen lake}, \text{taxi}\}$, while the reinforcement could be any real value. However, as Section 6.1.3 describes, each task had a different number of states and actions. The simulation was returning the current state, $s(t) \in \{0, ..., S^{\text{task}} - 1\}$, and reinforcement, $r^{\text{task}}(t) \in \mathbb{R}$, which were used by the tested algorithm to update its parameters and provide an action, $a(t) \in \{0, ..., A^{\text{task}} - 1\}$. The action was then given as input to the simulator, which returned a new state, $s(t+1)$, and a new reinforcement, $r^{\text{task}}(t+1)$, and so on. Each of these cycles is called an iteration, t, and a maximum of $T = 200$ iterations were run per episode, τ, unless the simulation was reaching a terminating state (earlier termination of the episode). Finally, a total of $\mathcal{T} = 3,000$ episodes were run before the final performance of each algorithm was recorded. To simplify the notation, states and actions for consecutive time steps are noted as $s = s(t)$, $a = s(t)$, $s' = s(t+1)$, and $a' = a(t+1)$.

6.4.2 *Representations of the state*

Regardless of the algorithm used to solve the tasks, the state that was provided by the simulation was transformed into a vector representation. Three alternative transformations of the state representation were used from the tested algorithms: the *one-hot*,

spatial, and *sparse*. In the one-hot representation, the state vector is the s^{th} row of the S^{task}-dimensional identity matrix, which can be expressed as,

$$x^{\text{task}}_{s,j} = (I_{S^{\text{task}}})_{s,j}, \qquad j \in \{0, ..., S^{\text{task}} - 1\}. \tag{6.10}$$

Note that the identity matrix is a square matrix where all its elements are zero except the ones in the diagonal, which are one.

In the spatial representation, the one-hot state vectors were augmented by adding correlations to the states with spatial proximity. This was done by activating elements related to adjacent locations in space, as described below,

$$x^{\text{task}}_{s,j} = \begin{cases} 1, & \text{if} \quad \rho_j = \rho_s \quad \text{and} \quad \kappa_j = \kappa_s, \\ 0.9, & \text{if} \quad \rho_j \neq \rho_s \quad \text{and} \quad \kappa_j \in \{\kappa_{s-1}, \kappa_s, \kappa_{s+1}\} \\ & \quad \text{or} \quad \kappa_j \neq \kappa_s \quad \text{and} \quad \rho_j \in \{\rho_{s-1}, \rho_s, \rho_{s+1}\}, \\ 0, & \text{otherwise}, \end{cases} \tag{6.11}$$

where ρ_* and κ_* denote the row and column in the grid of the environment associated with the state, $*$, and $j \in \{0, ..., S^{\text{task}} - 1\}$.

In the sparse representation, the idea was to transform the state into something similar to KC responses. Thus, the one-hot vector was used as the activity of the PNs, which was processed by Eq. (5.24) to produce the sparse code of the KCs. So, the sparse state vector would be calculated as,

$$x^{\text{task}}_{s,j} = \text{WTA}_{0.0025}\Big[\sum_{k=0}^{S^{\text{task}}-1} (I_{S^{\text{task}}})_{s,k}\, w^{p2k}_{k,j} \Big], \tag{6.12}$$

where $j \in \{0, ..., 4\, S^{\text{task}} - 1\}$, $\text{WTA}_{0.0025}(x)$ is the *winner-takes-all* (WTA) function that allows only $0.25\,\%$ of the KCs (which is, $\frac{1}{100} S^{\text{task}}$ KCs) to fire, and $w^{p2k}_{k,j}$ are the sparse weights connecting the PNs to KCs that were calculated as in Section 5.5.4.

6.4.3 *Eligibility trace*

When the eligibility trace option was used, temporal correlations were added to the state vectors (regardless of the representation). This was done as,

$$x^{\text{task}}_{s,j}(t) = \min[x^{\text{task}}_{s,j} + \lambda\, x^{\text{task}}_{s,j}(t-1), 1], \qquad j \in \{0, ..., S^{\text{task}} - 1\}, \tag{6.13}$$

where $\lambda = 0.9$ is the eligibility traces factor, $\min[*, 1]$ ensured that the state values would not exceed one, $0 \leq x^{\text{task}}_{s,j}(t) \leq 1$, time-independent $x^{\text{task}}_{s,j}$ is the state representation as calculated in the previous section, and time-dependent $x^{\text{task}}_{s,j}(t)$ includes the eligibility traces.

6.4.4 *The incentive circuit for multiple actions*

For the m-IC, the number of ICs used in each task was equal to the number of defined actions, $m = A^{\text{task}}$. Each IC was implemented independently from each other using the methods from Section 3.5, but the same KC responses, $x^{\text{task}}_{s,j}(t)$, were propagated in all circuits simultaneously. As the IC requires two different types of reinforcement, the real-valued reinforcement was transformed into a two-dimensional equivalent. One dimension represented positive reinforcements and the other one negative, which are defined by the positive and negative parts of the same value,

$$r^{\text{task}}_{+}(t) = \max[r^{\text{task}}(t), 0], \qquad r^{\text{task}}_{-}(t) = \max[-r^{\text{task}}(t), 0]. \tag{6.14}$$

These are then used as the reinforcements in the ICs.

The MBON responses were then processed, similarly to Eq. (5.28) for each IC as,

$$Q^{\text{task}}_{s,a} = \frac{m_{a,0}(t) - m_{a,1}(t) + m_{a,2}(t) - m_{a,3}(t) + m_{a,4}(t) - m_{a,5}(t)}{6} - 0.5, \tag{6.15}$$

where $m_{a,0}$ and $m_{a,1}$ are the susceptible, $m_{a,2}$ and $m_{a,3}$ are the restrained, and $m_{a,4}$ and $m_{a,5}$ are the LTM MBONs; and $a \in \{0, ..., A^{\text{task}} - 1\}$ are the different actions associated to the m-IC. The IC whose responses provided the highest value was the one that defined the chosen action, formally written as,

$$a^{\text{task}} = \underset{j \in \{0,...,A^{\text{task}}-1\}}{\arg\max} Q^{\text{task}}_{s,j}. \tag{6.16}$$

6.4.5 *Sarsa and Q-learning*

Both Sarsa and Q-learning use a look-up table (in the same way) to predict actions from states. They first calculate the Q-value of the state for the action taken as,

$$Q^{\text{task}}_{s,a} = \sum_{j \in \{0,...,S^{\text{task}}-1\}} x^{\text{task}}_{s,j} \, w^{\text{task}}_{j,a}. \tag{6.17}$$

The highest value for the given state provided the selected action using Eq. (6.16).

However, the two algorithms updated the values of these tables differently. The update rule for Sarsa was,

$$\Delta w_{s,a}^{\text{task}} = \eta \, [r^{\text{task}}(t) + \gamma \, Q_{s',a'}^{\text{task}} - Q_{s,a}^{\text{task}}], \tag{6.18}$$

while for Q-learning it was slightly different,

$$\Delta w_{s,a}^{\text{task}} = \eta \, [r^{\text{task}}(t) + \gamma \max_{j \in \{0,\ldots,A^{\text{task}}-1\}} Q_{s',j}^{\text{task}} - Q_{s,a}^{\text{task}}], \tag{6.19}$$

where $\gamma = 0.99$ is the discount factor and $\eta = 0.3$ is the learning rate.

6.4.6 *Calculating the average reward*

Fig. 6.3 and Fig. 6.4 summarise the average reward collected by the different algorithms and for the different parameters across all the iterations of the last episode of the experiment. Each experiment was run 21 times, and each time the average reward was calculated as,

$$\langle r \rangle^{\text{task}} = \frac{1}{200} \sum_{t=0}^{T-1} r_{\mathcal{T}-1}^{\text{task}}(t), \tag{6.20}$$

where $\mathcal{T} = 3,000$ are the number of episodes of the experiment, while $T = 200$ are the number of iterations across each experiment.

7

CONCLUSION

> "*Science is made up of mistakes, but they are mistakes useful to make, because they lead little by little to the truth.*"
>
> Jules Verne

This thesis worked towards bridging the gap between artificial *reinforcement learning* (RL) and natural learning by insects. It computationally explored the learning capabilities of the *mushroom bodies* (MBs) of the insect brain, and evaluated their performance in different tasks, aiming to frame them in the context of RL. This chapter summarises the key contributions and ideas of the thesis, as well as the possible future perspectives on how the work of this thesis can enhance the understanding of RL implementations in nature.

The main contributions of this thesis are: (1) The *dopaminergic plasticity rule* (DPR), which allowed for memory potentiation, depression, recovery, and saturation, and explained a variety of learning phenomena when the appropriate temporal resolution was used (for example, in backward conditioning). (2) The *incentive circuit* (IC), which introduced memory dynamics in the MB of the fruit-fly brain; when combined with the DPR it could effectively build *short-term memories* (STMs), *long-term memories* (LTMs), and transfer between them. (3) Using the *mushroom body output neuron* (MBON) responses predicted from the model, the behavioural readout explained a large volume of data from olfactory conditioning and intervention paradigms, which demonstrated the roles of STM and LTM in providing flexible behaviour. (4) Regarding the visual input in the MBs, *principle component analysis* (PCA) whitening and *combinatorial sparse coding* were explored, with the aim of decorrelating the inputs from the insect visual system; they were respectively used to represent *visual projection neuron* (vPN) and *Kenyon cell* (KC) responses. (5) Due to its memory dynamics, the output of the proposed model (which is the combination of the IC and DPR) on the *visual place recognition* (VPR) task demonstrated an increasing output activity for consecutive familiar views and suggested a flexible control of the behaviour produced by the *central complex* (CX). (6) Finally, delayed reinforcements could not be used by the model successfully, and suggested that the MBs might have the role of an *adaptive* or *action-dependent adaptive critic* (ADAC) in a sequential RL architecture.

7.1.1 *The dopaminergic plasticity rule*

The proposed *dopaminergic plasticity rule* (DPR) was derived from calcium imaging data of the *dopaminergic neurons* (DANs) that terminate their axons in the MBs. Its function depended on the input stimulus (KC responses), the dopaminergic factor (which is a factor that depends on the reinforcement and is calculated from the DAN responses), and the target (KC → MBON) synaptic weight. Based on these variables, the DPR enabled different learning effects, such as depression, potentiation, recovery, or saturation. For a given input stimulus, depression and potentiation allow the synaptic weight to decrease or increase respectively, forming a memory that should contribute more weakly or strongly in predicting the output motivation (which is represented by the MBON responses). In the absence of an input stimulus, the depression effect is automatically transformed into the recovery effect, resetting the importance to its initial value. In a similar scenario, the potentiation effect is transformed into saturation, which further increases the value and strengthens the memory. Depending on the dopaminergic factor, the effects of the DPR provided the flexibility to build memories that are enhanced or recovered with the absence of input.

Given a *susceptible memory* (SM) microcircuit, where a DAN depresses the synaptic weight and the MBON inhibits the DAN, the DPR suggested a depression/recovery pair that results in the fast formation of STMs. In a higher temporal resolution, the positive dopaminergic factor (causing potentiation or saturation) follows a negative one (causing depression or recovery), which allowed for all the different effects of the plasticity rule to take place and provided insights for the mechanism behind backward conditioning (Handler et al., 2019). The effects of the DPR showed high dependency on the MBON → DAN feedback connections, suggesting that backwards conditioning might not occur in every compartment of the MB lobes, which has been tested in larval (Weiglein et al., 2021) but not in adult fruit flies. It also suggested that in the absence of feedback connections only specific pairs of dopaminergic effects can occur: potentiation and saturation, or depression and recovery. This suggests that by isolating the effects of a single synapse, only one of the two pairs should be detectable. If Dale's law is in place, this should be true at the level of neurons instead of synapses.

The DPR increases the flexibility of *dopamine* (DA) (as a re-enforcement signal) where feedback is provided through the appropriate circuitry—via KC, MBON, or *anterior paired lateral* (APL) neurons. In fact, these connections could allow the plasticity effect to approximate many of the plasticity rules described in Section 2.1. For example, the *reward prediction error* (RPE) could be implemented by negative or posi-

tive feedback from the target MBON, combined with potentiating or depressing DA effects respectively (Bennett, Philippides, and Nowotny, 2021; Springer and Nawrot, 2021). Inhibitory KC → DAN connections might allow an extension of the basic or template rules (although these connections have been suggested to be excitatory, Cervantes-Sandoval et al., 2017), while post-synaptic MBON modulation could allow for Hebbian plasticity (Takemura et al., 2017). Note that these circuits would implement the different plasticity rules through the DAN responses (affecting the magnitude of the dopaminergic factor in the DPR) without changing Eq. (2.10).

7.1.2 *The incentive circuit*

Following the success of the DPR in explaining backward conditioning (based on the SM microcircuit), the *incentive circuit* (IC) was proposed, which connected six different types of anatomically justified microcircuits of the MB in a single unified circuit. This circuit allowed for complicated memory dynamics through the activation of the DANs by feedback from the MBONs and provided a motivational output signal. The microcircuits that composed the IC were the *susceptible memory* (SM), *restrained memory* (RM), *long-term memory* (LTM), *reciprocal short-term memory* (RSM), *reciprocal long-term memory* (RLM), and *memory assimilation mechanism* (MAM). The SMs allowed for the formation of memories that are easily forgotten when the reinforcement is presented without the conditioned stimulus or with a different input. The RSM allowed for antagonistic learning between two STMs, and their activity was restrained by susceptible MBONs through the RM microcircuits. The DANs of the RSM charged the LTM microcircuits through the potentiating and saturating effects, while the RLM microcircuits added competition to the LTMs by allowing the depressing and recovering effects of the DPR to take place. Finally, the MAM ensured that the memory transferred to the LTMs was eventually removed from the STMs, allowing the acquisition of new experiences. During an aversive olfactory conditioning paradigm, the modelled responses of the neurons in the IC were compared to the ones recorded from fruit flies, showing major similarities in their response trends, especially for the susceptible and STM neurons. The LTM neurons were harder to compare as unknown long-term experiences of the animals could have potential effects on their responses.

The microcircuits of the IC did not implement any of the plasticity rules discussed in the previous section—which are the RPE (via MBON → DAN; Bennett, Philippides, and Nowotny, 2021), basic and template (via inhibitory KC → DAN; Cervantes-Sandoval et al., 2017), or Hebbian (via post-synaptic MBON modulation; Takemura et al., 2017). However, it is possible that microcircuits that implement these rules also exist in the MBs or in other parts of the insect brain. Bennett, Philippides, and

Nowotny (2021) suggested MBON → DAN circuits that implement RPE in the MBs (although using a different plasticity rule), and they demonstrated impressive performance in explaining data from fruit flies. These circuits could be part of a bigger IC (involving more MBONs and DANs), and it might have a role in the overall motivational output of the animal. Connections involving other neurons, like the APL, KCs, or DANs (X. Liu and Davis, 2009; Y. Wu et al., 2012; C.-L. Wu et al., 2013; and involving plastic connections; Zhou et al., 2019), or axo-axonic KC → KC connections (Eichler et al., 2017; Takemura et al., 2017) were not considered here, but they could also have a role in motivational output, especially regarding temporal correlations and delayed reinforcements (as discussed in Chapter 6).

The IC demonstrated simple mechanisms for memory dynamics through MBON → DAN feedback connections, which get more interesting when weaved into the *incentive wheel* (IW) of Appendix A. These could contribute to implementing an equivalent to the mammalian limbic system in the insect brain, suggesting higher-order memories that are paired with emotions or feelings. Although this might look impossible, it would be interesting to start exploring whether such a hypothesis could be valid.

7.1.3 *Memory dynamics for olfactory conditioning*

In experiments of both neuroscience and computational modelling, a common approach that is used to describe the behaviour of the animal is to predict the *preference index* (PI) by using the MBON responses. By taking the same approach, the IC could accurately predict the observed behaviour of fruit flies in a variety of elemental olfactory conditioning paradigms. It could also predict the behaviour of fruit flies in a large number of neural activity intervention experiments, where an MBON or DAN was forced to be active or inactive during classical olfactory conditioning. This is the first time that a single model of the MB can explain so many different olfactory conditioning paradigms. To challenge the model further, a set of closed-loop behavioural experiments were performed; the model was used to control freely moving (simulated) flies in an arena with two odours that were partially paired with some reinforcement. This experiment provided further insights into the function of the IC, demonstrating the role of LTM and MAM.

The PIs estimated by the activities of the model predicted in favour of elemental, two-element, mixture, and overlap learning, and against negative patterning, biconditional discrimination, and blocking. These findings were in line with the observed fly behaviour (J. Young et al., 2011). The PIs also predicted that the behaviour was sensitive to the presenting order of the odours during multi-element training and

testing, suggesting that the behavioural effect of more recent pairings of a *conditioned stimulus* (CS) with the *unconditioned stimulus* (US) were stronger, which was also in line with experimental data (Yin et al., 2009). This effect also caused a bimodal distribution for the PIs of the positive patterning results, which explained the ambiguous results observed when flies were tested in this condition (J. Young et al., 2011). This could be tested experimentally by specifically looking for bimodal distributions in such assays or by analysing positive patterning experiments on a case-by-case basis.

The experiments run with the simulated flies suggested that overlap of KC activity for the different CSs could cause second-order conditioning. Associations related to second-order conditioning were formed directly in the LTM (bypassing the STM), which was in line with findings in fruit flies (Tabone and J. S. d. Belle, 2011). Finally, this experiment provided a potential explanation of why olfactory conditioning showed a stronger behavioural effect when punishments were used compared to rewards (Krashes and Waddell, 2011a; Krashes and Waddell, 2011b). This was because the experiment itself causes asymmetry in the behaviour, as flies prefer to spend more time in positively associated or neutral odours compare to the negatively associated ones. The IC suggests that an asymmetric circuit or plasticity rule is not needed to explain this phenomenon.

7.1.4 *The visual place recognition task*

Inspired by desert ants, the IC was also employed in solving the *visual place recognition* (VPR) task, which is more applicable in robotics; the circuit had to recognise previously learnt (familiar) views among unfamiliar (but spatially proximal) views. Two processing stages were proposed that increased the separability of the visual representations. PCA whitening was suggested to happen prior to the vPNs. This substantially reduced the correlation among the different visual patterns (as proposed before by Laurent, 2002), and it could be implemented, for example, through a calibration process during early excursions of the desert ants (Kühn-Bühlmann and Wehner, 2006). Furthermore, a combinatorial sparse encoding was implemented for the KCs (as suggested by Litwin-Kumar et al., 2017) where a heuristic approach was used to allocate a unique pattern of synapses with one-to-six vPNs for each KC. This process further decreased the correlation among the views in a single route, fully utilising the sparse encoding of the KCs. By using this input, the IC uniquely predicted an increasing MBON response when consecutive familiar views were experienced. This feature could be interpreted as an increased motivation to continue in the same direction, or change the direction when the motivation drops. A potential implemen-

tation of this mechanism using the *lateral horn* (LH) and CX was also suggested but not implemented.

The proposed rendering with a more realistic compound eye (compared to more standard image-like rendering; Baddeley et al., 2012; Ardin, Peng, et al., 2016; Möel and Wystrach, 2020; Sun, Yue, and Mangan, 2020) allowed for changing of the visual focus towards regions of interest (like the horizon). However, varying focuses were not explored in this work; instead, a homogeneous rendering of the environment was used. The ommatidial location and distribution were based on the Fibonacci series and the golden number, which was different to the grid-like distribution of image-like renderings. The previously proposed *Zernike moments* (ZMs) for visual pre-processing (Stone, Differt, et al., 2016; Stone, Mangan, et al., 2018; Sun, Yue, and Mangan, 2020) did not result in rotational invariant features for this rendering method (see Appendix C), suggesting that this property might be sensitive to the rendering of the visual input.

In contrast to random sparse connections, the proposed combinatorial encoding in the KC effectively increased the sparseness of the visual inputs without penalising their separability. The results of Chapter 5 and 6 suggested that this encoding might keep relatively stable correlations among input signals (by decreasing the correlation among correlated visual inputs and increasing it for one-hot encoded states).

The IC suggested an increasing motivation for consistent familiar views that might be an effect of the charging LTM, and it could be related to the increasing value observed in ADAC frameworks. This effect could be particularly useful as input to the *fan-shaped bodies* (FBs) of the CX (which could take the role of the actor), where it could be integrated with actions and innate behaviours coming from the *lateral accessory lobe* (LAL) and LHs (different types of oscillations, like scans; Schwarz, Clement, et al., 2020; normal oscillations; Clement, Schwarz, and Wystrach, 2022; and transverse oscillations; Stankiewicz and Webb, 2021), implementing an RL framework for route following. Compared to previous models that integrated the MBs and CX (Sun, Yue, and Mangan, 2020; Goulard et al., 2021), the proposed (but not implemented) solution would more clearly put the role of the different brain areas in the context of different modules in an RL framework. For example, the MB (adaptive critic) might be able to explain how classical conditioning works in the brain, but the CX (actor) might be required to explain operant conditioning (reinforced *behaviours* will be repeated; Skinner, 1938). Such a model would benefit both fields of RL and neuroethology in the context of computational neuroethology. This also provides a testable hypothesis, where knocking out the MBs should impair learnt motivations (but not completely impair navigation), while knocking out the CX should impair navigation, but do not affect much the motivation of the animal (encoded by the MBON activity).

7.1.5 *Delayed reinforcements*

Solving tasks with temporally sparse or delayed reinforcements is a challenging problem in RL and requires the algorithm to follow a strategy (or policy) to produce actions. When substituted for a sequential RL algorithm (like Sarsa and Q-learning), the m-IC (extension of the IC for m actions) was not able to solve these tasks. Instead, it produced behaviour that avoided punishing (and promoted rewarding) actions in specific states, which seemed closer to an associative RL algorithm. This suggested that the different ICs of the model might be connected (through MBON $\rightarrow$ DAN connections), allowing their value to affect the plasticity on the synapses of other ICs, implementing something similar to the *temporal difference* (TD) learning. Alternatively, a similar model to the one discussed before could be in place, where the MBs took the role of an adaptive critic (or an ADAC), which motivates the CX (actor) to repeat or change its actions.

Temporal correlations in the state representation (introduced through eligibility traces in the KC activity) seemed to not affect the performance of the m-IC in any of the tested delayed reinforcement tasks. However in fruit flies, the pairing between the odour and the reinforcements was unaffected when the reinforcement was introduced up to $90\,\mathrm{sec}$ after the odour (Grover et al., 2022), but it was affected when it was introduced before the odour occurrence (Dylla et al., 2017). Sustained *adenylyl cylase* and *protein kinase C* in the KC $\rightarrow$ MBON synapses (hypothesised to be emitted due to the modulation of the synaptic strength) suggested that eligibility traces should be implemented in the synaptic level and not in the KC activity (which is similar to the suggestion of Wessnitzer et al., 2012). This might partially explain the unsuccessful results of the IC. Another reason was that learning the different state-action pairs independently (without connections among the different ICs of the model) resulted in a non-sequential RL algorithm. The LTMs seemed to encode a similar value to the cumulative sum of the reinforcements for the different state-action pairs, but taking into account only the known values related to the same action (which is represented by the same IC) and ignoring the values related to other actions (other ICs). This suggested that if there was feedback from the LTM MBONs to the respective charging DANs of all the ICs, it would be possible to implement TD learning and enable sequential RL. However, MBON $\rightarrow$ DAN connections are quite specific (rather than all-to-all; Aso, Hattori, et al., 2014; Li et al., 2020), which suggests that the above solution is less plausible. In line with the previous section, it is more plausible that the MB is an adaptive critic in a bigger sequential RL framework, and the CX is an actor that uses the motivational output to perform the actions. Thus, the MBON $\rightarrow$ DAN connections could allow for the calculation of the values (through the LTM) or the

Q-values (through connections from LTM MBONs to charging DANs of different IC motifs, like in Appendix A).

7.2 FUTURE PERSPECTIVES

This thesis explored some interesting properties of the MBs in the context of RL, making important contributions towards understanding their function and demonstrating their capabilities. However, there are still many hypotheses that need to be explored and future directions of study.

The DPR seems to be a powerful plasticity rule that works nicely with the specific circuits proposed in this thesis and provides high flexibility in the synaptic changes. It was used to explain a large number of observed learning phenomena with mostly low temporal resolution. High temporal resolution was partially explored in Section 2.2.2 and provided insight into how backward (or relief) conditioning works, which was verified using detailed calcium imaging (Handler et al., 2019). An interesting future direction would be to explore the capabilities of the DPR in a bigger variety of problems with high temporal resolution, which might provide more holistic insights into the function of DA in the brain.

The proposed IC was built on six MBONs and six DANs, which proved sufficient to encode the demonstrated memory dynamics and explain the behavioural data. However, thirty-four MBONs and a-hundred-and-thirty DANs interact with the MBs of the adult fruit flies, which suggests that a more complicated circuit might be in place, allowing for dynamic memory capabilities, producing more complicated relations among the different motivations. Appendix A suggested a way that the IC could be used as a motif to build a bigger circuit that encodes more motivations, while Section 3.4.3 suggested that the MBONs interacting with the α/βlobe of the MBs might encode a higher level of motivations (which could be analogous to emotions). These discussions are of course speculative and they need further investigation, providing a good basis for future work. Other than MBON $\rightarrow$ DAN connections, KC $\rightarrow$ KC, KC $\rightarrow$ DAN, DAN $\rightarrow$ MBON, and KC $\rightarrow$ APL $\rightarrow$ KC connections have been found in the MBs. Although the function of some of these connections was abstracted in the models of this thesis, a more thorough exploration of these connections (in the context of IC function) would be another interesting future direction.

Section 5.4.4 discussed a hypothesis of how the route following behaviour of desert ants could be implemented in the FBs of the CX. It specifically suggested that the incremental MBON activity for consecutive familiar views and the LH output could be integrated with the *protocerebral-bridge fan-shape-body noduli* (PFN) neural activity in the FB, driving holonomic motion for the animal towards a goal position (for exam-

ple, the nest). Implementing this quite plausible circuit is a reasonable future direction that would demonstrate how the motivational output of the MB could be useful for the navigation-related circuit of the CX, providing insights into the behaviour of central-place foraging insects.

Temporally correlated visual inputs and plastic KC → KC connections were demonstrated (computationally) to encode temporal dynamics in the MBs (Zhu, Mangan, and Webb, 2020). Other than the VPR task, these temporal dynamics might also prove useful for the delayed reinforcements tasks explored in Chapter 6, allowing for higher temporal sparsity, and providing the essential properties to encode eligibility traces in the KC → MBON synaptic weights. These properties might also be related to other MBON → MBON or MBON → DAN connections, which could transfer the values of previous experiences in the DPR through the dopaminergic factor, allowing the implementation of sequential RL. Further investigation of the biological plausibility and implementation of these connections could be another future perspective.

The study conducted during this thesis brought up a more general point about the tasks that computational models of the MBs have been challenged to do so far. Usually, these models try to solve olfactory conditioning (or VPR) tasks, which involve simple associations between the input and output of the model. The structure and function of the MBs seem overly complicated to just be doing simple associative learning. Hence, challenging future models with more complex tasks might be crucial for improving our understanding of their function.

7.3 EPILOGUE

Aiming to bridge the gap between RL and the brain, this thesis proposed a novel plasticity rule and circuit that encode the motivations of insects, realised them computationally, and speculated on their role both in the insect brain and on an RL framework. This theoretical study provided RL solutions that were constrained by their biological plausibility. The biological plausibility was an important factor for this study, as it made it useful in both the fields of biology and *artificial intelligence* (AI). This study provides testable hypotheses for biologists and neuroscientists, which might lead to important discoveries in insect behaviour and cognition. Further, it might contribute to allowing RL models to be implemented in electronic circuits and neuromorphic chips (massively reducing the size and energy consumption of the future autonomous machines, while staying competitive with respect to the high standards of modern AI), providing smaller (and greener) robotic solutions for society.

APPENDIX

A

THE INCENTIVE WHEEL

It has been shown that the *incentive circuit* (IC) can explain classical conditioning experiments that have been done with adult fruit flies, and its neurons can replicate the responses of the *mushroom bodies* (MBs) in the fruit-fly brain. Three types of memories are stored in this model (susceptible, restrained and long-term) for each of the two represented motivations of the animal (attraction or avoidance) guided by reinforcements (reward or punishment). Although this model was sufficient to explain the behaviour of the animals in the laboratory, where the animal was exposed to controlled portions of chemicals and the results are translated into a simple attraction to or avoidance from a source, in the wild, there are more than two motivations that modulate the behaviour of the animal either synergistically or opponently.

Real-life experiences are complicated and rich in information. This could produce a whole spectrum of reinforcements and motivations that guide the behaviour of animals. Data shows that animals respond differently to different reinforcements, which cannot be represented just by the magnitude of a single variable (more or less rewarding or punishing). For example, different concentrations of salt (Y. V. Zhang, Ni, and Montell, 2013), or sugar (Colomb et al., 2009) might be combined with the sated state of the animal, activate different subsets of *dopaminergic neurons* (DANs) and trigger different behaviours, such as feeding or escaping. When the male fruit fly is exposed to female pheromones, courtship behaviour is triggered through P1 neurons (Kallman, H. Kim, and Scott, 2015; Sten et al., 2020), which can be translated to attraction, but has nothing to do with the appetite of the animal. On the other hand, other male pheromones trigger avoidance which suggests that a similar to the IC could explain this behaviour. MBs have been proven to contribute to many behaviours other than olfactory classical conditioning (including visual navigation) and their output neurons encode richer information that is very close to human decision-making (Heisenberg, 2003).

It is reasonable to think that the thirty-four *mushroom body output neurons* (MBONs) and a-hundred-and-thirty DANs interacting with the MBs in the brain of fruit flies are not all used in order to discriminate odours and assign a positive or negative reinforcement to them (driving attraction and avoidance). For this reason, the different MBONs might not represent different odours like it has been proposed before (Huerta, Nowotny, et al., 2004) neither split into two groups (for example, attraction

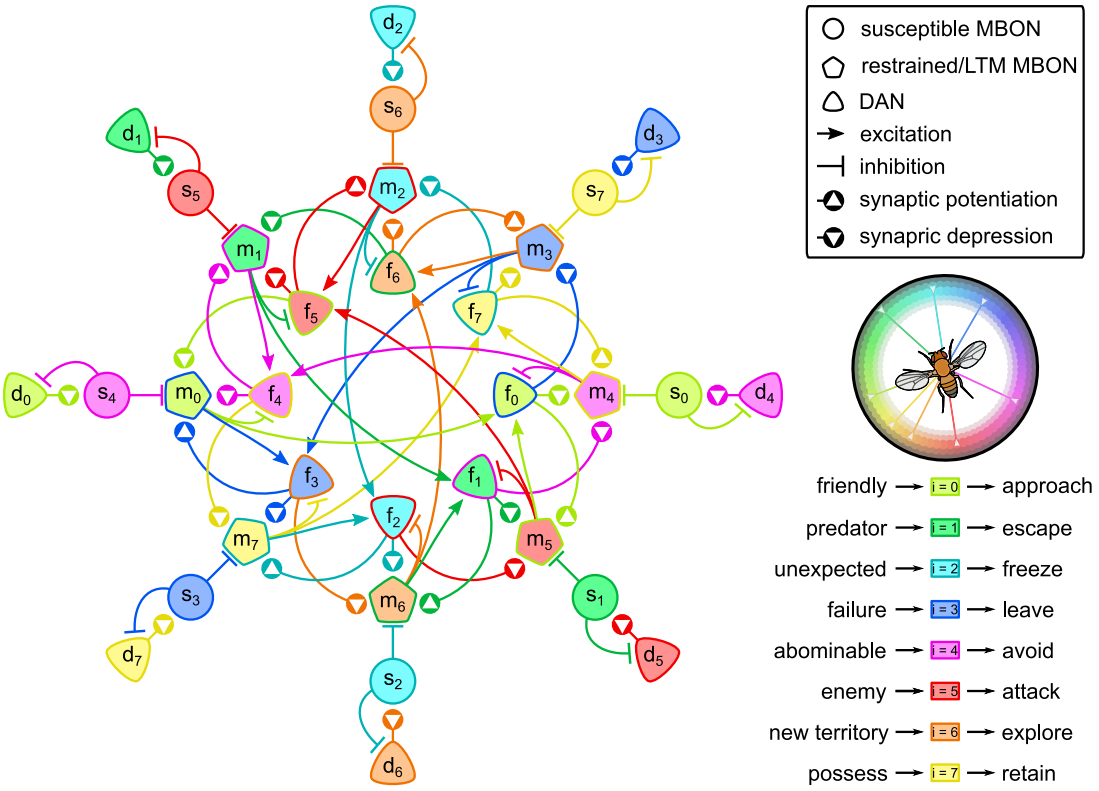

Figure A.1: The *incentives wheel* (IW) model. This model supports that reinforcement is not binary but it draws its values from a spectrum. Different types of reinforcements trigger different *dopaminergic neurons* (DANs) that enable learning in different parts of the *mushroom body* (MB). Colours show the variety of motivations that the model can encode associated with the human 'wheel of emotions' (Plutchik, 2001); for example, *light green*: trust, *green*: fear, *light blue*: surprise, *blue*: sadness, *pink*: disgust, *red*: anger, *orange*: anticipation, *yellow*: joy. Neurons of more than one colour are part of multiple circuits that contribute to different motivations.

or avoidance; Schwaerzel et al., 2003; Schroll et al., 2006; Waddell, 2010), but rather represent different motivations of the animal that all together guide its overall behaviour (Heisenberg, 2003; Krashes, DasGupta, et al., 2009). These motivations might be associated with different contexts, represented by the responses of the *Kenyon cells* (KCs) (as it has been proposed by Cohn, Morantte, and Ruta, 2015, who showed that the same output neurons in the γ compartment respond differently when a different context is given). These contexts enable or disable different microcircuits of the MBs (similar to the ones described in Section 3.3.1) that result in the activation of a subset of MBONs that represent different motivations, while the overlapping microcircuits result in what sometimes is called "noisy" or "insignificant" changes in the behaviour.

Fig. A.1 illustrates such a model, which is named the *incentive wheel* (IW). In this model, four identical ICs were used (C0/4, C1/5, C2/6 and C3/7), where the *reciprocal short-term memories* (RSMs) microcircuit of the one is the *reciprocal long-term memory*

(RLM) microcircuit of another. As the structure of RLM is identical to the one of RSM, it allows for RLM of circuit C0/4 to be the RSM of circuit C1/5, RLM of circuit C1/5 to be the RSM of circuit C2/6, and so on. This way, the different ICs are weaved into an IW with opposing motivations. The reinforcements that trigger the DANs in this model are drawn from a spectrum and the output of the MBONs of the model trigger different motivations. The *long-term memories* (LTMs) and *restrained memories* (RMs) can both exist in the same neurons of the core of the model, representing different motivations in different contexts. This might cause changes in the behaviour of the circuits that are irrelevant to the associated reinforcement, but relevant to a neighbouring reinforcement of the spectrum.

The IW is an example of how the IC can be part of a bigger circuit that can provide a variety of motivations to the animal. An extension of it could have susceptible MBONs, s_i, connecting to other susceptible MBONs from a parallel IW model with higher-order motivations. In Fig. A.1, the different motivations are associated with the primary human emotions from the 'wheel of emotions' (Plutchik, 2001). Higher-order motivations could exist by combining primary motivations (as if they were emotions) and result in more complicated behaviours for the animal.

B

LATERAL INHIBITION AMONG OMMATIDIA

In Chapter 5, the performance of the *incentive circuit* (IC) was challenged in the *visual place recognition* (VPR) task, where the *visual projection neurons* (vPNs) encoded uncorrelated features extracted by raw responses of the ommatidia. Here, the same experiment is tested, but assuming further preprocessing of the ommatidia responses in the *lamina* (LA) of the *optic lobes* (OLs) that introduces *lateral inhibition* (LI). This processing was inspired by the work of Goulard et al. (2021), where they used edge detection features as the input to the *mushroom bodies* (MBs). Edge-related features can be approximated by LI, which is easier to implement for the structure of the compound eye.

B.1 RESULTS AND DISCUSSION

Fig. B.1 summarises the encoded familiarity for both *parallel displacements* and *glancing looks* experiments when LI pre-processing was added prior to the *principle component projection neuron* (PCPN) encoding. Not surprisingly, irrespective of the model that was used to estimate the familiarity—*perfect memory* (PM), *Willshaw network* (WN), or IC—the LI overall caused steeper distributions over the parallel routes compared to the ones shown in Fig. 5.6. This suggests that LI increases the separability of the views further and that even for small lateral displacements it is more difficult for the models to find similar features between familiar and novel views. Such an increase in separability has some advantages and disadvantages, which are discussed below.

There are several reasons why an increased separability of the different visual scenes would be a desirable property. One of them is that it is harder for the models to make an incorrect prediction that the perceived view is learnt, increasing the true positive (predicting that the view is known when it is) and true negative rates (predicting that the view is unknown when it is) in a classification task. This reduces ambiguity in the predictions of the model, which makes its predictions more trustworthy for the animal. However, in the context of VPR, another desirable property is to know how similar is the perceived view to the learnt (familiar) views. The sharp changes between 'familiar' and 'novel' predictions that the LI introduced (see Fig. B.1) do not allow for such a similarity estimation. Thus, it becomes harder for a familiarity gra-

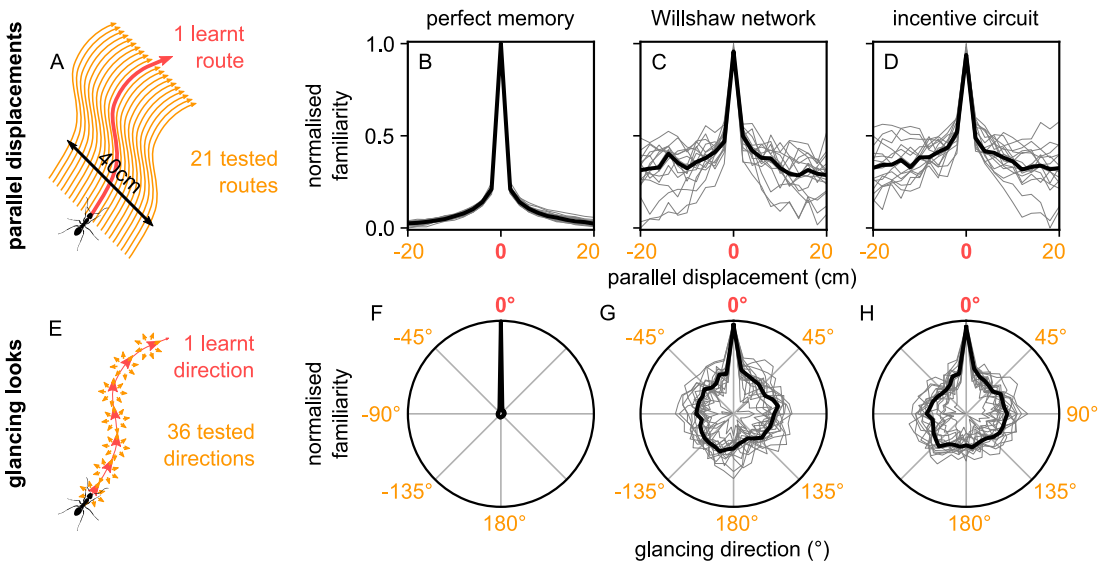

Figure B.1: Comparison of the familiarity predictions among the models, when additional *lateral inhibition* (LI) was applied before the *component projection neurons* (PCPNs), for *parallel displacements* and *glancing looks* experiments. (A) Schematic description of the parallel displacements task. Each model is trained and tested on the same route and additionally tested on 20 shifted copies of this route for up to 20 cm to the left or right. This experiment was repeated for 14 routes, which were recorded from different desert ants, and the average distribution of familiarity across the displacements when using (B) *perfect memory* (PM), (C) the *Willshaw network* (WN), and (D) the *incentive circuit* (IC), is illustrated. (E) Schematic description of the glancing looks task. Each model is trained on a route, like before, and tested on the same route but facing 36 directions in regular intervals, starting from the original direction. The average distribution throughout each route is then plotted for (F) PM, (G) WN, and (H) IC.

dient to be formed around a familiar route, which would be useful for the animal to follow (for example, using transverse oscillations) and meet the route. This is against the idea that the animals can recognise views from far away and use them to cross a familiar route (as suggested by T. S. Collett, Graham, and R. A. Harris, 2007). This suggests that the animals might need intensity information for this task (allowing them to compute the similarity of different views), while LI features might be more useful for the binary classification of the perceived views. It is widely accepted that both intensity and edges (provided by LI) are processed downstream of the OLs, and both of them may enter the MBs allowing both functions to affect the result.

B.2 METHODS

The LI was implemented as synaptic weights in the OLs, where the relative orientation of the individual ommatidia was used (in the form of quaternions), q_ι, similarly to Section 5.5, with $\iota \in \{0, \ldots, n_{\text{omm}} - 1\}$ being the identity of an ommatidium and

$n_{\text{omm}} = 1,000$ the total number of ommatidia. The relative three-dimensional position of each ommatidium was then calculated as,

$$(w_\iota, x_\iota, y_\iota, z_\iota) = q_\iota \cdot i, \tag{B.1}$$

where i is the imaginary number and x_ι, y_ι, and z_ι define the three-dimensional position of the ommatidium in the coordinate system of the eye.

The angular distance between the ommatidium ι and ommatidium ξ can then be calculated as,

$$c_{\iota,\xi} = \cos^{-1}(x_\iota\, x_\xi + y_\iota\, y_\xi + z_\iota\, z_\xi). \tag{B.2}$$

By using this distance, the synaptic weights of a weights matrix that allow LI in the visual input, $\mathbf{W}^{\text{LI}} \in \mathbb{R}^{n_{\text{omm}} \times n_{\text{omm}}}$, can be defined as,

$$w^{\text{li}}_{\iota,\xi} = \begin{cases} 6, & \text{if} \quad \iota = \xi, \\ -1, & \text{if} \quad 0 < c_{\iota,\xi} \leq \min^{7}_{\zeta \in \{0,\dots,n_{\text{omm}}-1\}} c_{\iota,\zeta}, \\ 0, & \text{otherwise}, \end{cases} \tag{B.3}$$

where the function $\min^{7}_{\zeta \in \{0,\dots,n_{\text{omm}}-1\}} c_{\iota,\zeta}$ returns the seventh lowest angular distance across all the ommatidia for the ommatidium ι. The above equation essentially assigns a synaptic weight of strength 6 as self-excitation and inhibitory synaptic weights of strength 1 from the six closest ommatidia. Note that the total incoming synaptic strength sums to zero, which ensures that the overall activity is not changing.

The ommatidia responses were consequently calculated by using the raw ommatidia responses and the above synaptic weights as,

$$r^{\text{li}}_{\iota} = \sum_{\xi=0}^{n_{\text{omm}}-1} r^{o}_{\xi}\, w^{\text{li}}_{\iota,\xi}, \tag{B.4}$$

where r^{o}_{ξ} is the raw response of the ommatidium ξ. For the results of Fig. B.1, the raw ommatidia responses are replaced by the LI responses in the methods described in Section 5.5 before the calibration process of the PCPNs.

C

ZERNIKE MOMENTS FROM OMMATIDIA VALUES

In Chapter 5, the performance of the *incentive circuit* (IC) was challenged in the *visual place recognition* (VPR) task, where the *visual projection neurons* (vPNs) were encoding uncorrelated features that were extracted by the raw responses of ommatidia. Here, the same experiment was tested but assuming a different processing (before the vPNs) that encodes rotational invariant features and corresponds to the *Zernike moments* (ZMs). The ZM features were successfully explored by Stone, Mangan, et al. (2018) and Sun, Yue, and Mangan (2020) for the VPR task. Here their performance was challenged when applied to the non-array-like structure of compound eyes.

C.1 RESULTS AND DISCUSSION

The performance of all three models tested in Section 5.3.3—*perfect memory* (PM), *Willshaw network* (WN), and IC—was tested when the visual input was encoded by a variation of ZM descriptors, which were adapted for the ommatidial input of a compound eye. The familiarity predicted by the PM suggested that the ZMs increased the tolerance of the model towards both translations and rotations. Although Fig. C.1F suggested that the ZM features (as adapted for the compound eyes) were less rotational specific than the *principle component projection neurons* (PCPNs), they were not rotational invariant (as suggested by Stone, Mangan, et al., 2018; Sun, Yue, and Mangan, 2020), which is evident in Fig. C.1F by a dramatic drop of the familiarity for directions that diverged for more than $40°$ from the familiar direction. This provided a useful rotational gradient to the familiarity that was not observed in other vPN encodings, and it suggests that the ZMs-related rotational invariant features observed before (Stone, Mangan, et al., 2018; Sun, Yue, and Mangan, 2020) also depended on the array-like structures of the images they used as the visual input. The ZMs also introduced a wider familiarity distribution for the parallel displacements experiments when the PM was used. Although the ZM representation was not completely rotational invariant (as described before), it demonstrated a major advantage over the PCPN representation when the PM model was used; it demonstrated that the extracted features shared characteristics in both rotational and translational variations.

The performance of the WN and the IC was not as successful (see Fig. C.1C, D, G, and H). This was because the responses introduced by the ZMs were highly cor-

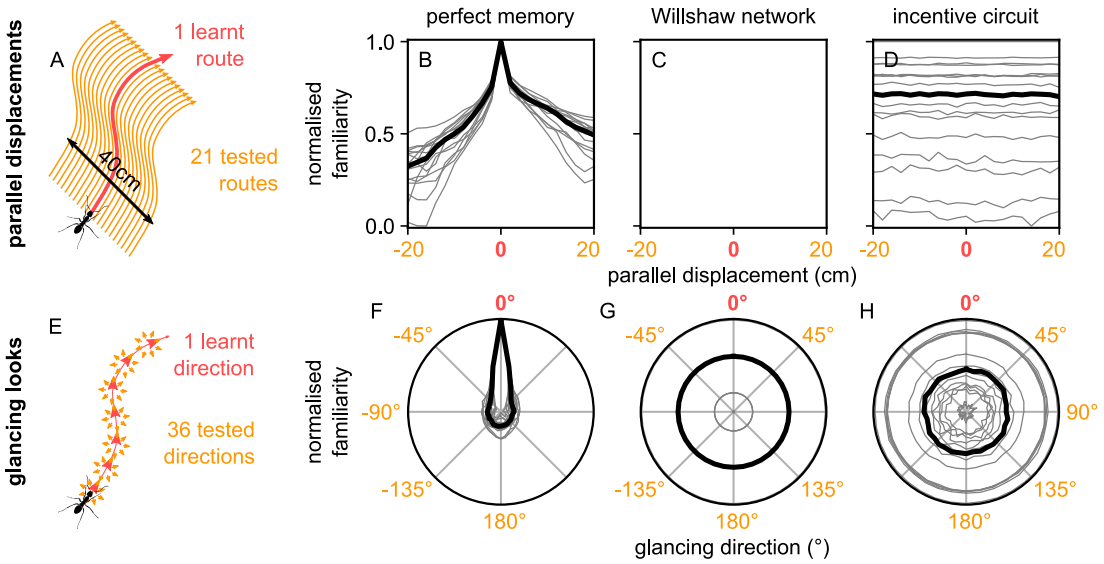

Figure C.1: Comparison of the familiarity predictions among the models, when using the *Zernike moments* (ZMs) instead of the *principal component projection neurons* (PCPNs), for *parallel displacements* and *glancing looks* tasks. (A) Schematic description of parallel displacements task. Each model is trained and tested on the same route and additionally tested on 20 shifted copies of this route for up to 20 cm to the left or right. This experiment was repeated for 14 routes, which were recorded from different desert ants, and the average distribution of familiarity across the displacements when using (B) *perfect memory* (PM), (C) the *Willshaw network* (WN), and (D) the *incentive circuit* (IC), is illustrated. (E) Schematic description of the glancing looks task. Each model is trained on a route (like before) and tested on the same route but facing 36 directions in regular intervals, starting from the original direction. The average distribution throughout each route is then plotted for (F) PM, (G) WN, and (H) IC.

related among the different views (for example, a small number of neurons always dominating the responses). In the frequency domain, it is common that only a few frequencies are high and the rest almost zero (especially for such high-contrast visual inputs). This can potentially provide a sufficient difference among the views for the PM, but it is less useful for the binary encoding in the WN and IC (different frequencies representing the responses of different neurons). Although this was not tested explicitly, it is possible that further decorrelation of this encoding (for example, through a subsequent layer of PCPNs) could resolve this problem and provide more useful input for the *mushroom bodies* (MBs).

C.2 METHODS

For the implementation of the ZMs, the approach used in Sun, Yue, and Mangan (2020) was modified for the customised compound eye. Thus, $n_{\text{zm}} = (16/2+1)^2 = 81$ ZM coefficients were extracted from $n \in \{0, ..., 16\}$ ZM orders. In order to calculate

the Zernike polynomials, the polar coordinates of the ommatidia were needed. These were calculated by transforming the spherical coordinates of the ommatidia to polar coordinates, where the radius, ρ_ι, was calculated by the ommatidial elevation, e_ι, ranging from zero (zenith) to one (nadir), while the polar angle, θ_ι, was equivalent to the ommatidial azimuth, a_ι, but in radians. These transformations were formalised as

$$\rho_\iota = \frac{90^\circ - e_\iota}{180^\circ}, \qquad \theta_\iota = \frac{a_\iota\,\pi}{180^\circ}. \tag{C.1}$$

For each order (n), repeat (m), and ommatidium (ι) the *radial polynomial* was calculated as,

$$R_\iota^{n,m} = \sum_{s=0}^{n-|m|/2} (-1)^s \frac{(n-s)!}{s!(\frac{n+|m|}{2}-s)!(\frac{n-|m|}{2}-s)!}\,\rho_\iota^{n-2s}, \tag{C.2}$$

where '!' denotes the factorial operation. This polynomial was subsequently used to calculate the *Zernike polynomial* as,

$$V_\iota^{n,m} = R_\iota^{n,m}\,e^{-jm\theta_\iota}, \tag{C.3}$$

where j is the imaginary number. Therefore, for each order and repeat, the *Zernike moment coefficient* was calculated by using the Zernike polynomial and the responses of each ommatidium, r_ι^o, as described below,

$$Z_{n,m} = \frac{n+1}{\pi} \sum_{\iota=0}^{n_{\text{omm}}-1} r_\iota^o\,V_\iota^{n,m}, \tag{C.4}$$

where n_{omm} is the total number of ommatidia. The $n_{\text{zm}} = 81$ vPN responses, were the amplitudes of the ZM coefficients for each order and repeat as described below,

$$r_{n,m}^{\text{zm}} = |Z_{n,m}|, \qquad \forall m \in \{..., n-2, n\} \quad \text{where} \quad n \in \{0, ..., 16\}. \tag{C.5}$$

Note that, for each order, only half of the possible repeats were used alternating between 'use' and 'do not use'. Also note that only positive repeats were considered $(m \geq 0)$, as their negative counterparts would result in the same coefficients.

D

THE POLARISED-LIGHT COMPASS

Orientation cues are required for spatial behaviours from following a straight line (Souman et al., 2009) to migrating across continents (Åkesson and Hedenström, 2007) (for a theoretical proof see Cheung and Vickerstaff, 2010). Idiothetic cues such as those generated in the mammalian vestibular system are useful for short periods but are inherently problematic because of accumulating errors. To avoid this limitation, many animals (in particular insects) have developed an array of sensory systems to detect allothetic directional cues in their environments: magnetic (butterflies, Etheredge et al., 1999; moths, Baker and Mather, 1982; ants, Freas and Schultheiss, 2018), wind (moths, Chapman et al., 2008; ants, Wystrach and Schwarz, 2013), and visual—solar compass (honeybees, Greiner et al., 2007; crickets, Labhart, 1988; locusts, Heinze, Gotthardt, and Homberg, 2009; butterflies, Heinze and Reppert, 2011; ants, Labhart, 2000; dung beetles Dacke, Jundi, et al., 2014), star compass (dung beetle, Dacke, Baird, et al., 2013). The benefit of such accurate external compass systems is exemplified in the behaviour of desert ants, who utilise the sky polarisation pattern (Müller and Wehner, 1988). In a habitat with few if any landmarks, these ants can integrate the distance and directions travelled on a tortuous search path of up to a kilometre in length and make a direct return home when food is found (Huber and Knaden, 2015). Our main aim in this paper is to study the potential accuracy with which an insect can estimate its allocentric direction from the sky polarisation pattern, given realistic constraints on the environmental cues, the sensory system, and the various sources of disturbances.

The primary directional cue used for *path integration* by central place foraging insects (such as desert ants; Wehner, 2001) is the sky (sometimes called the *celestial compass*). The position of the sun (or moon) in the sky—as well as providing a direct directional reference point—determines the properties of light across the skydome including intensity and chromatic gradients, and a specific pattern of polarisation. Linear polarisation of light is the alignment of the orientation of oscillation of the electromagnetic wave to a single plane. As light from the sun passes through the earth's atmosphere it undergoes a scattering process (Strutt, 1871; J. W. S. B. Rayleigh, 1871; L. Rayleigh, 1881; L. Rayleigh, 1899) producing different levels of polarisation across the sky-dome relative to the position of the sun. From the point of view of an earth-based observer, as the angular distance from the sun increases from $0°$ to

90°, the degree of linear polarisation in skylight increases, with the principal axis of polarisation perpendicular to the observer-sun axis, forming concentric rings around the sun. Angular distances above 90° have decreasing degree of linear polarisation (Coulson, 1988).

The insect celestial compass has been studied extensively in the honeybee *Apis mellifera* (Menzel and Snyder, 1974; Greiner et al., 2007), cricket *Gryllus campestris* (Labhart, 1988; Labhart, 1999; Labhart, Petzold, and Helbling, 2001; Heinze and Homberg, 2007), locust *Schistocerca gregaria* (Vitzthum, Müller, and Homberg, 2002; Heinze, Gotthardt, and Homberg, 2009), monarch butterfly *Danaus plexippus* (Heinze and Reppert, 2011), dung beetle *Scarabaeus lamarcki* (Dacke, Jundi, et al., 2014) and desert ant *Cataglyphis bicolor* (Labhart, 2000). Insects perceive polarised light through a specially adapted region of their upper eye known as the *dorsal rim area* (DRA). For the ommatidia in this area, the light-sensing elements (microvilli) do not twist relative to each other, resulting in units that are sensitive to specific polarisation angles. Ommatidia in the DRA are connected to *polarisation sensitive* (POL) neurons in the *medulla* (ME) of the insect *optic lobe* (OL) which follow a sinusoidal activation profile under a rotating linear polarisation input (Labhart, 1988). The maximum and minimum activation is separated by 90°, consistent with an antagonistic input from at least two polarisation-sensitive channels with orthogonal *e-vector* tuning orientation (the e-vector is the electric vector component of the light's electromagnetic energy and is orthogonal to the direction of propagation). The identity and spike rate of each POL neuron thus encodes information about the *angle* (AOP) and *degree of polarisation* (DOP) respectively for the specific region of sky from which the associated ommatidia samples. An array of such sensing elements arranged appropriately may hence be sufficient to decode the sun position, without using any additional sky cues (Bech, Homberg, and Pfeiffer, 2014).

From the OL the pathway for polarised light processing has been traced through several neuropils in the insect brain. Key processing stages for polarised light are the dorsal rim lamina and medulla, specific layers of the lobula, the anterior optic tubercle, and the lateral complex with giant synapses; before reaching a highly structured midline neuropil known as the *central complex* (CX) (Vitzthum, Müller, and Homberg, 2002; Homberg et al., 2011). A variety of neuron types within the CX region have been shown to have polarisation-dependent responses, including CX inputs in the form of three types of tangential neurons (Vitzthum, Müller, and Homberg, 2002) (TL1, TL2, and TL3) which synapse with columnar neurons (Heinze, Gotthardt, and Homberg, 2009; Turner-Evans et al., 2017; Green et al., 2017) in the ellipsoid body/lower division of the central body. Most strikingly, intracellular recordings from neurons in the CX *protocerebral bridge* (PB) have revealed an orderly polari-

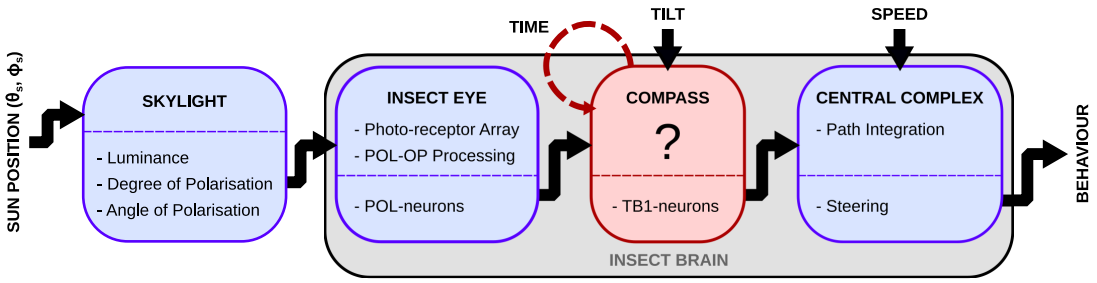

Figure D.1: Overview of the modelling pipeline. The simulation consists of four consecutive models (left to right). Given the position of the sun, a realistic skylight model Wilkie et al. (2004) provides the predicted luminance, degree and angle of linear polarisation for every point in the sky-dome. This provides input to the eye model, based on biological data from insects (Labhart, 1986; Labhart, 1988), which defines an array of polarisation-sensitive photo-receptors facing different parts of the sky, and uses opponent processing to produce luminance-independent POL-neuron responses. The compass model provides a hypothesis for the unknown neural process that converts the POL-neuron response to a true compass signal in the TB1 neurons; this also utilises information about the tilt of the sensor array and allows for the movement of the sun with passing time. Finally, the compass neurons' output is used along with speed as the input to an anatomically grounded model of the *central complex* (CX) (Stone, Webb, et al., 2017) which performs path integration and produces an output signal that can steer the insect back home. Blue boxes represent known systems in the pathway of the skylight; the red box represents a fuzzy/unknown system, which is the main focus of modelling in this paper.

sation tuning preference across the eight columns, described as an internal compass (Heinze and Homberg, 2007). More recently, the same structure has been observed through neurogenetic imaging in *Drosophila* to act as a *ring attractor* encoding the heading direction of the insect relative to a prominent visual cue (Seelig and Jayaraman, 2015). However, there remains an inconsistency between these observations, as the polarisation tuning appears to range from $0°$ to $180°$ with each successive column tuned $\sim 22.5°$ degrees apart (Heinze and Homberg, 2007), whereas the fly's compass covers at least $270°$, changing by around $35°$ per column (Seelig and Jayaraman, 2015) and might be assumed to form a $360°$ representation of space (Green et al., 2017).

In a recent model, we used anatomical constraints for processing within the CX to explain how compass information in the PB could be combined with speed information to carry out path integration, and subsequently steer home (Stone, Webb, et al., 2017). This model assumed a $360°$ compass across the 8 tangential cells (TB1) of the PB (PB) could be derived from sky polarisation cues. In this paper, we first determine whether in principle such a signal can be recovered from a simulated array of POL-neurons stimulated by a realistic sky polarisation pattern. We further investigate whether this signal can deal with, or be plausibly corrected for, potential disturbances such as partially obscured sky, tilting of the sensor array, and the movement of the sun with passing time. We evaluate the potential accuracy of compass information

DRA layout			photo-receptors		optic lobe neurons	tilt compen-sation	time compen-sation	compass in CX
n	ω	shape	λ_{max}	ρ				
60	56°	fan-like	350 nm	5.4°	POL-OP	ocelli, vis. horizon	ephemeris function	heading code
(*desert ant*)			(*desert ant*)		(*desert ant* & *cricket*)	(*desert ant* & *other*)	(*desert ant* & *honeybee*)	(*desert locust*)

Table D.1: The cross-insect properties of our model. n, number of ommatidia on the sensor; ω, receptive field of the sensor; λ_{max}, maximum spectral sensitivity (wavelength) of a single photo-receptor; ρ, acceptance angle of a single photo-receptor. Specific species of insects for each property: *dorsal rim area* (DRA) layout—*C. bicolor* & *C. fortis* (Wehner, 2003), *C. albicans*, *C. bicolor* & *C. fortis* (Zollikofer, Wehner, and Fukushi, 1995). photo-receptors—*C. bicolor* (Labhart, 1986). *optic lob* (OL) neurons—*C. bicolor* & *G. campestris* (Nilsson, Labhart, and E. Meyer, 1987). tilt compensation—*C. bicolor* (Fent and Wehner, 1985), various insects (Goodman, 1970). time compensation—*C. bicolor* & *C. fortis* (Wehner and Müller, 1993), *A. mellifera ligustica L.* (Towne, 2008). compass in *central complex* (CX)—*S. gregaria* (Heinze and Homberg, 2007).

both in absolute terms and in the context of path integration. Finally, we show how the discrepancy in biological data for the PB tuning pattern (Heinze and Homberg, 2007; Seelig and Jayaraman, 2015) might be resolved, by testing our model with artificial polarisation patterns.

D.1 METHODS

D.1.1 *Overview*

This study investigates how navigating insects can transform solar light into an earth-based compass signal that is sufficiently stable and accurate to drive precise path integration behaviour. Fig. D.1 provides an overview of the modelling pipeline. We start with a physical model of the skylight, which is used as input to a biomimetic sensor array based on the desert ant eye. We then take a more direct computational approach to generate compass output from the insect eye input, by defining a hypothetical neural architecture that will reconstruct the sun position from this input, with additional mechanisms to correct for tilt and for passing time. As the precise neural connectivity underlying these transformations in the insect is unknown, this is a proof of principle that can provide hypotheses for future investigation of this circuit (see discussion). We then use the output of the compass as input to an existing biologically constrained model of path integration in the CX and test it in a closed-loop agent simulation. The properties of our model are drawn from a variety of insects that are shown to have

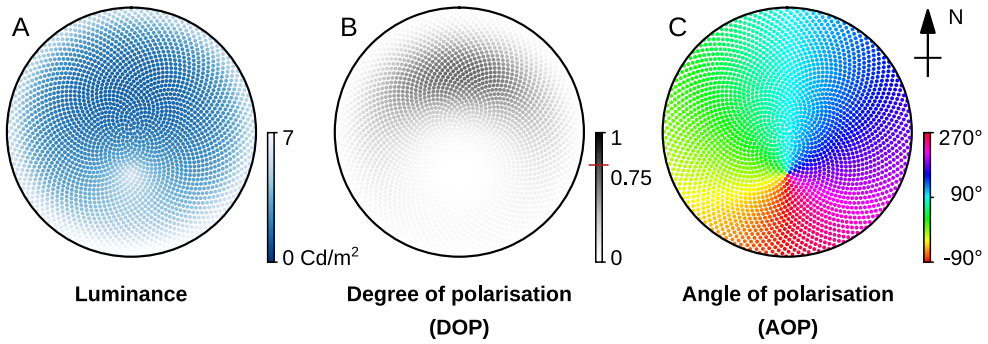

Figure D.2: Sample output from the skylight dome model (Wilkie et al., 2004). (A) The luminance pattern of the sky is proportional to its intensity and describes the amount of light per area unit existing in a specific direction. Along with the chromaticity coordinates, it can provide spectral information. (B) The degree of linear polarisation pattern in the sky based on the scattering of the light on atmospheric particles. It is defined by the fraction of the polarised portion over the total intensity. The red line on the colour bar showing the $d = 0.75$ indicates the maximum *degree of polarisation* (DOP) observed in the skylight simulation. (C) The angle-of-polarisation pattern in the sky is defined by the average e-vector (electric part of an electromagnetic wave) orientation of the photons. The black circle in the figures denotes the horizon. In all panels the sun position is $30°$ south.

a celestial compass, as detailed in Table D.1, but with a specific focus towards the desert ant.

D.1.2 *Skylight*

To test our neural model, we need to simulate the incoming light using a skylight dome model. Previous computational studies of the insect POL-system have often copied the typical stimulus input from experimental studies (rotating linear polariser; Sakura, Lambrinos, and Labhart, 2008). However, the topology of the ommatidia and the neural processing of the compass system in the insect brain have evolved under real sky conditions, hence using a more realistic input can be critical to understanding the function. Specifically, as we will show, the real sky pattern breaks the symmetry conditions that inherently prevent $360°$ directions from being recovered from a simple linear polarisation cue.

We use the skylight dome model described in Wilkie et al. (2004), which gives a very realistic luminance and linear polarisation information pattern (a sample of its output is illustrated in Fig. D.2). This model is the most accurate description of the skylight distribution currently available, and for a detailed description, we refer the reader to the original work (Wilkie et al., 2004) which we follow directly. Given the position of the sun and a set of points in the sky, this model generates the *luminance*, *degree* (DOP) and *angle of polarisation* (AOP) for those points. Tuning by geo-referenced

input parameters allows realistic sky patterns to be estimated for specific locations. Therefore, plugging into the model the location from our own desert ant fieldwork site (Seville, Spain) allows us to run simulated experiments for desert ants. This way, we can study the response of their POL-sensitive neurons using near-natural stimuli.

D.1.3 *The insect eye*

Dorsal rim ommatidia

We developed a simulated sensory unit to mimic the function of the ommatidia in the DRA of the desert ant (Fig. D.3A and B). We consider each ommatidium to be an individual sensorial unit that can be freely arranged in space to match the compound structure of the real ant's DRA (Fig. D.3D and E).

For an insect, the acceptance angle of the ommatidia affects the volume of the skylight perceived and the maximum polarisation contrast sensitivity. This is defined by the optical properties of the cornea and the crystalline cones. Light passing through these lenses is focused through a light guide onto 8 light-sensitive cells (rhabdoms) that have a preference for one of two groups of perpendicularly oriented microvilli, that work as linear polarisation filters (Fig. D.3C).

For each simulated ommatidium in our model, we define a fixed acceptance angle, $\rho = 5.4°$, and spectral sensitivity, $\lambda_{\text{max}} = 350\,\text{nm}$ wavelength, based on the ant eye (Labhart, 1986) and use 2 perpendicularly arranged photo-receptor channels. We notate $s_{\parallel}$ ($s_{\perp}$) the stimulus of the photo-receptor channels that have parallel (perpendicular) aligned polarisation filters with respect to the ommatidium orientation. These follow a sinusoidal response curve to the e-vector orientation with the highest value when fully aligned with the preferred polarisation angle and the lowest when perpendicular. The photo-receptor neuron transforms this raw input using a square-root activation function, $r = \sqrt{s}$ which reduces right skewness, transforming the sinusoid as shown in Fig. D.3G. This transformation is essential in order to penalise the high illuminations of the bright sky. It acts similarly to the logarithmic transformation $r = \log(s)$, introduced by Nilsson, Labhart, and E. Meyer (1987) and used on the Sahabot robot (Lambrinos, Kobayashi, et al., 1997). The main advantage of the square-root transformation is that it can be applied to zero values (darkness or linearly polarised light perpendicular to the filter orientation). As the exact transformation of the photo-receptor signal in the insect's eye remains unknown (Rossel and Wehner, 1986), any transformation that reduces the right skewness could be theoretically correct.

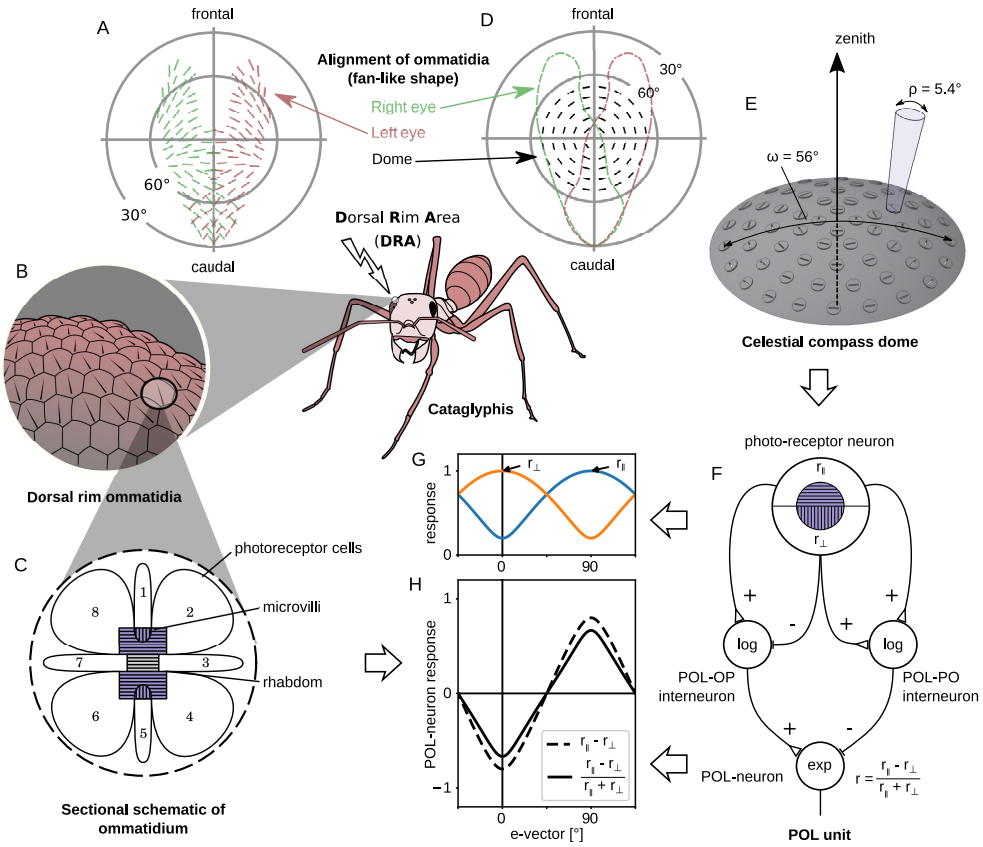

Figure D.3: Processing stages of light in the biological and artificial *dorsal rim area* (DRA). (A) Top view of the fan-like arrangement of the ommatidia on the *Cataglyphis* DRA for both the right (green) and left (red) eyes; adapted and modified after Wehner (2003). (B) A closer look at the DRA, which is composed of hexagonal ommatidia. (C) An ommatidium on the DRA of the compound eye of the *Cataglyphis* has 8 photo-receptor cells, with parallel microvilli direction in 2, 3, 4, 6, 7 and 8, and perpendicular in 1 and 5; the colour violet indicates sensitivity to ultraviolet light. (D) Top view of the fan-like arrangement of the units on our sensor. The dashed lines show the overlap with the areas of the left (red) and right (green) *Cataglyphis* DRAs. (E) 3D representation of the sensor array in the eye model with visual field $\omega = 56°$: the 60 discs on the dome are different units (ommatidia) with acceptance angle $\rho = 5.4°$; the orientation of the lines on the circles denote the direction of the main (parallel) polarisation filter. (F) Model of a POL-unit: the photo-receptor neurons combine an ultra-violet (UV) sensor (photo-receptor) and a polarisation filter (microvilli), and have a square-root activation function. The normalised difference of the photoreceptor neurons is calculated by the POL interneurons. The empty triangular and dashed synapses denote excitatory and inhibitory connections respectively. (G) Simulated response of the two photo-receptors in one unit in partially linearly polarised light of intensity $I = 1$ and *degree of polarisation* (DOP) $d = 0.9$ against different e-vector orientations. (H) Simulated response of the POL-neuron to the input of Fig. D.3G; the dashed line shows the response of the POL-OP interneuron, and the solid line is the response of the POL-neuron (normalised difference). B and C figures adapted and modified from W. Zhang, Cao, X. Zhang, and Z. Liu (2015). F, G and H are after Labhart (1988).

POL-neurons: polarisation contrast

The first stage of the global-orientation encoding in insects is performed by the *polarisation* units in the medulla of the OL (Labhart, 1988). These units encode polarisation

information from different points of the insect's environment, creating a polarisation map. We follow the approach described in Nilsson, Labhart, and E. Meyer (1987) to define how light perceived from our sky model is transformed into POL-neuron activity. This produces luminance insensitivity in each POL-neuron, capturing only the polarisation contrast in specific e-vector directions.

The photoreceptor cells propagate their response to the POL-neuron output via two interneurons (see Fig. D.3F). The *polarisation-opponent* (POL-OP) interneuron (left) computes the difference between the responses of the photoreceptor neurons, $r_{\text{OP}} = r_{\parallel} - r_{\perp}$, similarly to the output (POL-neuron) in Lambrinos, Kobayashi, et al. (1997). The *polarisation-pooling* (POL-PO) interneuron (right) approximates the overall intensity of the input light, by pooling the responses of the opponent components, $r_{\text{PO}} = r_{\parallel} + r_{\perp}$, and is used as a normalisation factor that removes this luminance information from the signal. Both interneurons propagate the logarithm of their output signal to the POL-neuron.

The POL-OP interneuron excites the output POL-neuron, while the POL-PO interneuron inhibits it. As a result, the activity of the output neuron is independent of the luminance. This feature is very important, as the activity of the POL-neurons in insects' OL is not correlated to the light intensity (Labhart, 2000; Wehner, 2003). This has been previously modelled by normalising the response of POL-neurons using the values measured from different directions (Lambrinos, Kobayashi, et al., 1997; Labhart, 1999; Lambrinos, Möller, et al., 2000; F. Smith and D. Stewart, 2014) (scanning model). Finally, the output neuron transforms the response using an exponential activation function in order to bring back the sinusoidal shape:

$$r_{\text{POL}} = \exp\left[\log\left(\frac{r_{OP}}{r_{PO}}\right)\right] = \frac{r_{OP}}{r_{PO}} = \frac{r_{\parallel} - r_{\perp}}{r_{\parallel} + r_{\perp}} = \frac{\sqrt{s_{\parallel}} - \sqrt{s_{\perp}}}{\sqrt{s_{\parallel}} + \sqrt{s_{\perp}}} \tag{D.1}$$

The above set of interneurons, along with the photo-receptors and the output POL-neuron compose a POL unit (Fig. D.3F). Fig. D.3G illustrates the response of the antagonistic photo-receptor neurons in a POL unit when the perceived light is partially and linearly polarised (DOP, $d = 0.9$; which is higher than found in skylight, but possible in experimental situations). The dashed line in Fig. D.3H illustrates the response of the POL-OP interneuron and the solid line shows the normalised POL-neuron response after dividing (inhibiting) by the POL-PO signal.

The output signal of our POL-neuron thus matches the one found in ants' and crickets' POL neurons (Nilsson, Labhart, and E. Meyer, 1987). However, we note there is no specific evidence for the hypothesised interneurons in this model.

Dorsal rim area: layout

The layout of the sensor approximates a joint DRA (from both compound eyes; see Fig. D.3A) of an ant, leading to a cyclopic DRA. We approximate the sampling pattern of the cyclopic DRA by homogeneously distributing simulated ommatidia on the dome using the *icosahedron triangulation* method (Fig. D.3D), as widely used in computer graphics for sphere representations (detailed description of the design of the sensor in Fig. E.27 and a table with the spherical coordinates of all the POL units on the dome in Table E.3). The resulting pattern of polarisation preferences in our population of simulated ommatidia matches the overall specifications of the fan-like shape reported in ants (Labhart, 1986; Wehner, 2003; Zollikofer, Wehner, and Fukushi, 1995) (Fig. D.3A). More specifically, as Fig. D.3E indicates, each ommatidium, with *acceptance angle* $\rho = 5.4°$, is aligned to its respective concentric notional ring centred at the zenith of a dome. Using the above method, we place $n = 60$ ommatidia on a dome-shaped surface, and the outer notional ring with radius $28°$ results in an $\omega = 56°$ *receptive field* for the whole eye model. Note that the view of each ommatidium can partially overlap with its neighbours. The *receptive field* is around half that of the ants' DRA on the frontocaudal axis (see Fig. D.3D), and the total number of ommatidia are similarly 50% that of the real animal, keeping the resolution approximately the same.

The output of the ant eye model is a population $(n = 60)$ of neurons closely matching the known characteristics of the POL neurons of the medulla of the crickets' OL (Labhart, 1988). This population thus forms the biologically constrained input layer for our visual processing model.

D.1.4 *The compass*

We built a *compass model* to transform the responses of the POL-neurons into the desired activation patterns of the TB1-neurons used for path integration. Although the anatomical pathway is known (Schmitt et al., 2016), the neurobiological processes on this pathway are as yet uncertain, so we have taken an information processing approach: given biologically realistic POL-neuron responses gathered from the skylight simulation we examine whether this input provides enough information for a biomimetic CX model to drive steering. Fig. D.4 shows an overview of the model. The connection of POL-neurons to SOL-neurons in the *solar layer* implements a *sum-of-sinusoids* model to recover an estimate of the solar azimuth. The *gating function* adjusts the connection weights to compensate for the tilting of the sensor array. The *true compass layer* uses the confidence of the estimate to predict and compensate for

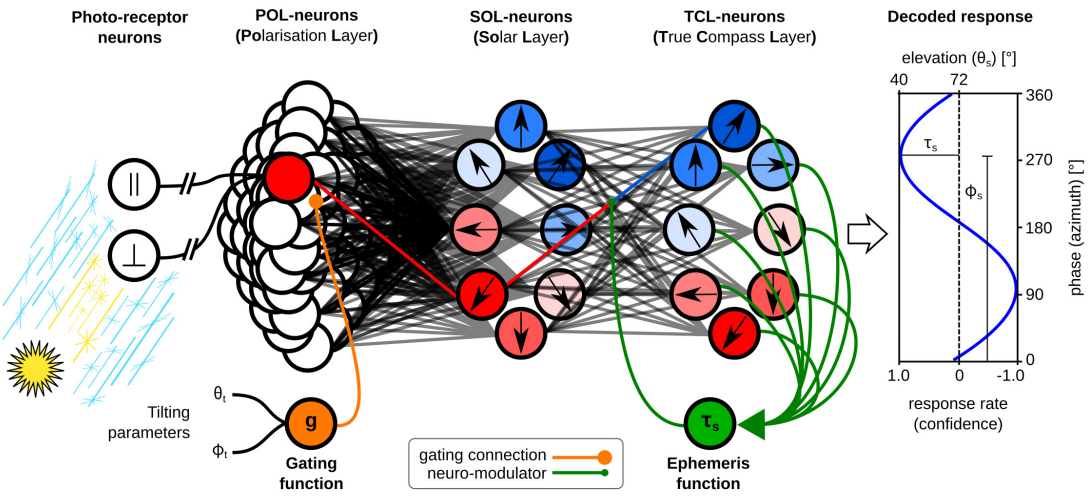

Figure D.4: Overview of the compass model. The light information stimulates the photoreceptor neurons. Using the eye model, this is transformed into a POL-neuron population code. Tilting information (orange) is propagated through the gating function and creates a mask for the POL-neurons' response, altering the relative weighting of information from different parts of the sensor array (see Fig. D.5). By combining information across the array, the response of SOL-neurons encodes an estimate of the solar azimuth and elevation—black arrows indicate the $(\phi_{\text{SOL}}^{\text{pref}})^{\xi}$ direction. The TCL-layer uses the elevation information along with passing time as an ephemeris function (green) which modulates its input weights to rotate the SOL-neuron output and provide a true north estimate—black arrows indicate the $(\phi_{\text{TCL}}^{\text{pref}})^{k}$ direction. The response across the TCL-neurons is in the form of a sine wave which can be decoded to determine an estimate of the azimuth (phase) and of the confidence of that estimate (amplitude). Note however the *central complex* (CX) path integration circuit can use the TCL-neuron activity directly, without this explicit decoding.

the changing sun position over time. We describe each of these steps in more detail below.

Sum-of-sinusoids

The basic computation of the model is a *sum-of-sinusoids*, where the input is the 60 POL-neuron responses and the output is the position of the sun represented in an 8-neuron population code, named the *Solar Layer* (SOL). Each POL-neuron is connected to all SOL-neurons with a sinusoidal weighting function that represents the difference between the azimuthal direction of that POL-neuron's receptor in the sensor array, $\{\phi^j \in \mathbb{R} \,|\, 0° \leq \phi^j < 360°\}$, and the preferred direction of the SOL-neuron $(\phi_{\text{SOL}}^{\text{pref}})^{\xi}$, where $\{\xi \in \mathbb{N}^* \,|\, \xi \leq n_{\text{SOL}}\}$ is the index of the ξ^{th} SOL-neuron. Thus each SOL-neuron sums a set of sinusoids with the same frequency, different phases (depending on ϕ^j), and different amplitudes (depending on the POL-neurons activity),

resulting in another sinusoid. Specifically, the response of the neurons of the solar layer is given by the equation below,

$$r^{\xi}_{\text{SOL}} = \sum_{j=1}^{n_{\text{POL}}} \underbrace{\frac{n_{\text{SOL}}}{n_{\text{POL}}} \cdot \sin\left[\alpha^j - \left(\phi^{\text{pref}}_{\text{SOL}}\right)^{\xi}\right]}_{W^{j,\xi}_{\text{SOL}}} \cdot r^j_{\text{POL}} \tag{D.2}$$

where $n_{\text{POL}} = 60$ and $n_{\text{SOL}} = 8$ are the number of POL- and SOL-neurons respectively, r^j_{POL} and r^{ξ}_{SOL} are their responses and $W^{j,\xi}_{\text{SOL}}$ are the weights of the synapses connecting them. The weight depends on the orientation of the primary axis of the j^{th} ommatidium, $\alpha^j = \phi^j - 90°$, the preference angle of the respective ξ^{th} SOL-neuron, $\left(\phi^{\text{pref}}_{\text{SOL}}\right)^{\xi} = \xi \cdot 360°/n_{\text{SOL}}$, and the number of POL- and SOL-neurons. The more POL-neurons we have in the system the lower the synaptic weight connecting them to the SOL-neurons (smaller contribution), but the more SOL-neurons we have the higher the synaptic weight has to be (bigger contribution) amplifying the signal. Without this scaling factor, the signal is very weak and sometimes vanishes.

Effectively, this equation activates each SOL neuron in proportion to the DOP opposite its preferred direction. As the DOP is maximum at the *cross-solar* point, the response of the SOL-neurons contains information about the predicted solar azimuth, and the confidence of this prediction. By 'cross-solar' we mean the point that is $90°$ away from the sun, in the solar-antisolar meridian, passing through the zenith point. For analysis, we can decode the population code using Fast Fourier Transform (FFT),

$$R = \sum_{\xi=1}^{n_{\text{SOL}}} r^{\xi}_{\text{SOL}} e^{-i360°(\xi-1)/n_{\text{SOL}}} \tag{D.3}$$

where $R \in \mathbb{C}$. The angle of this complex number gives the estimated solar azimuth, ϕ'_s, while the magnitude implies the confidence of this prediction, τ_s,

$$\phi'_s = \arg(\bar{R}), \qquad \tau_s = ||R|| \tag{D.4}$$

where $\arg(\cdot)$ is the argument of a complex number, which gives the direction of the vector represented by it, and $\bar{R}$ is the complex conjugate of R. The confidence, τ_s, is just a factor and has no unit but it can be used to calculate the uncertainty, $\sigma_s = \frac{4°}{\tau_s} - 2°$, which is in degrees (see Appendix D.2). Note that the insect does not need to extract the solar azimuth explicitly in this manner to be able to use the SOL encoding in further processing, but it may need to extract the confidence factor as discussed later (see Appendix D.1.4).

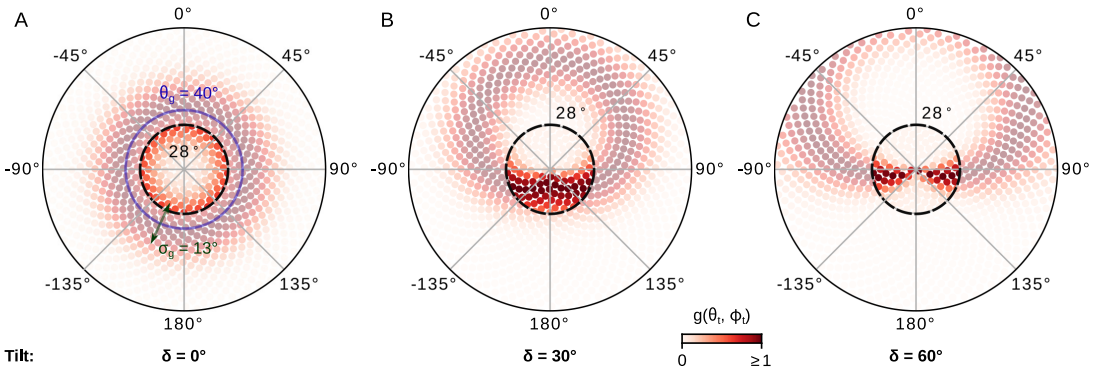

Figure D.5: The gating function that compensates for tilt. The differing weightings of ommatidia input under three levels of tilt—(A) $\delta = 0°$ ($\theta_t = 90°$), (B) $\delta = 30°$ ($\theta_t = 60°$), and (C) $\delta = 60°$ ($\theta_t = 30°$)—are shown, with darker shading indicating higher weighting. The inner dashed black circle delineates the actual receptive field of the simulated sensor ($28°$ radius, equivalent to $\omega = 56°$ receptive field). The extended array (greyed-out units) illustrates how this weighting adheres to a Gaussian function defined on the sky dome. The blue circle shows the dominant focus of the sensor ($\theta_g = 40°$), and the green arrow shows the smoothing parameter ($\sigma_g = 13°$).

Tilt compensation mechanism

The above computation of azimuth assumes the cyclopic DRA is aligned with the sky dome (its zenith point is always pointing towards the sky zenith and it is laterally aligned to the horizon). In nature, the head of the animals may not remain aligned with the horizon, particularly in walking animals such as ants, which do not fully stabilise their head position (Ardin, Mangan, et al., 2015). As the head deviates from the horizon, the predictions of the above model become less and less accurate. To compensate for this error because of tilting, we added a gating function that receives information about the sensor tilt and modulates the response of the solar layer (see Fig. D.4—orange connections).

Specifically, the gating function uses the known head-tilt angle (see Appendix D.3 for where this may come from for the animal) to preferentially select inputs from ommatidia facing towards the most interesting region in the sky, in the form of a Gaussian ring shape centred on the zenith point (see Fig. D.5):

$$g^j(\theta_t, \phi_t) = \exp\left\{ -\frac{1}{2}\left[\frac{\cos(\theta_g + \theta^j)\sin(\theta_t) - \sin(\theta_g + \theta^j)\cos(\theta_t)\cos(\phi_t - \phi^j)}{\sigma_g}\right]^2 \right\} \tag{D.5}$$

where θ^j and ϕ^j are the celestial coordinates (elevation and azimuth) of the relative position of the respective ommatidium, and (θ_t, ϕ_t) the celestial coordinates of the titling point. We notate $\delta = 90° - \theta_t$ the tilting polar angle, which is the tilting angle

with respect to the pole (zenith), such that zero tilt corresponds to the intuitively logical position of the sensor being centred on the zenith. The radius of this ring, $\theta_g = 40°$, denotes the dominant focus direction of the system (where the peak of the Gaussian bump should be placed). The width of the ring (variance), $\sigma_g = 13°$, stands for the soft elevational receptive field for the SOL-neurons (smoothing parameter). The above parameter values are the result of a global optimisation procedure (see Appendix D.2).

The sum-of-sinusoids from the previous section then changes to:

$$r^{\xi}_{\mathrm{SOL}} = \sum_{j=1}^{n_{\mathrm{POL}}} \underbrace{\frac{n_{\mathrm{SOL}}}{n_{\mathrm{POL}}} \cdot \sin\left[\alpha^j - \left(\phi^{\mathrm{pref}}_{\mathrm{SOL}}\right)^{\xi}\right]}_{W^{j,\xi}_{\mathrm{SOL}}} \cdot g^j(\theta_t, \phi_t) \cdot r^j_{\mathrm{POL}} \tag{D.6}$$

Time compensation mechanism

We extend the above model to provide a method by which insects might compensate for the changing celestial sky pattern over the course of the day and seasons. In particular, the aim is to provide a stable geocentric compass even though the solar azimuth varies compared to the true north. This is usually assumed to require an internal clock and calculation or learning of the ephemeris function (Wehner and Müller, 1993; Towne, 2008). Here we suggest a possible solution that uses the current polarisation information only to estimate the ephemeris function and thus continuously adjust the compass. This is based on the observation that the confidence of the estimate of the solar azimuth is related to the solar elevation (see Appendix D.3), from which we can infer the rate of change of the azimuth.

We add an extra layer to our model, the *true compass layer* (TCL), which is a copy of the SOL but the preference angle of its neurons changes through time (see Fig. D.4). The basic response of this layer is given by:

$$r^{k}_{\mathrm{TCL}} = \sum_{\xi=1}^{n_{\mathrm{SOL}}} \underbrace{\frac{n_{\mathrm{TCL}}}{n_{\mathrm{SOL}}} \cos\left[\left(\phi^{\mathrm{pref}}_{\mathrm{TCL}}\right)^{k} - \left(\phi^{\mathrm{pref}}_{\mathrm{SOL}}\right)^{\xi}\right]}_{W^{\xi,k}_{\mathrm{TCL}}} \cdot r^{\xi}_{\mathrm{SOL}} \tag{D.7}$$

where $\left(\phi^{\mathrm{pref}}_{\mathrm{TCL}}\right)^{k} = k \cdot 360°/n_{\mathrm{TCL}}$ is the preference angle of the k^{th} TCL-neuron and $n_{\mathrm{TCL}} = 8$ is the size of TCL-population.

As described earlier, we can derive from this neural representation the direction of the sun, ϕ'_s, and the confidence of the prediction, τ_s. In a low-disturbance environment

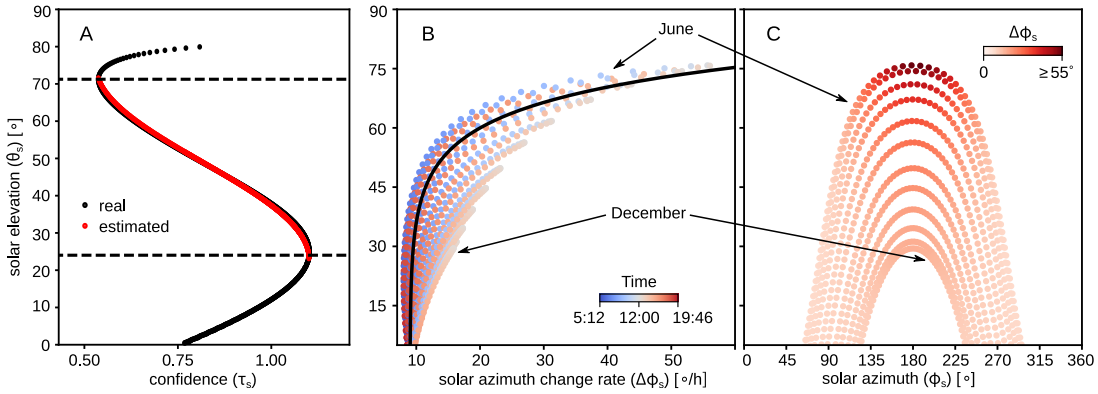

Figure D.6: Using confidence of the estimate to compensate for time. (A) The confidence value of the compass response varies with the solar elevation (black dots). Within the range $23° - 72°$ this relationship can be used to estimate the elevation using Eq. (D.8) (red dots). (B) The rate of change of the solar azimuth over time depends on the elevation [coloured dots represent different times of the day from morning (blue) to the evening (red)] and can be approximated using Eq. (D.9) (black line). (C) Showing the solar elevation with respect to the solar azimuth for different times of the year. Each of the 12 imaginary curves in B and C correspond to the 21st of every month; the sampling rate on each day is every 10 minutes from sunrise to sunset.

and for a specific set of responses, the confidence can be transformed to a prediction of the solar elevation using the function below:

$$\theta'_s = 75° + 26° \cdot \left[\frac{\sin^{-1}(2.855 - 3.5\tau_s)}{90°} - 1 \right] \tag{D.8}$$

with domain $\{\tau_s \in \mathbb{R} \,|\, 0.53 \leq \tau_s \leq 1.1\}$ and range $\{\theta'_s \in \mathbb{R} \,|\, 23° < \theta'_s < 72°\}$ (see Fig. D.6A). For values of τ_s outside its domain, this equation would give the respective highest or lowest possible θ'_s. The above equation has been derived heuristically and it is a very simplified approximation of the inverse of the *equation of time* or *astronomical equation* Scargle, 1982. When the confidence is high enough the predicted elevation, θ'_s, can be used to approximate the change rate, in degrees per hour (°/h), of the solar azimuth using the following equation (illustrated in Fig. D.6B):

$$\frac{\mathrm{d}\phi_s}{\mathrm{d}t} = \exp\left(\frac{\theta_s - 36°}{10°}\right) + 9°/\mathrm{h} \tag{D.9}$$

Integrating the above equation through time and using the information that the sun is always moving clockwise, we can approximate the shift of our predicted solar azimuth, ϕ'_s, required to get a global orientation (see Fig. D.6C). This is implemented by introducing an update rule for the TCL-layer:

$$\left(\phi_{\mathrm{TCL}}^{\mathrm{pref}}\right)^k \leftarrow \left(\phi_{\mathrm{TCL}}^{\mathrm{pref}}\right)^k + \frac{\mathrm{d}\phi_s}{\mathrm{d}t} \tag{D.10}$$

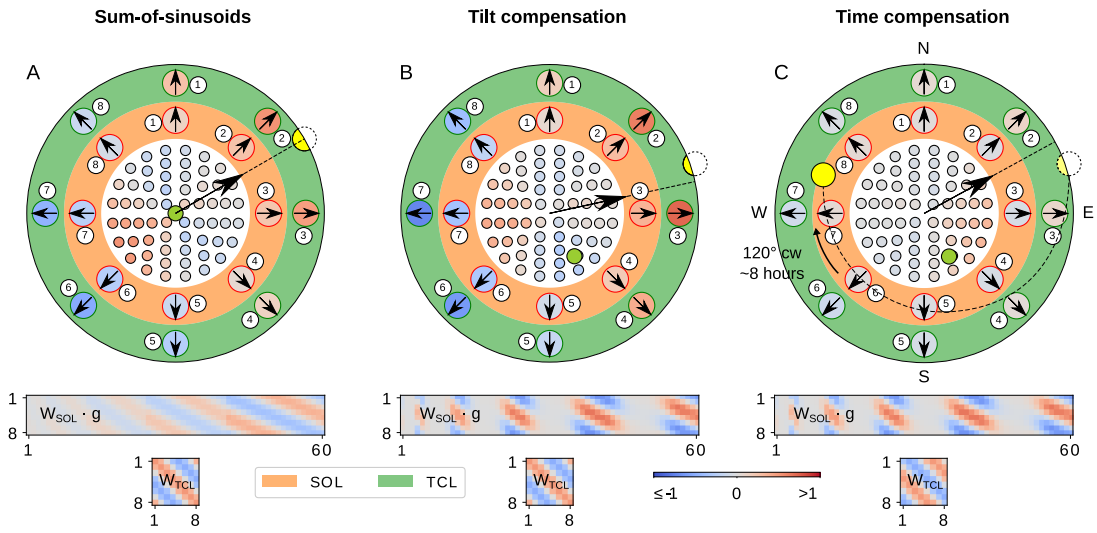

Figure D.7: Step-by-step processing of the compass model. The white, orange and green areas show the response of the POL-, SOL- and TCL-neurons respectively; red denotes excitation and blue suppression (see the colour bar for values). The set of synaptic weights connecting each layer is shown under the disc (values based on the same colour bar). Orange background means that the activity is affected by the gating function, and green that it is affected by the time compensation mechanism. In the white disc, different points are the relative positions of the POL units on the sensor; the numbering starts from the centre and unwraps clockwise towards the outline like a spiral; the round green mark is the point of the sensor that is aligned with the zenith of the sky, and the yellow circle is the sun position. The black arrow with the dashed line is the decoded prediction of the solar azimuth from the TCL neurons; the numbering refers to the identities of the neurons in the weight matrices below. The weight matrices show the synaptic weight between consecutive layers [defined by Eq. (D.6) and Eq. (D.7) for the SOL and TCL respectively; values based on the same colour bar]; the horizontal is the input and the vertical the output axis. (A) The sum-of-sinusoids mechanism detects the solar azimuth; the zenith (green) point is aligned with the sensor orientation; the solar azimuth is encoded in both the SOL- and TCL-layers and the activation code (phase) of the two layers look identical, as the time compensation mechanism has been deactivated. (B) The tilt compensation mechanism corrects the predicted solar azimuth using tilting information; the sensor has been tilted 30° NNW (so now the zenith point is 30° SSE); the gating function has changed the focus on the specific ommatidia, as shown in the $\mathbf{W}_{\mathrm{SOL}} \cdot \mathbf{g}$ matrix. (C) The time compensation mechanism corrects for the solar azimuth changes using the solar elevation; 8 hours have passed so the sun has moved 120° clockwise but the compass is still aligned to the same direction due to the updated $\mathbf{W}_{\mathrm{TCL}}$ weights (see also the difference in SOL and TCL responses compare to previous steps).

which updates the connections between the solar layer and the TCL-neurons through a recurrent connection (see Fig. D.4—green connections). Fig. D.7 shows the response and weights for all the different layers along with the corrections of the global direction when we add the tilt and time compensation mechanisms. The above integration gives an alternative answer to the question of how insects can develop their *ephemeris function* without using a clock.

D.1.5 *Central complex*

The TCL-neuron output described above provides the required compass input that was assumed to be available in a previous anatomically constrained model of path integration in the insect CX (Stone, Webb, et al., 2017). In that model, a set of 8 TB1 neurons with preferred directions $\{k \cdot 45^\circ \,|\, k = 1, ..., 8\}$ were activated with a sinusoidal relationship to the heading of the agent. We thus use an exact copy of this CX model, replacing its idealised input with our polarisation-derived compass signal to test its efficacy and robustness.

D.1.6 *Evaluation*

Measuring compass accuracy

In order to evaluate the performance of the model, we introduce an objective function, J, which measures the average error across multiple predictions of the solar azimuth. Each prediction is made for a different sun position and tilting orientation. More specifically, we tested 17 tilting orientations, namely 8 homogeneously distributed on a 60° tilted ring, 8 on a 30° ring and one pointing towards the zenith. For each of those tilting orientations, we sample 500 homogeneously distributed sun positions on the sky dome, giving a total of 8500 predictions. The error is given by the following equation,

$$J = \frac{1}{8500} \sum_{t=1}^{17} \sum_{s=1}^{500} ||[(\phi_{t,s} - \phi'_{t,s} + 180^\circ) \mod 360^\circ] - 180^\circ|| \tag{D.11}$$

where $\phi_{t,s}$ and $\phi'_{t,s}$ are the real and estimated solar azimuths.

We report the error as the mean absolute angular error plus/minus the standard error (MAE $\pm$ SE). We also report the confidence of the predictions, τ_s, as a value that shows how much we should rely on the respective predictions. Fig. D.8A gives a schematic representation of the objective function.

The same objective function was used to explore the parameter space in both the design and the computational model of the sensor. The parameters for the layout of our sensor along with those of the computational model are inspired by biological features in insects, and mainly the *Cataglyphis* desert ants (see Table D.1). Nevertheless, we are interested in exploring different set-ups that may perform better than the ones that biology indicates.

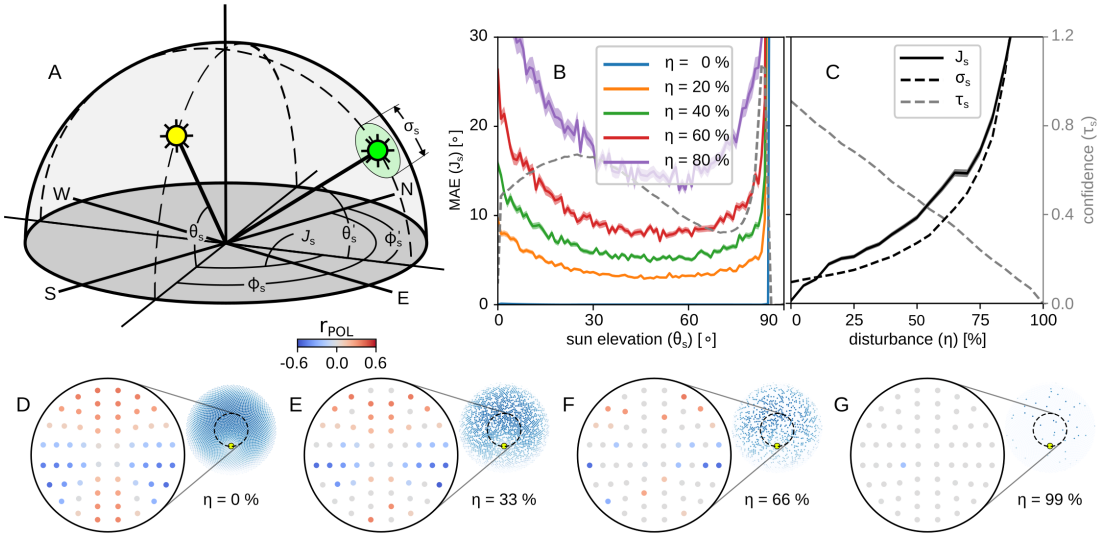

Figure D.8: The objective function and the accuracy of the compass. (A) Schematic representation of the objective function; the yellow and green suns illustrate the real and estimated sun position; the disk around the green sun denotes the uncertainty of the estimation, $\sigma_s = \frac{4^\circ}{\tau_s} - 2^\circ$. (B) Mean absolute angular error (coloured solid lines), MAE $\pm$ SE, and confidence (black dashed line), τ_s, against the solar elevation for different disturbance levels, when the sensor points towards the zenith. (C) The mean absolute angular error (black solid line), uncertainty (σ_s; black dashed line) and confidence (τ_s; grey dashed line) against different disturbance levels. On the bottom there are some examples of the responses of the POL units in (D) a non-disturbed condition; (E) with $\eta = 33\,\%$ disturbance; (F) with $\eta = 66\,\%$ disturbance; and (G) with $\eta = 99\,\%$ disturbance. The insets show a sample of the sky that caused the responses and the yellow mark on it shows the position of the sun.

Sensory input disturbance

By adding perturbations of the polarisation signal in the simulation, we can evaluate the robustness of the sensor in noisy environments. In a natural environment, this could be caused by a sensor malfunction, clouds, vegetation or other objects that can block the light from the sky or destroy the polarisation pattern (note however that in the UV spectrum, clouds have a relatively small effect on light propagation). In practice, a disturbance level, $\{\eta \in \mathbb{R} \,|\, 0\% \leq \eta \leq 100\%\}$, specifies the percentage of ommatidia that fail to contribute to the input. These are uniformly distributed across the surface of the eye. We tried different types of disturbance (including Gaussian noise) but found the type of disturbance shown here to be the most informative.

Behavioural simulations

The aim of our model was to fill in the missing link between the compass information available from the sky and the path integration behaviour of insects. We, therefore, assess the performance using a simulated ant in a simulated world using the accurate

sky model. The agent takes a predefined foraging path sampled from real ant-data (Mangan and Webb, 2012), which is translated into a sequence of directions and distances that the agent follows for its outward journey. It uses the insect eye and visual processing model described above as input to the CX path integration model (Stone, Webb, et al., 2017), and the consequent CX output to control its inward journey. The ability of the agent to return home by the direct path is assessed using the same methods as in Stone, Webb, et al. (2017) allowing direct comparison between the use of simulated input and our network. We also test with conditions of additional sensor disturbance and head tilt caused by uneven terrain.

Simulating neurophysiological experiments

As outlined in the introduction, neurophysiological recordings from the PB of the desert locust (Heinze and Homberg, 2007) appear to show a compass output that spans only $[0°, 180°]$. To see if we can account for this result, which is discrepant with our model, we perform a simulated neurophysiology study in which we record the response of the TCL-neurons in our model using the same stimulus conditions as the animal experiments. Thus, we expose the artificial DRA to a uniform light source filtered by a rotating linear polariser and construct tuning preference curves for the TCL-neurons.

D.2 RESULTS

D.2.1 *Compass accuracy without tilt*

We first evaluate our compass model in conditions where it always points towards the zenith. The average error in the absence of disturbance is $J = 0.28° \pm 0.1620°$ for $N = 1000$ sun positions homogeneously distributed on the sky-dome, and the average confidence was $\tau_s = 0.91$, which is quite high (see Fig. D.8B and C to compare with regular confidence levels).

The solar elevation strongly affects the polarisation pattern in the sky and as a result the accuracy of the compass predictions. Fig. D.8B reports the error measured for different solar elevations and different levels of disturbance. Note we measure the elevation as the angular distance from the horizon (thus the zenith corresponds to an elevation of $90°$). The blue line, which is hardly visible and lying at the bottom of the figure, is the error for samples without disturbance. The dashed line is the average confidence reported across all the reported disturbance levels. It is not hard to notice that, apart from the error, confidence is also affected by solar elevation.

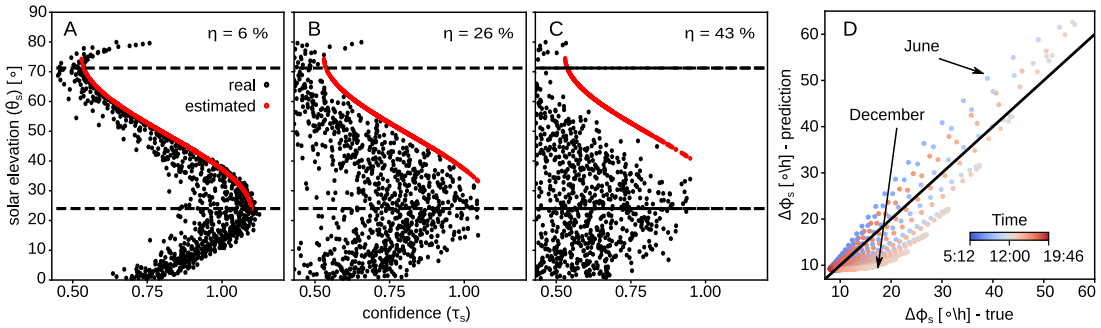

Figure D.9: Dealing with time and light disturbance. Transformation of the compass response to solar elevation, and to the derivative of the solar azimuth function. (A - C) The function of the elevation with respect to the response with disturbance $\eta = 6\%$, $\eta = 26\%$ and $\eta = 43\%$; the red dots are the predicted solar elevation given the compass response, and the black dots denote the real values; the dashed lines give the range of the function. (D) The function of the real derivative of the solar azimuth against its prediction using Eq. (D.9); black line shows a perfect match of the two derivatives; colour denotes the time: blue is for morning and red is for the evening.

In the model, we use this latter effect to our advantage, to estimate the elevation from the confidence. Fig. D.9A-C show how the different disturbance levels affect the estimation of the solar elevation, θ'_s (black dots—real, red dots—predicted), using the confidence, τ_s. Fig. D.9D shows the predicted against the real solar azimuth change rate.

Fig. D.8C gives a summary of the effect of the disturbance on our model's predictions. We notice that as the disturbance grows so does the error of the predictions, but the confidence drops. This suggests we should not trust the predictions of our model when the disturbance of the sensory input is more than 85%, but for lower disturbance levels the compass still gives predictions with less than $30°$ error, which can be sufficient for the path integration task (see later results).

D.2.2 *Effects of head tilt*

Navigating ants are subject to large changes in their head pitch angle, particularly when carrying objects such as food or nest mates Ardin, Mangan, et al., 2015. Here we assess how this might impact the accuracy of their celestial compass.

As we described in the methods, we filter the connections between the POL- and SOL-neurons using a gating function. With this function deactivated, and thus all ommatidia providing input to the solar layer, the performance of the model drops significantly. Fig. D.10A-C demonstrate the increased error of the predictions for different sun positions as the sensor is tilted for $\delta \approx 0°$, $30°$ and $60°$ respectively, and Table D.2 summarises the respective average error along with the overall error.

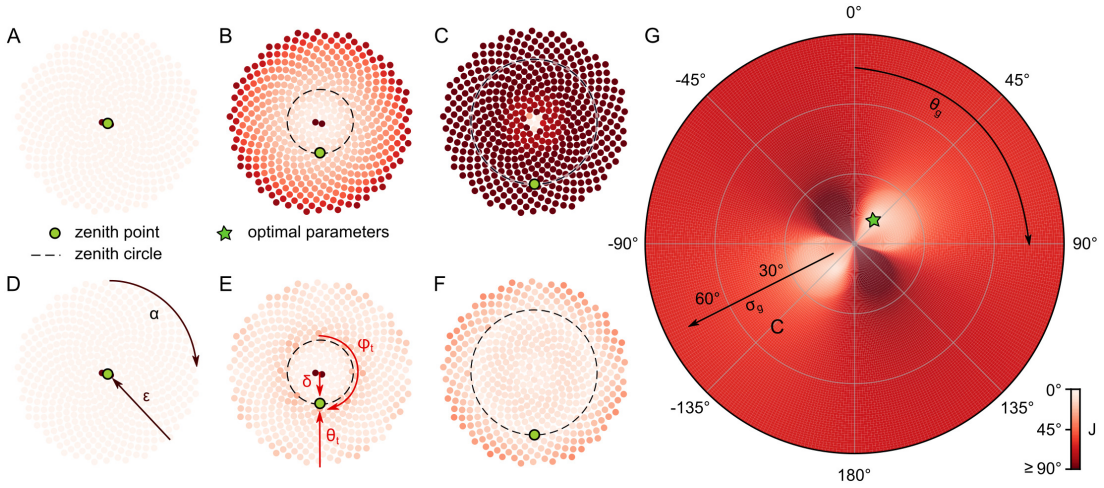

Figure D.10: Dealing with tilt for a variety of gating parameters. The angular error of the expected direction of the sensor for different tilted angles and gating parameters. Black arrows show the axes; red shading shows the value of the objective function (J)—darker shading is for higher error values; red shaded points show the values of J for different sun positions; green discs show the zenith angle, θ_t; black dashed lines visualise this angle for any tilt direction, ϕ_t (red arrows). (A-F) Visualisation of the azimuthal angular error with respect to the sun position—$\epsilon \in [0, 90]$ and $\alpha \in [0, 360)$—for three tilting angles of the sensor: (A, D) $\delta \approx 0°$ ($\theta_t = 90°$), (B, E) $\delta \approx 30°$ ($\theta_t = 60°$), (C, F) $\delta \approx 60°$ ($\theta_t = 30°$); without—top row (A-C)—and with gating—bottom row (D-F). (G) Average angular error for different gating parameters. The lowest cost (green star; $J = 10.47° \pm 0.12°$, $N = 8,500$) is for ring radius $\theta_g = 40°$ and width (variance) $\sigma_g = 13°$.

	before gating			after gating			N
$J_{\delta=0°}$	$0.63°$	$\pm$	$0.01°$	$0.47°$	$\pm$	$0.01°$	500
$J_{\delta=30°}$	$35.68°$	$\pm$	$0.01°$	$9.53°$	$\pm$	$< 0.01°$	$4,000$
$J_{\delta=60°}$	$104.01°$	$\pm$	$0.01°$	$13.16°$	$\pm$	$< 0.01°$	$4,000$
J	$65.78°$	$\pm$	$0.62°$	$10.47°$	$\pm$	$0.12°$	$8,500$

Table D.2: Mean absolute error before and after using the gating function. N, number of samples used for this error; $J_{\delta=0°}$, MAE $\pm$ SE for $\delta = 0°$ tilting; $J_{\delta=30°}$, MAE $\pm$ SE for $\delta = 30°$ tilting; $J_{\delta=60°}$, MAE $\pm$ SE for $\delta = 60°$ tilting; J, overall MAE $\pm$ SE; MAE, mean average azimuthal error; SE, standard error.

Activating the gating function, the influence of each ommatidium to the responses of the solar layer becomes a function of the tilting parameters, producing much more robust results (Fig. D.10D-F). More specifically, the overall average error drops to $J = 10.47° \pm 0.12°$ ($N = 8,500$).

The default parameters for the gating function were selected using exhaustive global optimisation. More specifically, we fixed the design and network parameters and explore the different combinations of the parameters, θ_g and σ_g, in the gating function. As the number of parameters in the function is very small and their range is also constrained, exhaustive global optimisation was not a very costly process. Fig. D.10G

illustrates that the current combination of sensor layout and compass model perform best for a gating function with a ring shape of $\theta_g = 40°$ radius and $\sigma_g = 13°$ thickness.

The radius of the ring could be interpreted as the dominant angle of focus or the most informative direction. Moreover, our optimal main focus angle is $40°$ from the zenith point differs to the $25°$ angle, suggested in Labhart, 1999.

D.2.3 *Exploration of the structural parameters*

The sensor layout and the number of neurons in the computational model were based on biological data, but it is of interest to examine the effect of varying these parameters. The performance error of the compass (including tilted conditions) for a range of layout parameters is illustrated in Fig. D.11A. The green line on this figure indicates the receptive field with the lowest error for the given number of units. The performance with respect to the receptive field seems to be independent of the number of units used. More specifically, the best performance on average was for $\omega = 55.99° \pm 0.33°$.

The error observed on the slice set by the green line in Fig. D.11A, where $\omega \approx 56°$, is illustrated in Fig. D.11B. This figure shows that there is a sharp drop of the error up to $n = 60$ units, after which there is not a significant improvement. A slice on the other axis, for $n = 60$, is illustrated in Fig. D.11C, which shows that the best design parameters for the sensor are $\omega \approx 56°$ receptive field and $n = 60$ number of units. The average error reported for these parameters is $J = 10.47° \pm 0.12°$. However, the lowest error reported was $J = 9.55° \pm 0.12°$ for $n = 336$ and $\omega = 56°$.

The number of SOL- and TCL-neurons were selected based on the CX model described in Stone, Webb, et al., 2017. However, we explored different populations of neurons and compare the performance to the one of the proposed model. Our results showed that less than 8 SOL interneurons increase the error to $J = 47.49° \pm 0.32°$, while more interneurons do not change the performance. Similarly, as long as we have at least 4 TCL-neurons, the performance of the sensor does not change for any number of SOL-neurons.

D.2.4 *Path integration*

To demonstrate the performance of the sensor in a more realistic scenario, we integrated the compass and path integration (Stone, Webb, et al., 2017) models, testing how the compass accuracy affects the foraging and homing paths. We create an environment with a simulated sky and let an agent navigate in it. We guide our agent to the food source, using 133 different routes from Spanish desert ants *Cataglyphis velox*

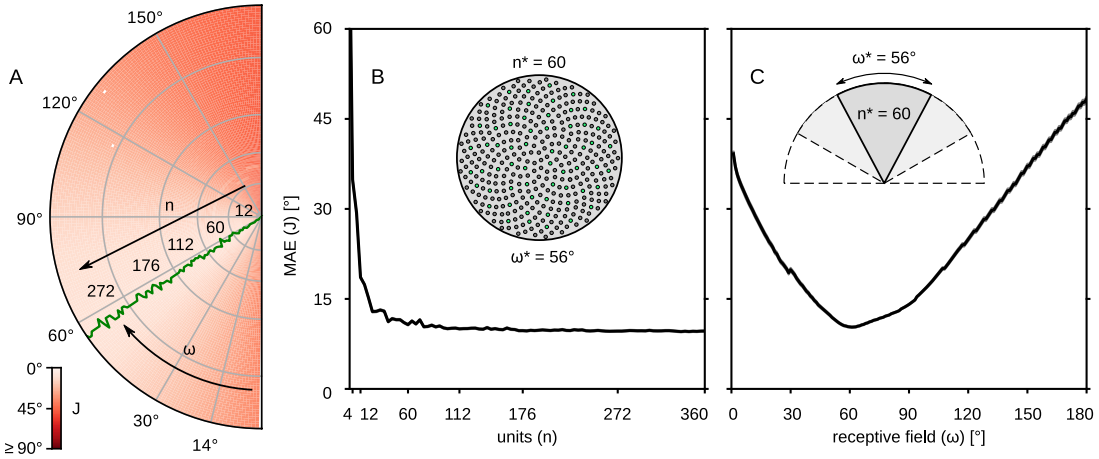

Figure D.11: Optimal compass structural parameters. The performance of the compass for different topological parameters. (A) Values of the objective function on the $\omega \times n$ plane; red shades illustrate the degree of error; black arrows show the axes; the green line shows the receptive field value associated with the minimum error for different numbers of units. (B) The error as a function of the number of units, n; the receptive field is fixed at $\omega = 56°$; inset demonstrates the $56°$ wide sensor with $n = 360$ units; with green are marked the 60 units closest to the ones we chose. (C) The error as a function of the receptive field, ω; the resolution (ratio between ω and n) is fixed so that the number of units is $n = 60$ for $\omega = 56°$; inset demonstrates the different visual fields including the optimal one with $n = 60$ units.

(Mangan and Webb, 2012) and let the agent return to the nest using its path integrator. In addition, we test the performance of the agent under different sky conditions, by adding disturbance to the polarisation pattern.

Fig. D.12 summarises the results of the above experiment. The faded coloured lines in Fig. D.12A (even terrain) and Fig. D.12B (uneven terrain, illustrated in Fig. D.12C) are the outward paths and the bold lines are the inward paths. The colour of the line identifies the different disturbance levels. We use similar evaluation methods to Stone, Webb, et al., 2017 to allow direct comparison. Fig. D.12D and E show the overall performance of the agent in the path integration task with respect to the tortuosity of the inward route, $\tau = \frac{L}{C}$, where L is the distance from the nest and C is the distance travelled.

The results show that in most cases, the agent is moving in the correct direction until it reaches the nest and then does a systematic search for the nest. An exception is for $\eta \approx 97\%$ (see Fig. E.29), where the agent continues moving in the same direction not realising how far it has travelled. Overall, for less than $\eta \leq 90\%$ disturbance the agent seems to navigate without noticeable problems. However, for higher disturbance levels we see a drop in the performance of the navigation task, walking at least twice the distance of the nest from the feeder for $\eta \approx 97\%$. The performance of the agent is affected very little by the uneven terrain, which shows that the tilting of the

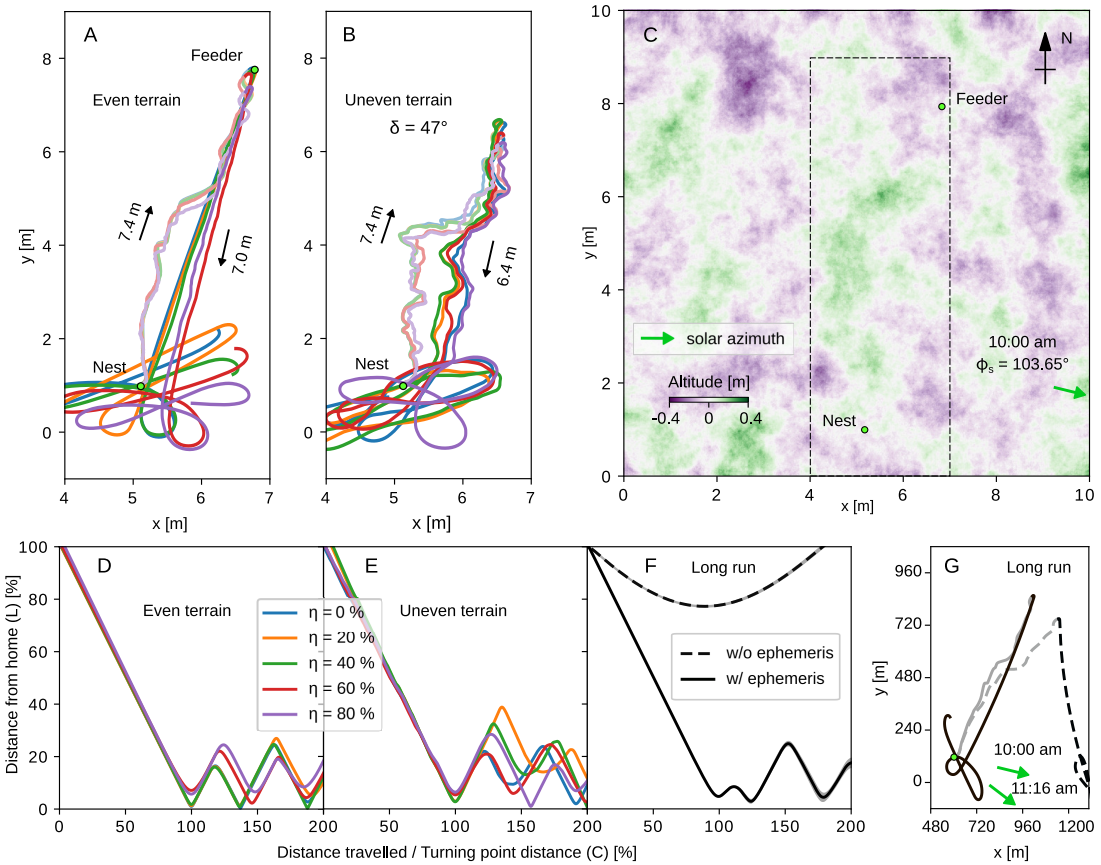

Figure D.12: Behavioural simulation for the path integration task. Testing the celestial compass on path integration tasks. We set up the experiments to take place at 10 am in Seville, Spain ($37°23'33.03''$N, $5°53'01.95''$W). The altitude variance is 0.8 m and the maximum tilting angle noticed in all the experiments is $\delta = 47°$. (A) Five representative routes of ants in different sky disturbance levels for an even—(B) uneven—terrain and their respective inward paths; different colours are for different disturbance levels (see legend); the faded lines are the outward paths and the bold ones are inward. (C) The uneven terrain map; the green colour denotes hills and purple valleys; the marked region is the one cropped for the A and B plots. (D) Deviation from the best possible route during homing for different disturbance levels for even—(E) uneven—terrain. We scale up our experimental arena (by a factor of 120) to enable longer runs that demonstrate the performance of the time compensation mechanism. (F) Comparison of the path integration performance in terms of tortuosity with (solid black line) and without using the time compensation mechanism (dashed line). (G) The actual paths generated by the above experiment; green arrows show the direction of the sun at the beginning (10:00 am, $103.65°$ clockwise from north) and end of the route (11:16 am, $127.47°$ clockwise from north).

agent is not a problem anymore. Fig. E.28 summarises the results for different input disturbance levels and steepness of the terrain. Fig. E.29 illustrates the corresponding detailed paths of all the agents.

The terrain used here is drawn from a normal distribution, allowing the agent to tilt for a maximum of $\delta = 47°$. The outward paths are consequently distorted by compass and distance errors while following the predefined sequence of directions

and distances, but the homing paths still successfully guide the agent back to the nest, suggesting that any systematic bias in compass or distance information caused by uneven terrain is balanced out between the outward and inward routes. However, it is clear that the uneven terrain introduces an extra level of moment-by-moment disturbance in the heading direction.

In addition, we tested the performance of the sensor in longer runs, which take more time and hence will test the operation of the sensor's time compensation mechanism (Fig. D.12F and G). We multiply the dimensions of the arena and the outward paths of the ants by a factor of 100, transforming the arena to 1 km $\times$ 1 km and the total run of the agent to 1 hour and 16 minutes. In this time (from 10 : 00 am to 11 : 16 am) the sun's position changes by 23.82° clockwise. Fig. D.12F and G show that including the time compensation mechanism the agent successfully returns to the nest, while without it the path integration mechanism leads it away from the nest due to the change of the sun position. For detailed paths of multiple ant-routes see Fig. E.30.

D.2.5 *Experimental paradigm*

We have noticed that the output of our compass model, the TCL-neurons, is not identical to the electrophysiological responses of the locusts' TB1-neurons reported in Heinze and Homberg (2007). However, this is not surprising, as the testing conditions of the two experiments were very different. More specifically, in the locust experiment, the animal was pinned in a vertical position and its DRA was exposed to uniform light passing through a rotating polariser. On the contrary, we expose our sensor to realistic sky-light facing upwards, assuming that the head of the hypothetical animal is aligned with the horizon. Therefore, we tried to simulate the former experimental environment and compare the responses of our TCL-neurons to the TB1-neurons recorded from the desert locust.

We found that the response of the simulated compass neurons closely resembles the double preference angles recorded in locust TB1-neurons (Heinze and Homberg, 2007) when stimulated by a rotating linear polariser under a uniform light source (Fig. D.13B and D). This contrasts dramatically with the response of the same simulated neurons when exposed to the natural skylight pattern (see Fig. D.13F). Moreover, calculating the preferred directions from the response of the simulated neurons under the linear polariser, for each column (Fig. D.13C), produces results rather comparable to the locust (Fig. D.13A). Note that for the linear polariser, the preferred direction has an inherent 180-degree ambiguity: for our simulation data (Fig. D.13C) we resolve this by taking the stronger of the two peaks (effectively a random choice as this difference results from noise); for the locust data (Fig. D.13A) we use the di-

rection chosen in the original paper. These data have been interpreted in Heinze and Homberg (2007) as supporting a $[0, \sim 180°]$ 'map' of polarisation directions across the PB, increasing by $\sim 22.5°$ per column (see fitted line, Fig. D.13A). Our results suggest this effect may be a consequence of the experimental procedure rather than revealing the true directional preference—relative to the sky pattern—of these neurons, which may instead resemble Fig. D.13E.

The responses of Fig. D.13A/B and Fig. D.13C/D are similar but not identical. For example, despite the similar range of their expected, $\overline{\phi}_{\max}$, and undisturbed preference angles, $\phi^*_{\max}$, in Fig. D.13A and C respectively ($[90°, 270°]$), their exact values are not identical. Note that the expected values try to approximate the undisturbed values of the real recordings. However, the number of recordings from the locust brain is limited and shows substantial variability (see the full set of simulation and locust responses in Fig. E.31). As a consequence, it does not seem productive to attempt to quantitatively replicate the details of this activity pattern with our model but rather would be more interesting to test the response of these neurons to a more realistic sky pattern, which we predict should have a substantial qualitative effect on the observed activation patterns.

D.3 DISCUSSION

To perform path integration, insects need to transform polarised skylight to a global orientation. We have proposed a mechanism to explain how this might occur in the insect brain, given known anatomical constraints of the OL and the PB. Our OL model contains a large number (~ 60) of polarisation-opponent neurons that respond to the degree and direction of polarisation received from each ommatidium. The PB should (to match previous work on path integration) contain exactly 8 compass neurons with sinusoidal tuning to 8 cardinal directions, ideally able to compensate for tilt and time. We show that this can be obtained by a weighting function that takes into account the fact that each ommatidium points at a specific region of the sky, and that the sky polarisation pattern has a specific relationship to the sun's position. Most previous models have either assumed a simple linear polarisation as input (Sakura, Lambrinos, and Labhart, 2008) or, if physically implemented and tested under the real skylight, have used a limited number of receptors (for a review see, Karman, Diah, and Gebeshuber, 2012). However, neurophysiological investigation of the receptive fields of CX neurons suggests they could act as matched filters to the specific pattern of polarised light in natural skylight (Bech, Homberg, and Pfeiffer, 2014). By combining a realistic sky model and receptor layout we show it is possible to obtain a good estimate of heading direction, sufficient for accurate path integration, robust to

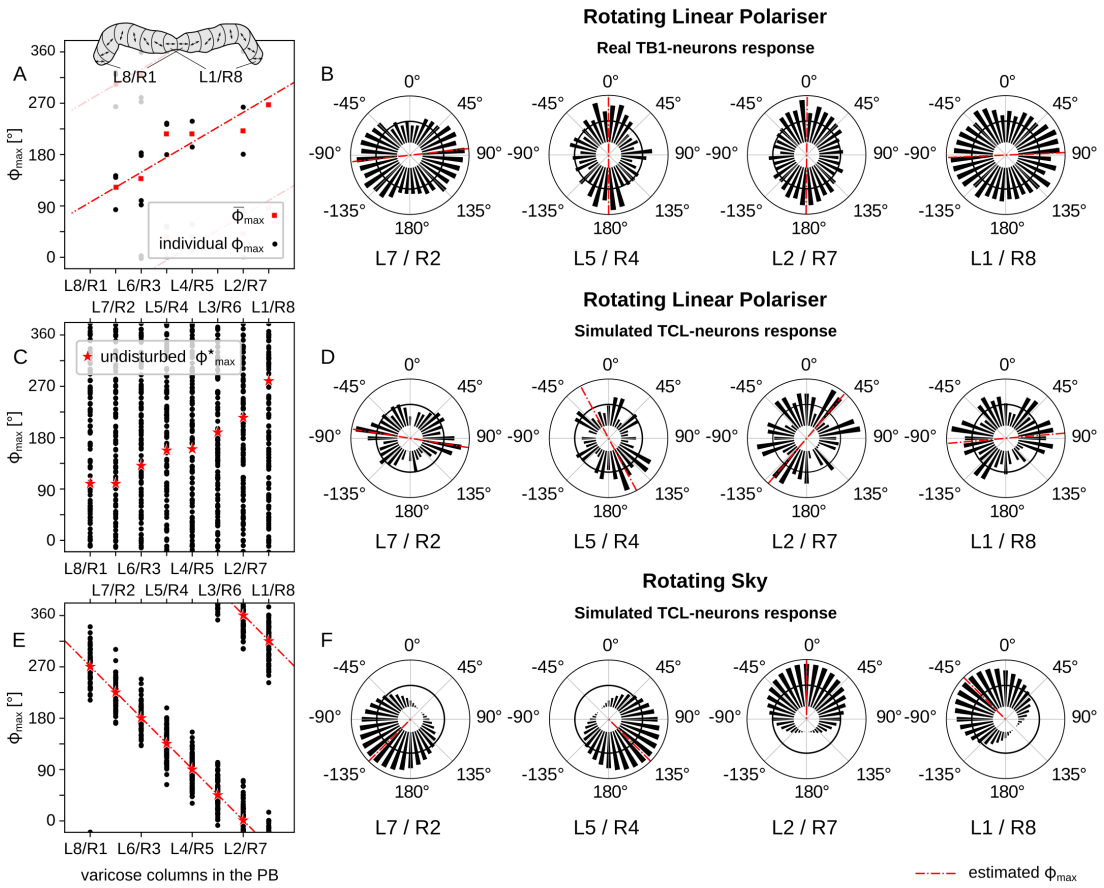

Figure D.13: Real and simulated response of compass neurons for artificial and natural polarised light. E-vector orientation resulting in maximum excitation, $\phi_{\max}$, for TB1-neurons in different columns of the *central complex* (CX) (A, B) and TCL neurons in our simulation, determined by circular statistics (Rayleigh test, Batschelet, 1981): (A) Real data of TB1-neurons ($N = 15$) in the *protocerebral bridge* (PB) of the locust brain, reconstructed from original data supplied by Stanley Heinze after Heinze and Homberg (2007) (inset shows the naming of the columns in the PB), (C) simulated TCL-neurons under a rotating linear polariser, and (E) under a rotating natural sky ($N = 100$, $\eta = 50\%$, $\theta_s = 30°$); black dots show individual samples, red squares show the mean $\phi_{\max}$ and red stars show the $\phi_{\max}$ in a condition without external disturbance ($\eta = 0\%$); red lines are best linear fit to the data points. (B, D, F) show corresponding examples for specific neurons from four of the CX columns; bin width $10°$; black solid lines indicate background activity; red lines indicate the $\phi_{\max}$ direction.

disturbances, able to compensate for a head tilt that might occur with rough terrain, and to adapt to sun movement, providing a powerful celestial compass sensor based on the polarisation pattern.

D.3.1 *Obtaining solar azimuth from polarisation information*

In our model, each neuron in the SOL layer integrates the signal from all sensory units. The relative azimuth of the sensor unit in the array to the preferred direction

of the compass unit determines its weighting. As a result, the response across the compass units effectively represents the direction in the sky with the highest DOP, the cross-solar azimuth, from which the solar azimuth can be directly inferred. Our model thus counters the common assumption that a polarisation-based compass sensor must inherently have a $180°$ degree ambiguity and requires some additional signal to resolve the $360°$ directionality (for example, Lambrinos, Möller, et al., 2000; El Jundi et al., 2014). Another consequence is that the best compass performance is not when the sun is on the horizon (thus producing the maximum DOP, largely in one direction, in the zenith) as has been sometimes assumed. In fact, for the sun exactly on the horizon or exactly on the zenith, precise symmetry in the resulting polarisation pattern will result in ambiguity and low confidence in our model. The highest confidence occurs when the cross-solar azimuth falls within the receptive field of the sensor (modulated by the gating function), which corresponds to a solar elevation $\theta_s \approx 28°$ (nearer to the horizon, but not on it).

Note however that for higher elevations, specifically, for $60° \leq \theta_s \leq 90°$ the sun itself would fall in the receptive field of the sensor. This suggests a parallel processing system based on sky luminance could form a complementary mechanism for determining the solar azimuth that would be accurate for solar elevations where the polarisation one is not. A speculative pathway for this could originate in two out of the eight photo-receptors in the ommatidia of the desert ants that are sensitive in a wider range of the spectrum Labhart, 2016 (see Fig. D.3C), which could detect a sufficient light intensity gradient, and (in a similar way to our POL to SOL processing) form a novel skylight intensity compass. Alternatively or in parallel, the position of the sun is likely to also be detected by the non-DRA ommatidia of the compound eye, which are much better equipped for this task due to their smaller acceptance angle (Labhart, 1986). The two pathways could then be combined in the CX to form a complete insect celestial compass, consistent with the observation Wystrach, Schwarz, et al., 2014; Labhart, 2016 that the insect's compass appears to integrate multiple modalities of light (Jundi et al., 2015; Pfeiffer and Homberg, 2007). Specifically, recent neurophysiological results show that all polarisation-sensitive neuron types in the CX also show azimuth-dependent responses to an unpolarised UV or green light spot (El Jundi et al., 2014; Pegel, Pfeiffer, and Homberg, 2018).

D.3.2 *Neurobiological plausibility*

Our model represents a proposed mapping from POL to TB1 neurons in a computational form (using a weight matrix derived from theoretical considerations rather

than following details of the neural connectivity in the insect brain). Here we consider whether there is a plausible neural substrate for this computation.

Tangential neurons (TL) of the *lower central body* (CBL) represent the actual input of the polarisation pathway to the CX, with at least three TL types which are all polarisation sensitive. As their name suggests, these neurons provide input tangentially across all columns in the CBL. As such, they could form the basis for the 'fully connected' mapping in our model between POL- and SOL-neurons, as CX columnar neurons innervating the same CBL region could potentially sample from all POL inputs. A plausible candidate would be the CL1 neurons: their receptive field is about 60° wide and their signal-to-noise ratio is relatively low (Heinze, Gotthardt, and Homberg, 2009). There is evidence that CL1-neurons may be homologous to the *ellipsoid-body protocerebral-bridge gall* (EPG) neurons of the *ellipsoid-body* (EB) in flies (Seelig and Jayaraman, 2015; Wolff, Iyer, and Rubin, 2015), which represent landmark orientation and are used for visual navigation. This suggests the same neurons may get information from the visual field, such as the horizon line, which could provide the pitch and roll information that our model assumes will be integrated at this stage of processing to correct for tilting. However, there is no direct evidence as yet to support the existence of our proposed gating function. A fascinating possibility is that this putative TL-CL1 mapping, which would need a rather well-tuned set of weights to extract the compass heading, could be a self-organising network, or at least could be calibrated by the experience of the animal (for example, if it makes coordinated rotations under the sky) as many insects have been observed to do (Wehner, 1997; Dacke, Baird, et al., 2013).

Columnar (CL1) neurons innervate the PB and are presumed to connect to TB1 neurons, hence this could form the physiological basis of our model's connections from SOL- to TCL-neurons. In other insects (for example, fruit flies) the equivalent neurons to CL1 have been shown to form part of a ring attractor network to encode heading relative to a visual target (Seelig and Jayaraman, 2015), which can also hold and update this information in the dark (based on self movement). The PB has recurrent connections to the CBL (EB in flies) which could provide the feedback hypothesised in our model to compensate for time (Su et al., 2017; Kakaria and Bivort, 2017). Alternatively, it has been noted Homberg, 2015 that there is a potential neural pathway from circadian pacemaker circuitry in the accessory medulla to the PB, which could provide an alternative way to adjust the compass with the time of day. As shown in our results (see also, Bech, Homberg, and Pfeiffer, 2014) it may be difficult to interpret the real encoding principles of these neurons using isolated cues if they have evolved to be tuned to the combined input pattern of the real sky, and are potentially modulated by time and the tilting orientation of the animal. Specifically, we see that

the robust $[0, \sim 360°]$ representation of direction in our simulated TCL neurons appears to be a noisy $[0, \sim 180°]$ representation when using a rotating polariser as the stimulus, resembling the data from Heinze and Homberg (2007).

D.3.3 *The sensor array*

We found that the resolution and receptive field of the sensor are optimal for $n = 60$ units resolution and $\omega = 56°$ receptive field. However, the DRA of *Cataglyphis* has more units (ommatidia) and a broader receptive field in the frontocaudal axis ($\omega_a \approx 120°$; see Fig. D.3A). This asymmetric design of the DRA might be explained if we assume the ant's head is tilted more often around the mediolateral axis; as they do not appear to significantly stabilise their head orientation while running up and down hills (Seidl and Wehner, 2008) or while carrying a load (Ardin, Mangan, et al., 2015); whereas there is some evidence that they do stabilise when their body is tilted around the frontocaudal axis Seidl and Wehner, 2008; Raderschall, Narendra, and Zeil, 2016. Therefore more samples on the frontocaudal axis would increase the confidence of their compass when running. Other insect species have distinctive differences in the layout of their dorsal rim: in the precise number of ommatidia, their alignment pattern, their acceptance angle and their spectral sensitives (see Table D.3; for a review see, Labhart and E. P. Meyer, 1999; Karman, Diah, and Gebeshuber, 2012). These might suggest similar adaptations to the specific requirements of habitat, foraging time, task and motor control.

D.3.4 *POL-compass design for robotics*

In this study, we used a somewhat generalised DRA, as it was specifically our intent to consider how we might construct an equivalent sensor for a robot, preferably at a low cost. The insect celestial compass has already inspired the design of a number of polarisation compass sensors, particularly as a robust alternative to magnetic compass sensing for robot applications (Karman, Diah, and Gebeshuber, 2012). The approach applied in the Sahabot (Lambrinos, Kobayashi, et al., 1997; Lambrinos, Möller, et al., 2000; F. Smith and D. Stewart, 2014) introduced a number of biomimetic aspects, including POL-OP units, but used only 3, each one with a relatively small acceptance angle, all pointing towards the zenith, and oriented in angles with $60°$ difference. The dominant polarisation direction was recovered either by rotational scanning or by using a look-up table, with additional light sensing used to decide between the solar and cross-solar directions. The output angular error reported for flat terrain experiments is $1.5°$. Chu et al. (2008) improved this design, using blue filters on photo-receptors

	model	ant	cricket	bee
n	60	~ 112	$530 - 680$	~ 120
$\rho_{\max}$	$5.4°$	$5.4°$	$35°$	$14°$
$\lambda_{\max}$	350 nm (UV)	350 nm (UV)	440 nm (blue)	350 nm (UV)

Table D.3: Properties of dorsal rim ommatidia in different species. n, number of facets; $\rho_{\max}$, acceptance angle; $\lambda_{\max}$, spectral sensitivity. Data from ants (Labhart, 2000; Wehner, 2003), crickets (Labhart, Hodel, and Valenzuela, 1984; Blum and Labhart, 2000; Labhart, 2016), bees (Greiner et al., 2007).

with wider acceptance angles, $\rho \approx 53°$, to obtain a minimum of $0.2°$ angular error. Ma et al. (2014) and Xian et al. (2014) followed the same line, optimising the DOP and AOP extraction using the least-squares method. More recently Dupeyroux, Diperi, et al. (2017) and Dupeyroux, Serres, and Viollet (2019) used UV sensors with orthogonally aligned HNP'B polarisers that were rotated $360°$ every 42 seconds using a stepper motor. Calculating a compass direction from this method produced from $0.3°$ to $1.9°$ peak errors in clear and cloudy skies respectively. Similar robot implementations include those by J. Chahl and Mizutani (2012) and S. Zhang et al. (2014). Alternative approaches use a camera (Horvath et al., 2002; Horstmeyer, Euliss, and Athale, 2009; Manakov et al., 2013; W. Zhang, Cao, X. Zhang, Y. Yang, et al., 2017; Stürzl, 2017) or multi-camera system (Horvath et al., 2002; S. Zhang et al., 2014; Stürzl and Carey, 2012), or specialised image sensors with different polarisation sensitivity for each pixel (Karman, Diah, and Gebeshuber, 2012; Sarkar et al., 2011). Good results are obtained by Stürzl (2017), who built a single camera sensor with 3 lenses and 3 polarisers oriented at different angles. The sensor estimates the angle and DOP of the skylight and fits a sky model on the input stream to estimate the parameters, which are the solar azimuth and elevation. In addition, they also estimate a covariance matrix that shows the confidence of their prediction. Their approach also works in tilted environments by integrating inertial measurement unit (IMU) data. Finally, Z. Yang et al. (2017) followed up Sahabot's work, used 2 POL-OP units oriented at different angles from the zenith, placed them on a plane, and used the scanning technique in order to get values from different angles. They show that their sensor can estimate the solar azimuth and elevation in clear sky conditions with MAE $0.2°$ and $0.4°$ respectively.

Our model suggests an alternative sensor design, in which a larger number of POL-OP sensors are used to sample specific areas of the sky, but these do not form a complete image as in the camera-based systems described above. Such a sensor could be built from off-the-shelf components; for example, using pairs of UV photodiodes and linear polarisation filters to imitate dorsal rim ommatidia photo-receptor neurons. The components could be mounted on a dome, creating a similar DRA to ant eyes.

As we have shown, the subsequent computation to recover the heading direction is relatively low-cost and could easily be carried out on a robot-compatible microprocessor, which could also run our path integration model. We hope to build this sensor and test it on a robot in future work.

D.3.5 *Conclusion*

As well as building a physical implementation, there are several other ways in which this model could be developed in the future. One consideration already discussed above would be to integrate parallel processing of luminance and spectral cues, which can provide complementary information to polarisation, and thus enhance the reliability with which the solar azimuth can be determined over a wider range of conditions. A second would be to examine whether a more direct mapping can be made between the computational processing we have proposed and the detailed neuroanatomy of the layers intervening between the POL neurons in the medulla and the compass neurons in the CX. Finally, we believe it is key that such a model remains grounded in the understanding of the real task constraints the circuit needs to support, which is the natural environment conditions under which insect path integration evolved and operates.

E

SUPPLEMENTARY MATERIAL

MC connection	IC connection	Microcircuit	Evidence	References
PPL1-01 $\multimap$ MBON-11	$d_{av} \multimap s_{at}$	SM	A, F	1, 3
MBON-11 $\dashv$ PPL1-01	$s_{at} \dashv d_{av}$	SM	A, F	1, 3
PAM-07 $\multimap$ MBON-05	$d_{at} \multimap s_{av}$	SM	A	1
MBON-05 $\dashv$ PAM-07	$s_{av} \dashv d_{at}$	SM	A, F	1, 4
MBON-11 $\dashv$ MBON-01	$s_{at} \dashv r_{av}$	RM	A, F	1, 5
MBON-05 $\dashv$ MBON-12	$s_{av} \dashv r_{at}$	RM	A	1
PPL1-03$_2$ $\multimap$ MBON-12	$c_{av} \multimap r_{at}$	RSM	A, F	1, 6
MBON-12 $\rightarrow$ PAM-02	$r_{at} \rightarrow c_{at}$	RSM	A, F	1, 6
PAM-02 $\multimap$ MBON-01	$c_{at} \multimap r_{av}$	RSM	A, F	1, 6
MBON-01 $\rightarrow$ PPL1-03$_2$	$r_{av} \rightarrow c_{av}$	RSM	A	2
PPL1-03$_2$ $\multimap$ MBON-15	$c_{av} \multimap m_{av}$	LTM	A	1
MBON-15 $\rightarrow$ PPL1-03$_2$	$m_{av} \rightarrow c_{av}$	LTM	A	1, 2
PAM-02 $\multimap$ MBON-02	$c_{at} \multimap m_{at}$	LTM	A	1
MBON-02 $\rightarrow$ PAM-02	$m_{at} \rightarrow c_{at}$	LTM	A	1, 2
PPL1-03$_1$ $\multimap$ MBON-15	$f_{at} \multimap m_{av}$	RLM	A	1
MBON-15 $\rightarrow$ PAM-04	$m_{av} \rightarrow f_{av}$	RLM	A	2
PAM-04 $\multimap$ MBON-02	$f_{av} \multimap m_{at}$	RLM	A	1
MBON-02 $\rightarrow$ PPL1-03$_1$	$m_{at} \rightarrow f_{at}$	RLM	A	2
PPL1-03$_1$ $\multimap$ MBON-12	$f_{at} \multimap r_{at}$	MAM	A	1
PAM-04 $\multimap$ MBON-01	$f_{av} \multimap r_{av}$	MAM	A	1

Table E.1: Connections among neurons in the fruit-fly *mushroom body* (MB) mapped to the connections of the *incentive circuit* (IC). Connection types: '$\multimap$', modulates the synaptic weights of the KC $\rightarrow$ MBON connections terminating in that *mushroom body output neuron* (MBON); '$\rightarrow$', excitatory connection; '$\dashv$', inhibitory connection. Microcircuit—SM: susceptible memory, RM: restrained memory, RSM: reciprocal short-term memories, LTM: long-term memory, RLM: reciprocal long-term memories, MAM: memory assimilation mechanism. Evidence—A: anatomical connection is known (using light or electron microscopy), F: functional connection is known (whether activating the presynaptic neuron leads to an excitatory or inhibitory effect on the postsynaptic neuron, and/or the neurotransmitter released by the presynaptic neuron). References: (1) Aso, Hattori, et al. (2014), (2) Li et al. (2020), (3) Pavlowsky et al. (2018), (4) Cohn, Morantte, and Ruta (2015), (5) Felsenberg, Jacob, et al. (2018), (6) McCurdy et al. (2021).

Driver	Cluster	Neuron name	#cells	Short name	Alternative names
MB011B	M4/M6	MBON-γ5β′2a	1	MBON-01	MB-M6
MB399B	M4/M6	MBON-β2β′2a	1	MBON-02	
MB298B	MV2	MBON-γ4>γ1γ2	1	MBON-05	
MB112C		MBON-γ1pedc>α/β	1	MBON-11	MB-MVP2
MB077B	V3/V4	MBON-γ2α′1	2	MBON-12	
MB050B	V2	MBON-α′1	3	MBON-15	
MB109B	PAM	PAM-β′2a	6-9	PAM-02	
MB301B	PAM	PAM-β2β′2a	> 3	PAM-04	subset of MB-M8
MB312B	PAM	PAM-γ4<γ1γ2	13-17	PAM-07	subset of MB-AIM?
MB320C	PPL1	PPL1-γ1ped	1	PPL1-01	MB-MP1, MP
$MB296B_1$	PPL1	$PPL1\text{-}\gamma2\alpha'1_1$	1	PPL1-03	MB-MV1
$MB296B_2$	PPL1	$PPL1\text{-}\gamma2\alpha'1_2$	1	PPL1-03	MB-MV1

Table E.2: Additional information on the *mushroom body* (MB) neurons used for the *incentive circuit* (IC), sorted by their short name. Information about the neurons has been collected from McCurdy et al. (2021) and Aso, Hattori, et al. (2014).

ID (i)	θ^i	ϕ^i	ID (i)	θ^i	ϕ^i	ID (i)	θ^i	ϕ^i
1	5.13	45.00	21	15.39	−105.00	41	25.65	9.00
2	5.13	135.00	22	15.39	−75.00	42	25.65	27.00
3	5.13	−135.00	23	15.39	−45.00	43	25.65	45.00
4	5.13	−45.00	24	15.39	−15.00	44	25.65	63.00
5	10.26	22.50	25	20.52	11.25	45	25.65	81.00
6	10.26	67.50	26	20.52	33.75	46	25.65	99.00
7	10.26	112.50	27	20.52	56.25	47	25.65	117.00
8	10.26	157.50	28	20.52	78.75	48	25.65	135.00
9	10.26	−157.50	29	20.52	101.25	49	25.65	153.00
10	10.26	−112.50	30	20.52	123.75	50	25.65	171.00
11	10.26	−67.50	31	20.52	146.25	51	25.65	−171.00
12	10.26	−22.50	32	20.52	168.75	52	25.65	−153.00
13	15.39	15.00	33	20.52	−168.75	53	25.65	−135.00
14	15.39	45.00	34	20.52	−146.25	54	25.65	−117.00
15	15.39	75.00	35	20.52	−123.75	55	25.65	−99.00
16	15.39	105.00	36	20.52	−101.25	56	25.65	−81.00
17	15.39	135.00	37	20.52	−78.75	57	25.65	−63.00
18	15.39	165.00	38	20.52	−56.25	58	25.65	−45.00
19	15.39	−165.00	39	20.52	−33.75	59	25.65	−27.00
20	15.39	−135.00	40	20.52	−11.25	60	25.65	−9.00

Table E.3: Spherical coordinates of the exact positions of the POL units on the compass dome. ID (i), the identity of the *polarisation sensitive* (POL) unit referring to the figures of the main text; θ^i, the zenith distance of the i^{th} unit; ϕ^i, the azimuth of the i^{th} unit. The orientation of each unit is a function of its relative azimuth to the heading direction of the dome: $\alpha^i = \phi_i - 90°$.

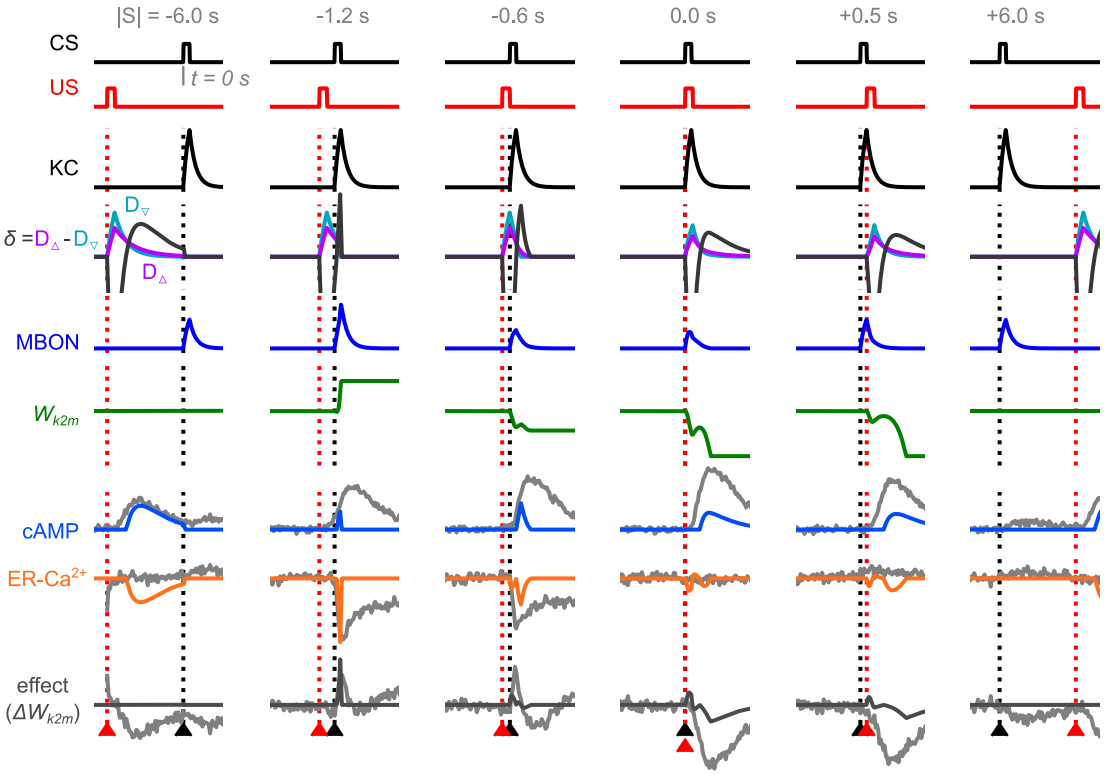

Figure E.1: All the chemical levels and neural activities calculated based on the order of the *conditioned* (CS) and *unconditioned stimuli* (US). The black arrowhead marks the time of the CS (duration $0.5\,\text{sec}$); the red arrowhead marks the time of the US (duration $0.6\,\text{sec}$), similar to Handler et al. (2019)—Figure 5D. Predicted ER$-$Ca^{2+}, cAMP and the plasticity effect responses are drawn on top of the data from the original paper (Handler et al., 2019)—grey lines.

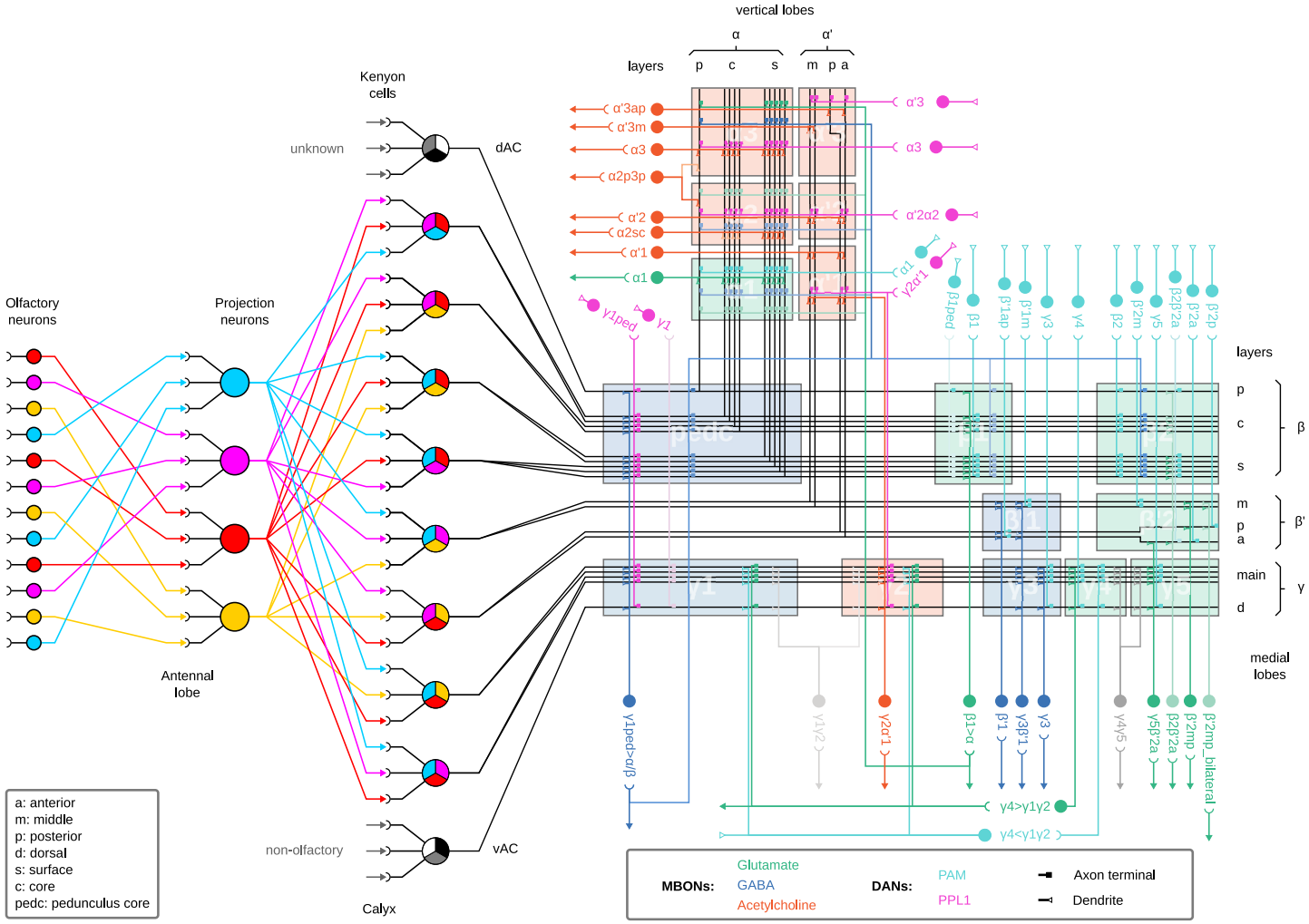

Figure E.2: Anatomy of olfactory pathways in the fly brain. A schematic representation of the key cellular components and information flow during the processing of olfactory inputs to the *mushroom body* (MB). Olfactory receptor neurons expressing the same odourant receptor converge onto a single glomerulus in the *antennal lobe* (AL). A small number (generally 3-4) of *projection neurons* (PNs) from each of the 51 AL glomeruli innervate the MB calyx where they synapse on the dendrites of the $\sim$ 2000 *Kenyon cells* (KCs) in a globular structure, the calyx. Each KC exhibits, on average, 6.4 dendric 'claws', and each claw is innervated by a single PN. There is little order in the connection patterns of PNs to KCs. The axons of the KCs project in parallel anteriorly through the pedunculus to the lobes, where KCs terminate onto the dendrites of *mushroom body output neurons* (MBONs). Adapted and modified from Aso, Hattori, et al. (2014).

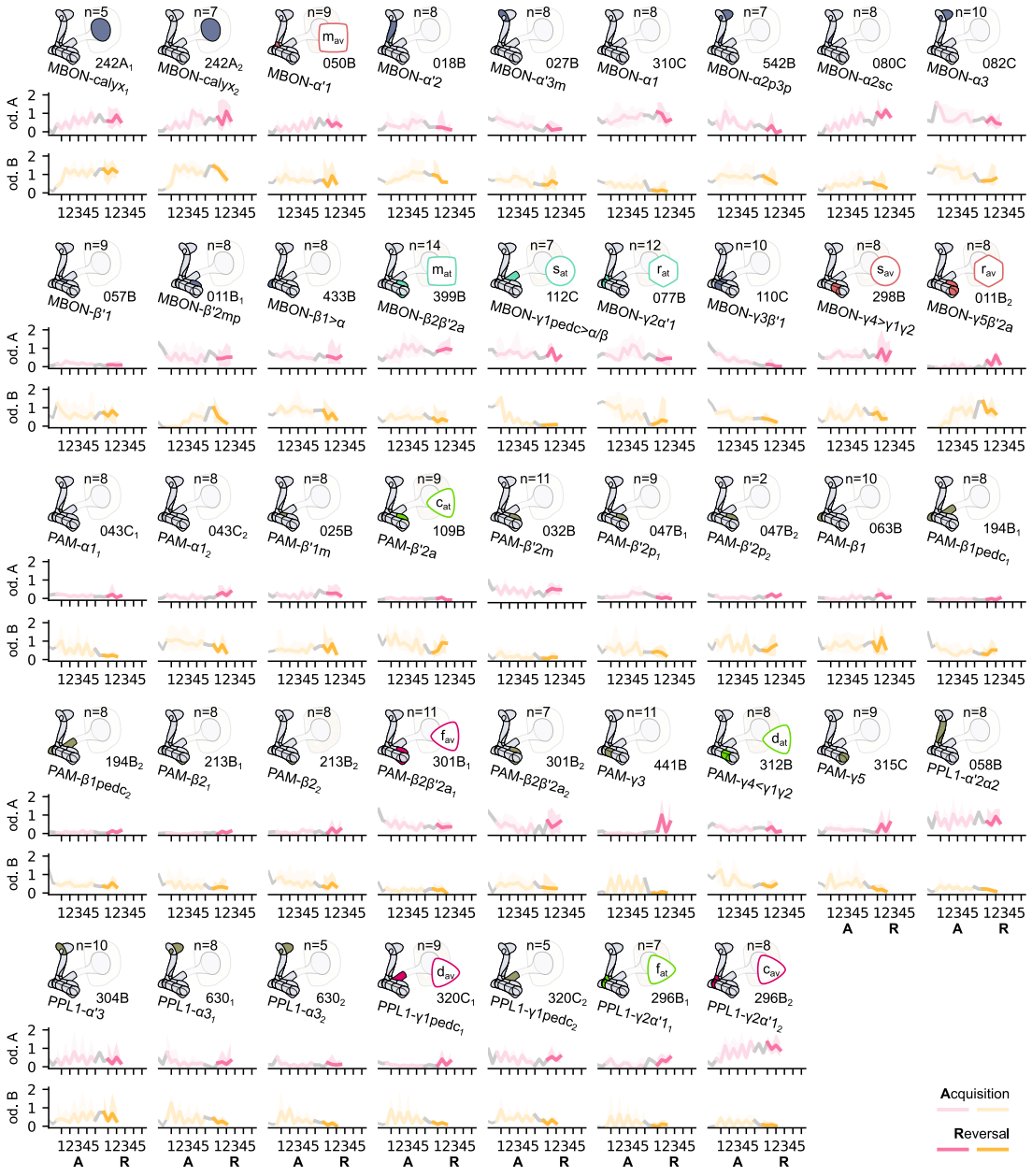

Figure E.3: The responses from all the recorded neurons in the *Drosophila melanogaster mushroom body* (MB) during the experiment described in Fig. 3.6B. Each row presents the driver of the recorded neurons and a schematic representation of where they innervate the MB; the median responses of the neuron for each odour (coloured as pink for odour A or yellow for odour B) over a number of flies, denoted as n, and the 25% and 75% quantiles marked by the coloured region.

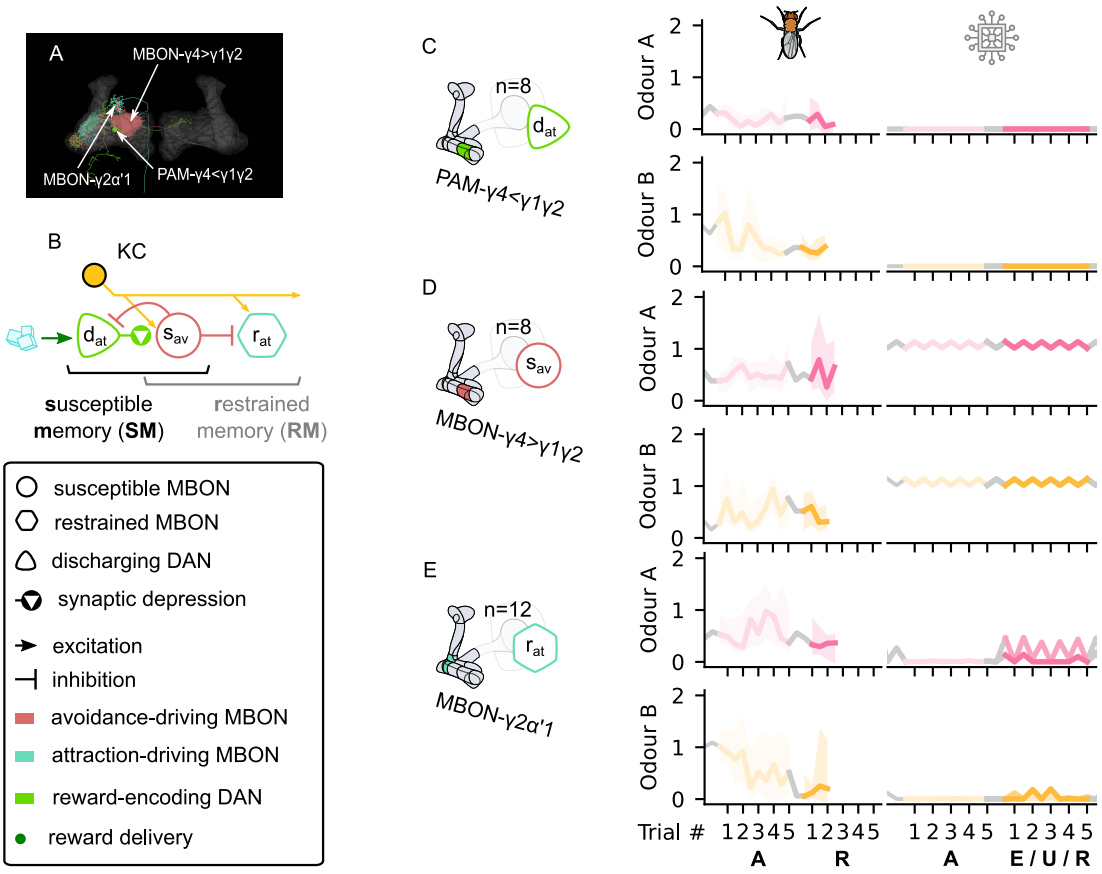

Figure E.4: The *susceptible* (SM) and *restrained memory* (RM) microcircuits of the *mushroom body* (MB). (A) Image of the avoidance-driving SM and attraction-driving RM microcircuits made of the PAM-γ4<γ1γ2, MBON-γ4>γ1γ2 and MBON-γ2α′1 neurons—created using the Virtual Fly Brain software (Milyaev et al., 2012). (B) Schematic representation of the SM and RM microcircuits connected via the susceptible *mushroom body output neuron* (MBON). The responses of (C) the reward-encoding discharging *dopaminergic neuron* (DAN), d_{at}, (D) the avoidance-driving susceptible MBON, s_{av}, and (E) the attraction-driving restrained MBON, r_{at}, generated by experimental data (left) and the model (right) during the olfactory conditioning paradigms of Fig. 3.7D. The lightest shades denote the extinction, the mid shades the unpaired and the dark shades the reversal phase. Each trial happens in two consecutive time steps: the off-shock (odour only) followed by the on-shock (paired odour and shock) when available (odour B in acquisition and odour A in reversal phase) otherwise a second off-shock time-step (all the other phases).

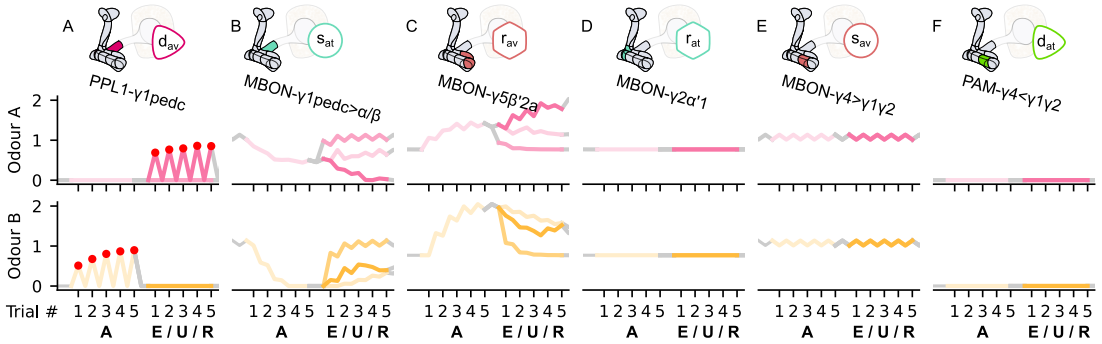

Figure E.5: The responses of the neurons of the *incentive circuit* (IC) using only the connections of the *susceptible* (SM) and *restrained memory* (RM) microcircuits. The responses of (A) the punishment-encoding discharging *dopaminergic neuron* (DAN), d_{av}, (B) the attraction-driving susceptible *mushroom body output neuron* (MBON), s_{at}, (C) the avoidance-driving restrained MBON, r_{av}, (D) the attraction-driving restrained MBON, r_{at}, (E) the avoidance-driving susceptible MBON, s_{av}, and (F) the reward-encoding discharging DAN, d_{at}, generated by experimental data (left) and the model (right) during the olfactory conditioning paradigms of Fig. 3.7D. The lightest shades denote the extinction, the mid shades the unpaired and the dark shades the reversal phase. The first row of responses (coloured pink) corresponds to responses associated with odour A, and the second (coloured yellow) to those associated with odour B. Each trial happens in two consecutive time steps: the off-shock (odour only) followed by the on-shock (paired odour and shock) when available (odour B in acquisition and odour A in reversal phase) otherwise a second off-shock time-step (all the other phases).

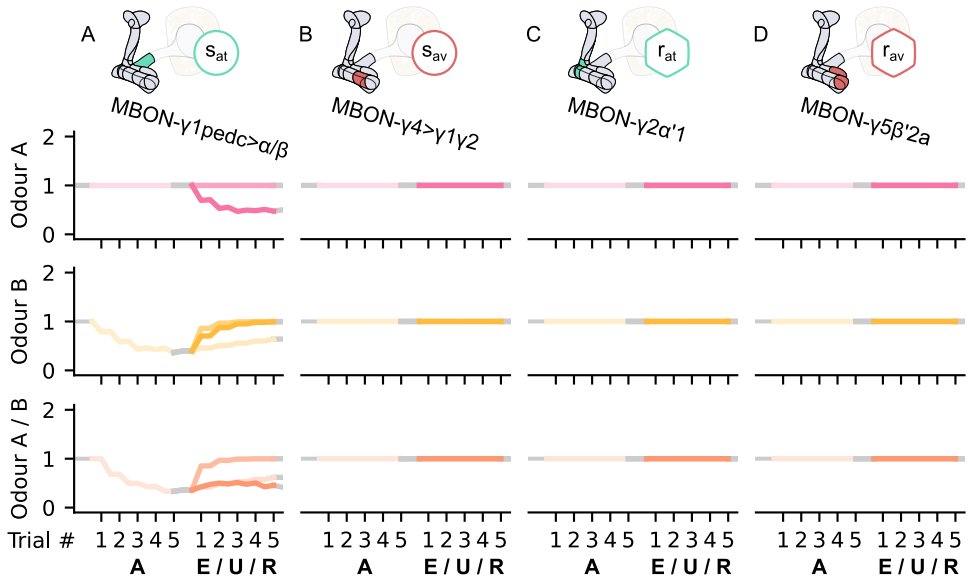

Figure E.6: The KC → MBON synaptic weights of the neurons of the *incentive circuit* (IC) using only the connections of the *susceptible* (SM) and *restrained memory* (RM) microcircuits. The synaptic weights of (A) the attraction-driving susceptible *mushroom body output neuron* (MBON), s_{at}, (B) the avoidance-driving susceptible MBON, s_{av}, (C) the attraction-driving restrained MBON, r_{at}, and (D) the avoidance-driving restrained MBON, r_{av}, generated by experimental data (left) and the model (right) during the olfactory conditioning paradigms of Fig. 3.7D. The lightest shades denote the extinction, the mid shades the unpaired and the dark shades the reversal phase. The first row of weights (coloured pink) corresponds to *Kenyon cells* (KCs) associated with odour A, the second (coloured yellow) to KCs associated with odour B and the third (coloured orange) to KCs associated with both odours. Each trial happens in two consecutive time steps: the off-shock (odour only) followed by the on-shock (paired odour and shock) when available (odour B in acquisition and odour A in reversal phase) otherwise a second off-shock timestep (all the other phases).

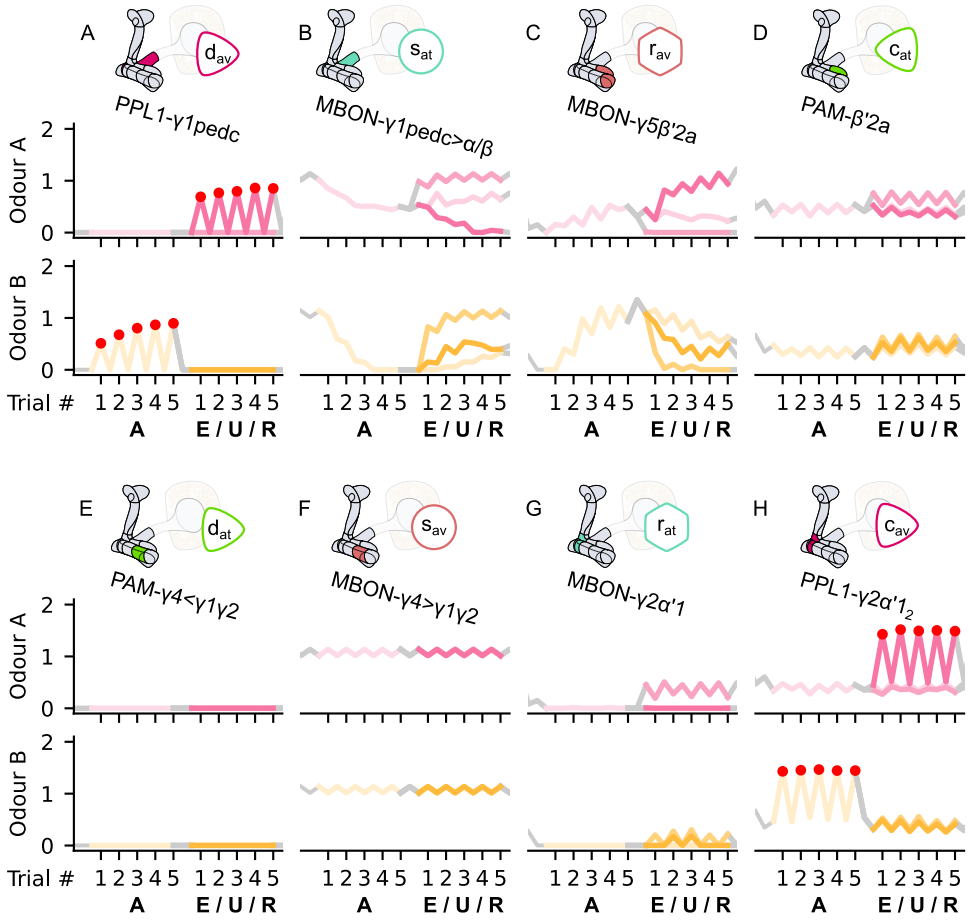

Figure E.7: The responses of the neurons of the *incentive circuit* (IC) using only the connections of the *susceptible* (SM), *restrained* (RM), and *reciprocal short-term memories* (RSM) microcircuits. The responses of (A) the punishment-encoding discharging *dopaminergic neuron* (DAN), d_{av}, (B) the attraction-driving susceptible *mushroom body output neuron* (MBON), s_{at}, (C) the avoidance-driving restrained MBON, r_{av}, (D) the reward-encoding charging DAN, c_{at}, (E) the reward-encoding discharging DAN, d_{at}, (F) the avoidance-driving susceptible MBON, s_{av}, (G) the attraction-driving restrained MBON, r_{at}, and (H) the punishment-encoding charging DAN, c_{av}, generated by experimental data (left) and the model (right) during the olfactory conditioning paradigms of Fig. 3.7D. The lightest shades denote the extinction, the mid shades the unpaired and the dark shades the reversal phase. The first row of responses (coloured pink) corresponds to responses associated with odour A, and the second (coloured yellow) to those associated with odour B. Each trial happens in two consecutive time steps: the off-shock (odour only) followed by the on-shock (paired odour and shock) when available (odour B in acquisition and odour A in reversal phase) otherwise a second off-shock time-step (all the other phases).

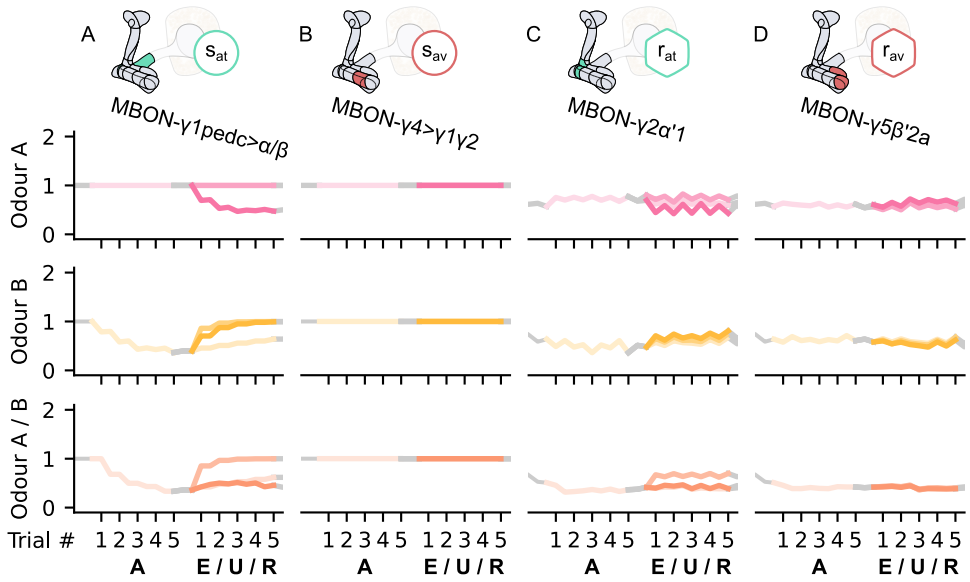

Figure E.8: The KC → MBON synaptic weights of the neurons of the *incentive circuit* (IC) using only the connections of the *susceptible* (SM), *restrained* (RM), and *reciprocal short-term memories* (RSM) microcircuits. The synaptic weights of (A) the attraction-driving susceptible *mushroom body output neurons* (MBON), s_{at}, (B) the avoidance-driving susceptible MBON, s_{av}, (C) the attraction-driving restrained MBON, r_{at}, and (D) the avoidance-driving restrained MBON, r_{av}, generated by experimental data (left) and the model (right) during the olfactory conditioning paradigms of Fig. 3.7D. The lightest shades denote the extinction, the mid shades the unpaired and the dark shades the reversal phase. The first row of weights (coloured pink) corresponds to *Kenyon cells* (KCs) associated with odour A, the second (coloured yellow) to KCs associated with odour B and the third (coloured orange) to KCs associated with both odours. Each trial happens in two consecutive time steps: the off-shock (odour only) followed by the on-shock (paired odour and shock) when available (odour B in acquisition and odour A in reversal phase) otherwise a second off-shock time-step (all the other phases).

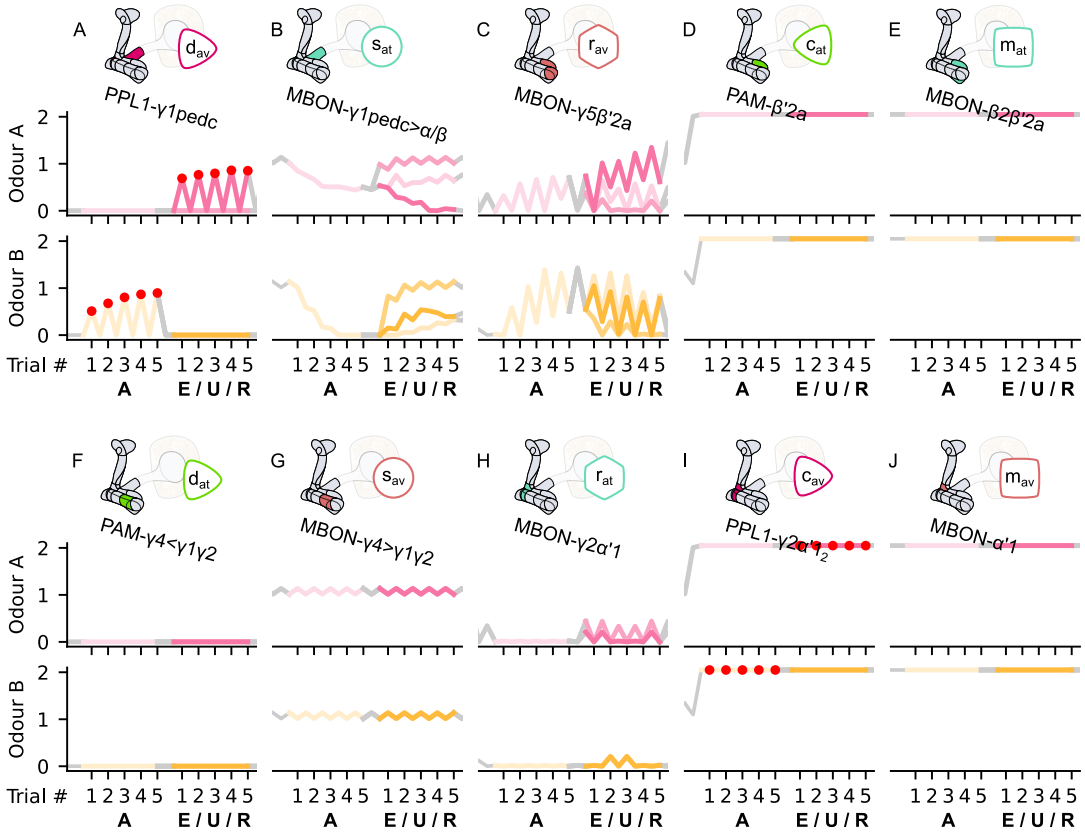

Figure E.9: The responses of the neurons of the incentive circuit using only the connections of the *susceptible* (SM), *restrained* (RM), and *reciprocal short-term* (RSM), and *long-term memories* (LTMs) microcircuits. The responses of (A) the punishment-encoding discharging *dopaminergic neuron* (DAN), d_{av}, (B) the attraction-driving susceptible *mushroom body output neuron* (MBON), s_{at}, (C) the avoidance-driving restrained MBON, r_{av}, (D) the reward-encoding charging DAN, c_{at}, (E) the attraction-driving LTM MBON, m_{at}, (F) the reward-encoding discharging DAN, d_{at}, (G) the avoidance-driving susceptible MBON, s_{av}, (H) the attraction-driving restrained MBON, r_{at}, (I) the punishment-encoding charging DAN, c_{av}, and (J) the avoidance-driving LTM MBON, m_{av}, generated by experimental data (left) and the model (right) during the olfactory conditioning paradigms of Fig. 3.7D. The lightest shades denote the extinction, the mid shades the unpaired and the dark shades the reversal phase. The first row of responses (coloured pink) corresponds to responses associated with odour A, and the second (coloured yellow) to those associated with odour B. Each trial happens in two consecutive time steps: the off-shock (i. e., odour only) followed by the on-shock (i. e., paired odour and shock) when available (i. e., odour B in acquisition and odour A in reversal phase) otherwise a second off-shock time-step (i. e., all the other phases).

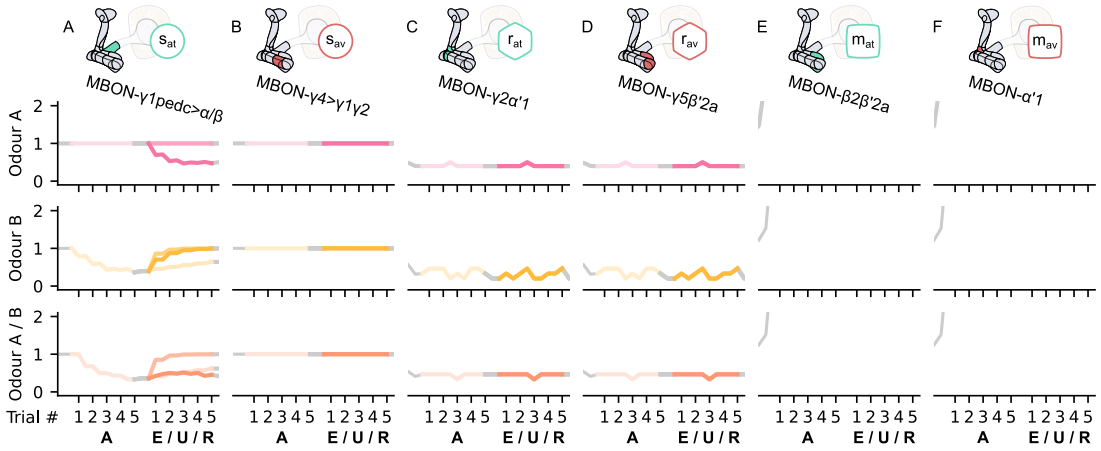

Figure E.10: The KC → MBON synaptic weights of the neurons of the *incentive circuit* (IC) using only the connections of the *susceptible*, *restrained* (RM), *reciprocal short-term* (RSM) and *long-term memories* (LTMs) microcircuits. The synaptic weights of (A) the attraction-driving susceptible *mushroom body output neuron* (MBON), s_{at}, (B) the avoidance-driving susceptible MBON, s_{av}, (C) the attraction-driving restrained MBON, r_{at}, (D) the avoidance-driving restrained MBON, r_{av}, (E) the attraction-driving LTM MBON, m_{at}, and (F) the avoidance-driving LTM MBON, m_{av}, generated by experimental data (left) and the model (right) during the olfactory conditioning paradigms of Fig. 3.7D. The lightest shades denote the extinction, the mid shades the unpaired and the dark shades the reversal phase. The first row of weights (coloured pink) corresponds to *Kenyon cells* (KCs) associated with odour A, the second (coloured yellow) to KCs associated with odour B and the third (coloured orange) to KCs associated with both odours. Each trial happens in two consecutive time steps: the off-shock (odour only) followed by the on-shock (paired odour and shock) when available (odour B in acquisition and odour A in reversal phase) otherwise a second off-shock time-step (all the other phases).

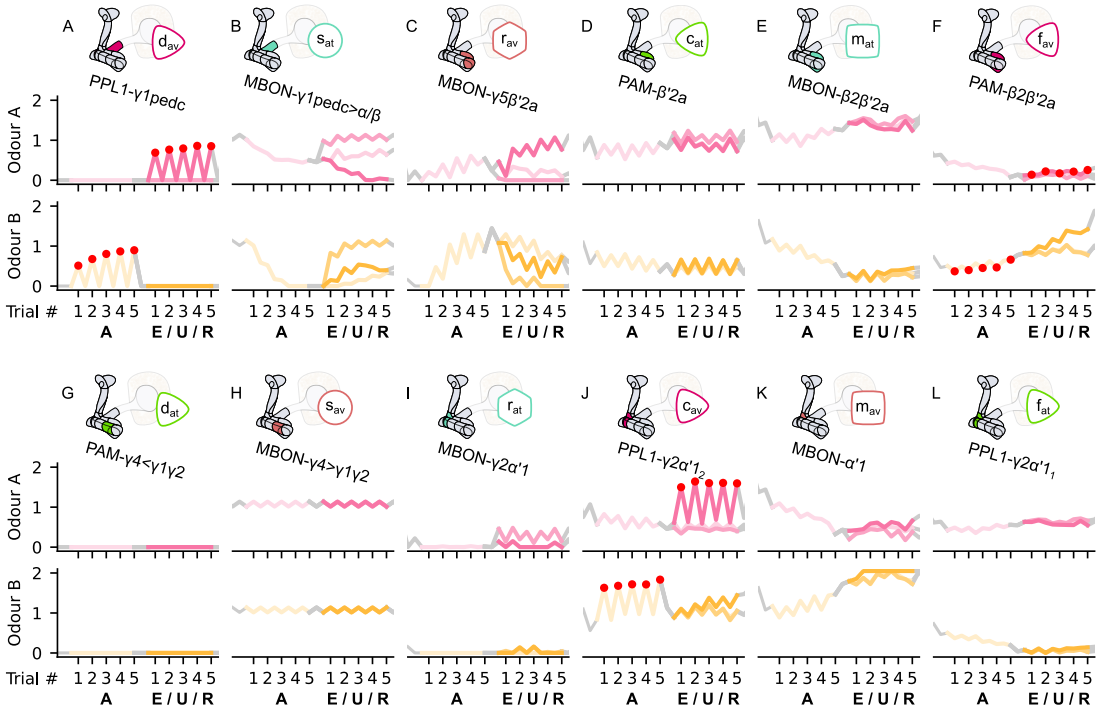

Figure E.11: The responses of the neurons of the *incentive circuit* (IC) using all the connections except of the *memory assimilation mechanism* (MAM) microcircuit. The responses of (A) the punishment-encoding discharging *dopaminergic neuron* (DAN), d_{av}, (B) the attraction-driving susceptible *mushroom body output neuron* (MBON), s_{at}, (C) the avoidance-driving restrained MBON, r_{av}, (D) the reward-encoding charging DAN, c_{at}, (E) the attraction-driving *long-term memory* (LTM) MBON, m_{at}, (F) the avoidance-driving forgetting DAN, f_{av}, (G) the reward-encoding discharging DAN, d_{at}, (H) the avoidance-driving susceptible MBON, s_{av}, (I) the attraction-driving restrained MBON, r_{at}, (J) the punishment-encoding charging DAN, c_{av}, (K) the avoidance-driving LTM MBON, m_{av}, and (L) the attraction-driving forgetting DAN, f_{at}, generated by experimental data (left) and the model (right) during the olfactory conditioning paradigms of Fig. 3.7D. The lightest shades denote the extinction, the mid shades the unpaired and the dark shades the reversal phase. The first row of responses (coloured pink) corresponds to responses associated with odour A, and the second (coloured yellow) to those associated with odour B. Each trial happens in two consecutive time steps: the off-shock (odour only) followed by the on-shock (paired odour and shock) when available (odour B in acquisition and odour A in reversal phase) otherwise a second off-shock time-step (all the other phases).

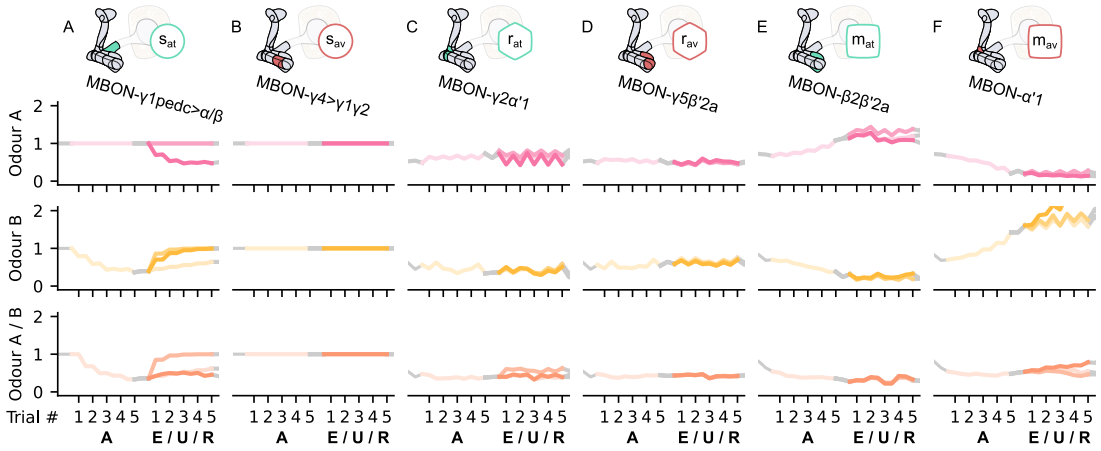

Figure E.12: The KC → MBON synaptic weights of the neurons of the *incentive circuit* (IC) using all the connections except of the *memory assimilation mechanism* (MAM) microcircuits. The synaptic weights of (A) the attraction-driving susceptible *mushroom body output neuron* (MBON), s_{at}, (B) the avoidance-driving susceptible MBON, s_{av}, (C) the attraction-driving restrained MBON, r_{at}, (D) the avoidance-driving restrained MBON, r_{av}, (E) the attraction-driving *long-term memory* (LTM) MBON, m_{at}, and (F) the avoidance-driving LTM MBON, m_{av}, generated by experimental data (left) and the model (right) during the olfactory conditioning paradigms of Fig. 3.7D. The lightest shades denote the extinction, the mid shades the unpaired and the dark shades the reversal phase. The first row of weights (coloured pink) corresponds to *Kenyon cells* (KCs) associated with odour A, the second (coloured yellow) to KCs associated with odour B and the third (coloured orange) to KCs associated with both odours. Each trial happens in two consecutive time steps: the off-shock (odour only) followed by the on-shock (paired odour and shock) when available (odour B in acquisition and odour A in reversal phase) otherwise a second off-shock time-step (all the other phases).

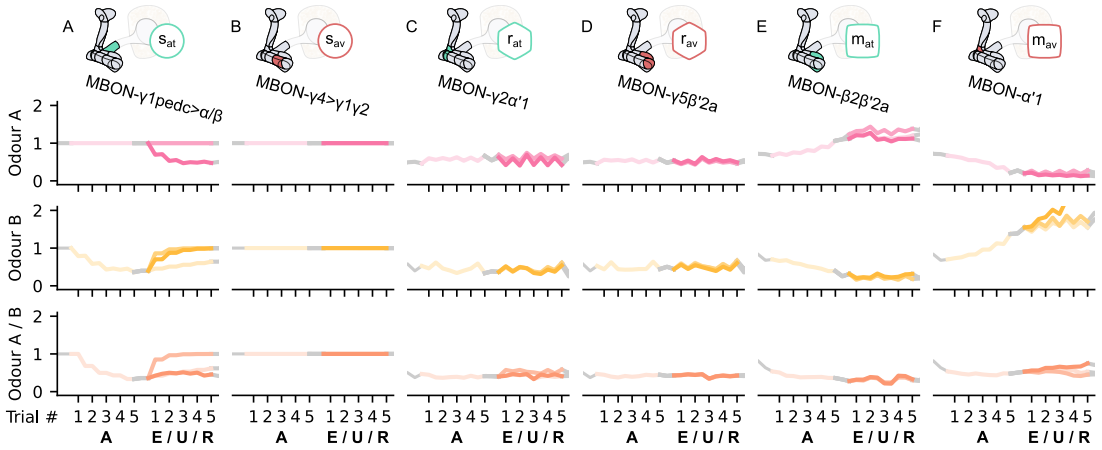

Figure E.13: The KC → MBON synaptic weights of the neurons of the *incentive circuit* using all its connections. The synaptic weights of (A) the attraction-driving susceptible *mushroom body output neuron* (MBON), s_{at}, (B) the avoidance-driving susceptible MBON, s_{av}, (C) the attraction-driving restrained MBON, r_{at}, (D) the avoidance-driving restrained MBON, r_{av}, (E) the attraction-driving *long-term memory* (LTM) MBON, m_{at}, and (F) the avoidance-driving LTM MBON, m_{av}, generated by experimental data (left) and the model (right) during the olfactory conditioning paradigms of Fig. 3.7D. The lightest shades denote the extinction, the mid shades the unpaired and the dark shades the reversal phase. The first row of weights (coloured pink) corresponds to Kenyon cells (KCs) associated with odour A, the second (coloured yellow) to KCs associated with odour B and the third (coloured orange) to KCs associated with both odours.

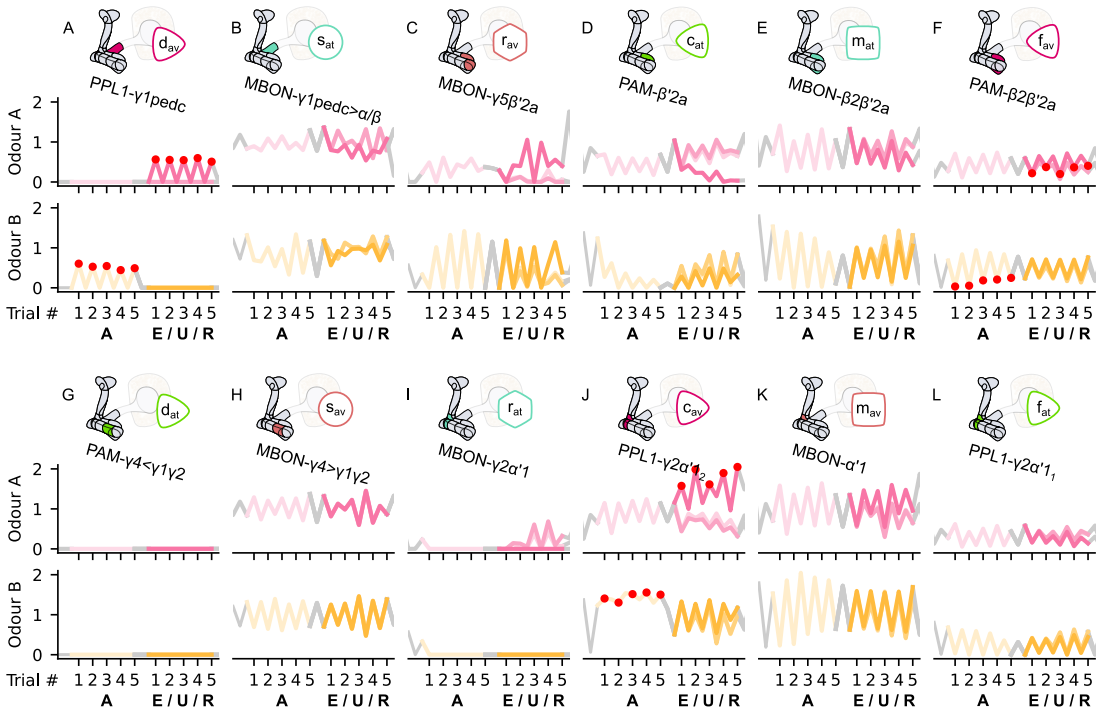

Figure E.14: The reconstructed responses of the neurons of the *incentive circuit* using the *reward prediction error* (RPE) plasticity rule. The reconstructed responses of (A) the punishment-encoding discharging *dopaminergic neuron* (DAN), d_{av}, (B) the attraction-driving susceptible *mushroom body output neuron* (MBON), s_{at}, (C) the avoidance-driving restrained MBON, r_{av}, (D) the reward-encoding charging DAN, c_{at}, (E) the attraction-driving *long-term memory* (LTM) MBON, m_{at}, (F) the punishment-encoding forgetting DAN, f_{av}, (G) the reward-encoding discharging DAN, d_{at}, (H) the avoidance-driving susceptible MBON, s_{av}, (I) the attraction-driving restrained MBON, r_{at}, (J) the punishment-encoding charging DAN, c_{av}, (K) the avoidance-driving LTM MBON, m_{av}, and (L) the reward-encoding forgetting DAN, f_{at}, generated by experimental data (left) and the model (right) during the olfactory conditioning paradigms of Fig. 3.7D. The lightest shades denote the extinction, the mid shades the unpaired and the dark shades the reversal phase.

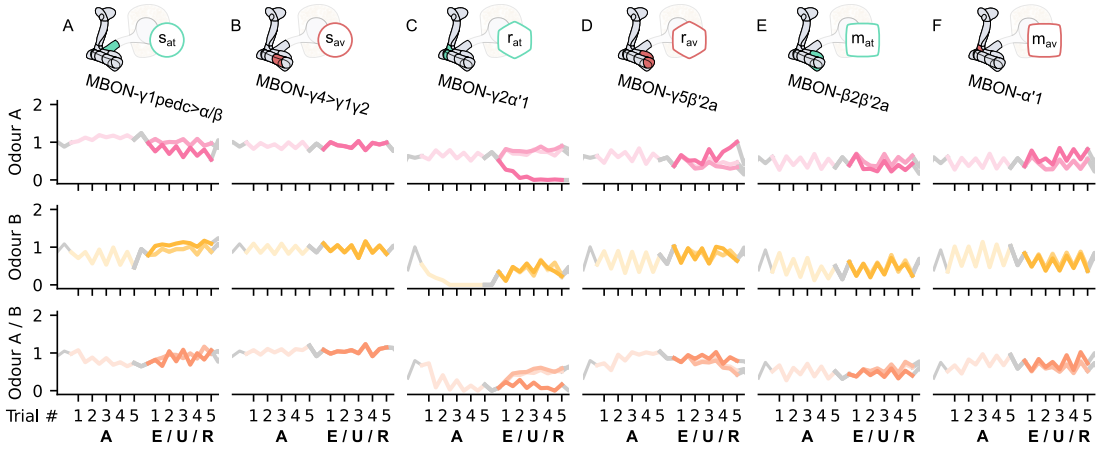

Figure E.15: The KC → MBON synaptic weights of the neurons of the *incentive circuit* (IC) using the *reward prediction error* (RPE) plasticity rule. The synaptic weights of (A) the attraction-driving susceptible *mushroom body output neuron* (MBON), s_{at}, (B) the avoidance-driving susceptible MBON, s_{av}, (C) the attraction-driving restrained MBON, r_{at}, (D) the avoidance-driving restrained MBON, r_{av}, (E) the attraction-driving *long-term memory* (LTM) MBON, m_{at}, and (F) the avoidance-driving LTM MBON, m_{av}, generated by experimental data (left) and the model (right) during the olfactory conditioning paradigms of Fig. 3.7D. The lightest shades denote the extinction, the mid shades the unpaired and the dark shades the reversal phase. The first row of weights (coloured pink) corresponds to *Kenyon cells* (KCs) associated with odour A, the second (coloured yellow) to KCs associated with odour B and the third (coloured orange) to KCs associated with both odours.

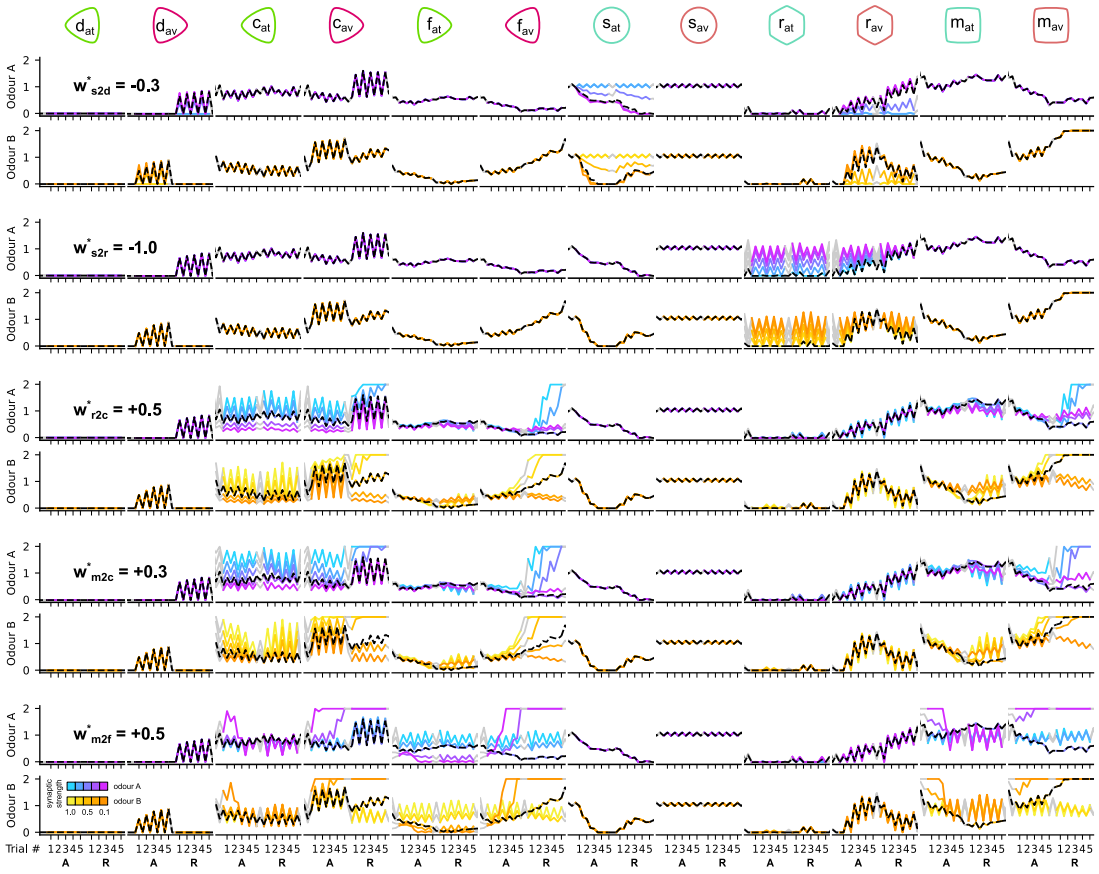

Figure E.16: The responses of the *dopaminergic neurons* (DANs) and *mushroom body output neurons* (MBONs) of the circuit when altering the pre-synaptic strengths of MBONs during the reversal condition. Each column corresponds to the responses of a different neuron. Odd and even rows show the responses of the neurons to odour A and B respectively. Pairs of rows (consecutive odour A and odour B) show the responses of the neurons for the different values of the target parameter (indicated in the first column). The Black dashed line shows the responses of the neurons for the chosen (*) parameter. The different colour codes show the responses of the neurons for the different (absolute) values of the target parameter as indicated in the colour bar on the bottom left.

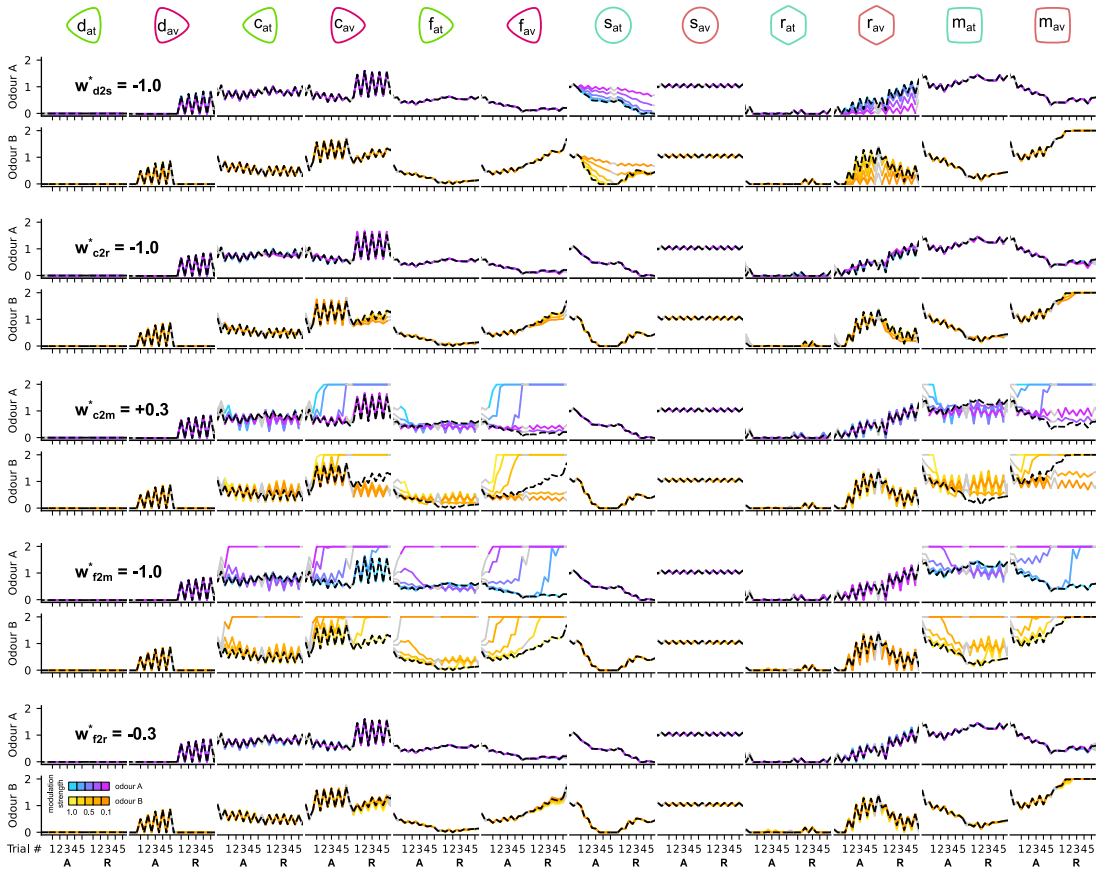

Figure E.17: The responses of the *dopaminergic neurons* (DANs) and *mushroom body output neurons* (MBONs) of the circuit when altering the *dopaminergic* (DA) modulation strengths during the reversal condition. Each column corresponds to the responses of a different neuron. Odd and even rows show the responses of the neurons to odour A and B respectively. Pairs of rows (consecutive odour A and odour B) show the responses of the neurons for the different values of the target parameter (indicated in the first column). The Black dashed line shows the responses of the neurons for the chosen (*) parameter. The different colour codes show the responses of the neurons for the different values of the target parameter as indicated in the colour bar on the bottom left.

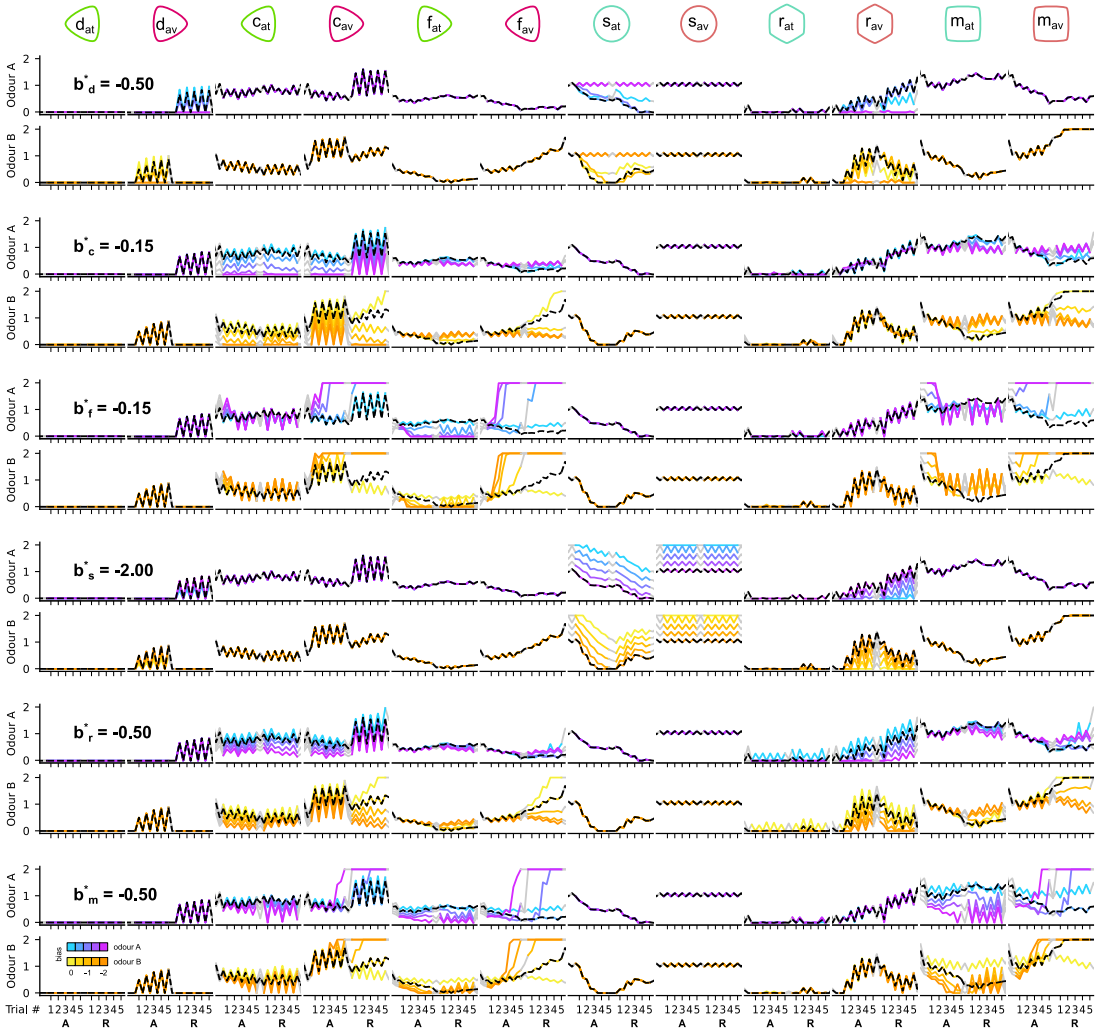

Figure E.18: The responses of the *dopaminergic neurons* (DANs) and *mushroom body output neurons* (MBONs) of the circuit when altering the DAN and MBON biases during the reversal condition. Each column corresponds to the responses of a different neuron. Odd and even rows show the responses of the neurons to odour A and B respectively. Pairs of rows (consecutive odour A and odour B) show the responses of the neurons for the different values of the target parameter (indicated in the first column). The Black dashed line shows the responses of the neurons for the chosen (*) parameter. The different colour codes show the responses of the neurons for the different values of the target parameter as indicated in the colour bar on the bottom left.

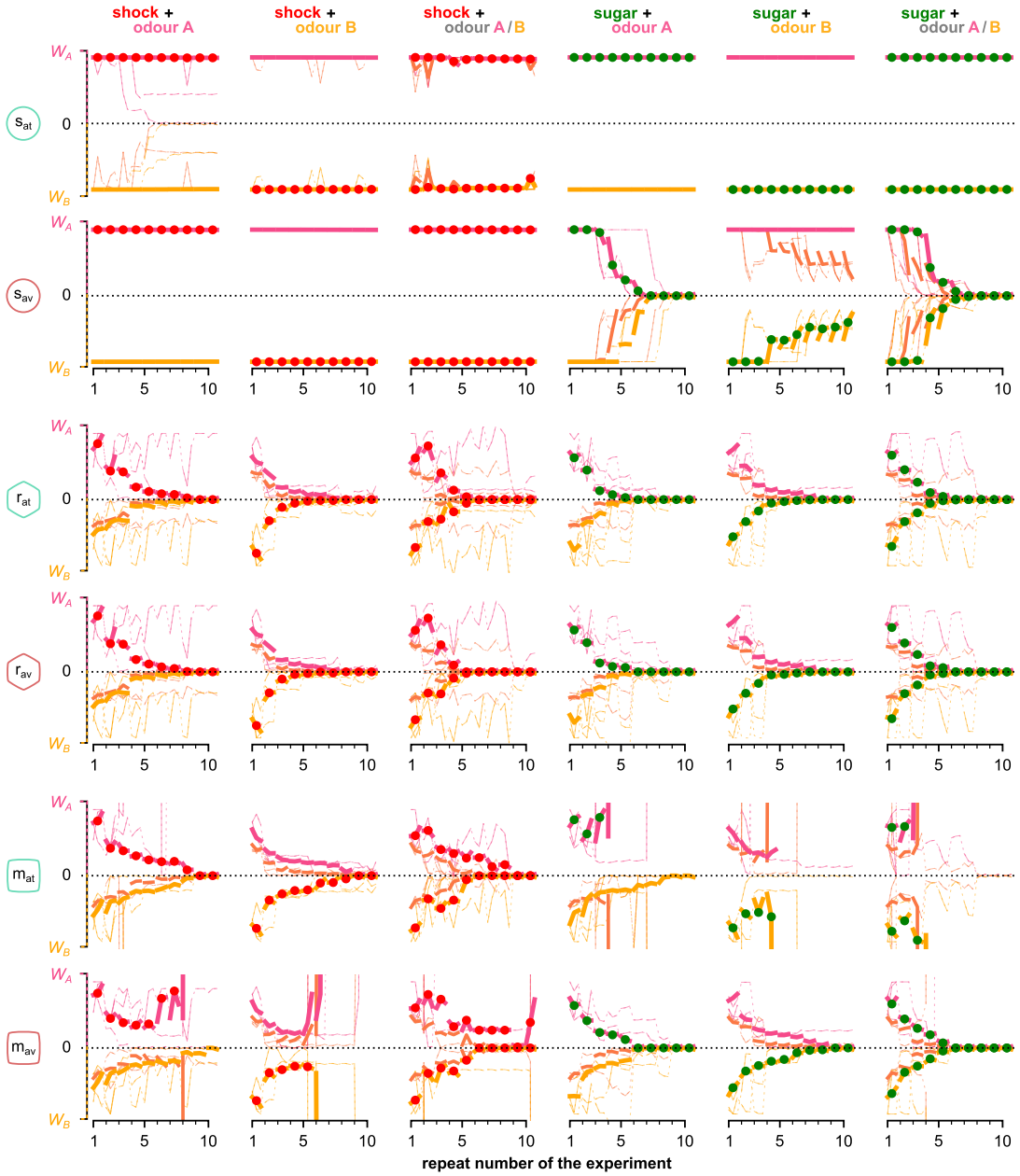

Figure E.19: The mean KC → MBON synaptic weights over the simulated flies that visited both odours and for each neuron, phase and repeat of the experiment. Different rows correspond to KC → MBON weights associated with a different type of *mushroom body output neuron* (MBON), from top to bottom: attraction- and avoidance driving susceptible, restrained and *long-term memory* (LTM) MBONs. Pink and yellow lines show synaptic weights associated with odours A and B respectively, and dashed lines show weights associated with both odours. Thin lines show 3 representative examples of synaptic weights in single simulated flies. Thick lines show the median of the synaptic weight over all the simulated flies that visited both odours.

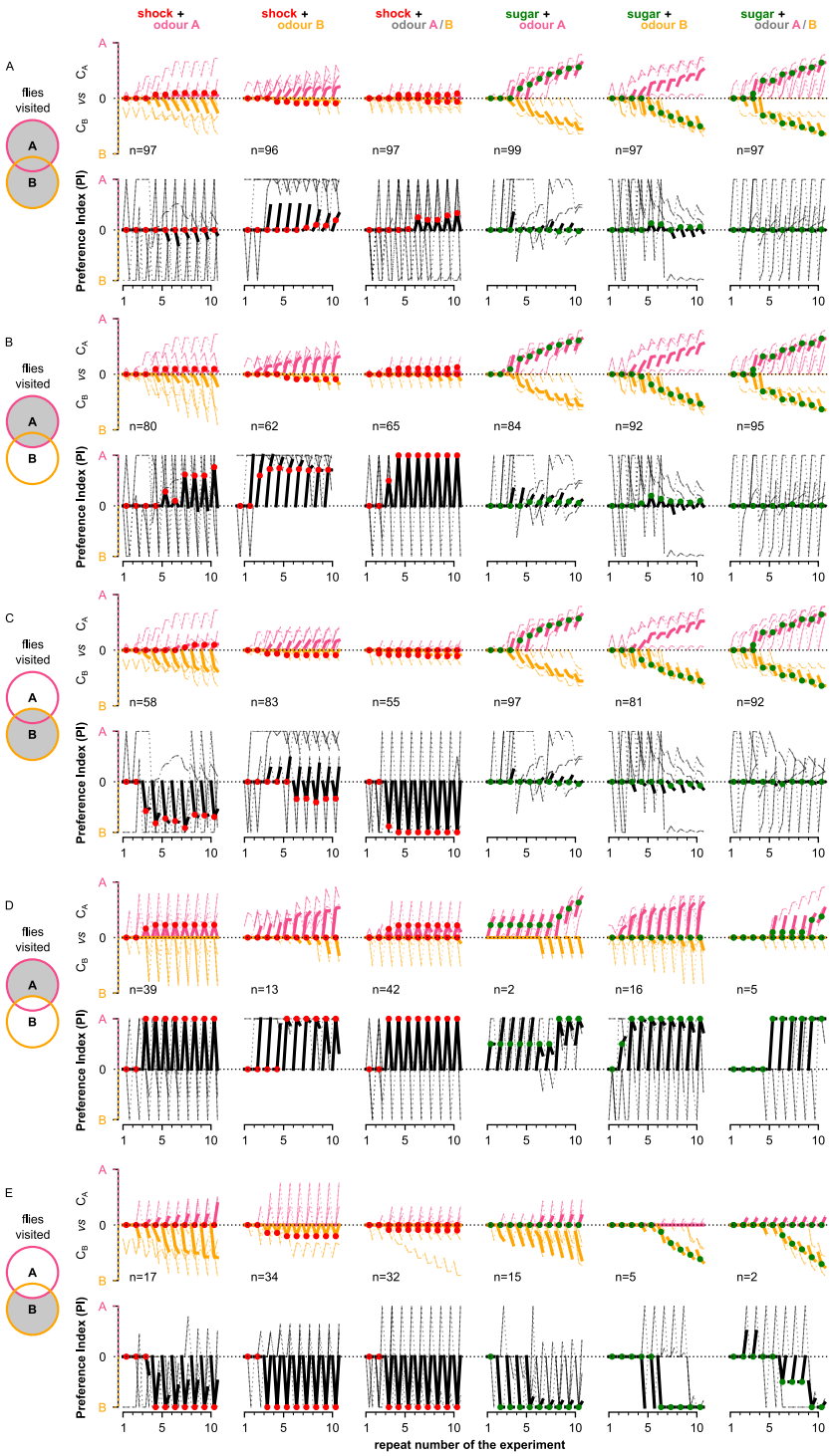

Figure E.20: Behavioural summary of simulated flies grouped by the areas that they visited: (A) at least one of the two odours, (B) odour A, (C) odour B, (D) only odour A and (E) only odour B. In each panel, columns show the different conditions and the population for each group. Top row: the normalised cumulative time spent exposed in odour A (pink lines) or odour B (yellow lines—note this line is reversed). For each repeat, 3 values are presented (averaged over all the pre-training, training and post-training time-steps respectively) where the values associated with the training phase are marked with red or green dots when punishment or reward has been delivered to that odour respectively. Thin lines show 3 representative samples of individual flies. Thick lines show the median over the simulated flies that visited both odours. Bottom row: the *preference index* (PI) to each odour extracted by the above cumulative times.

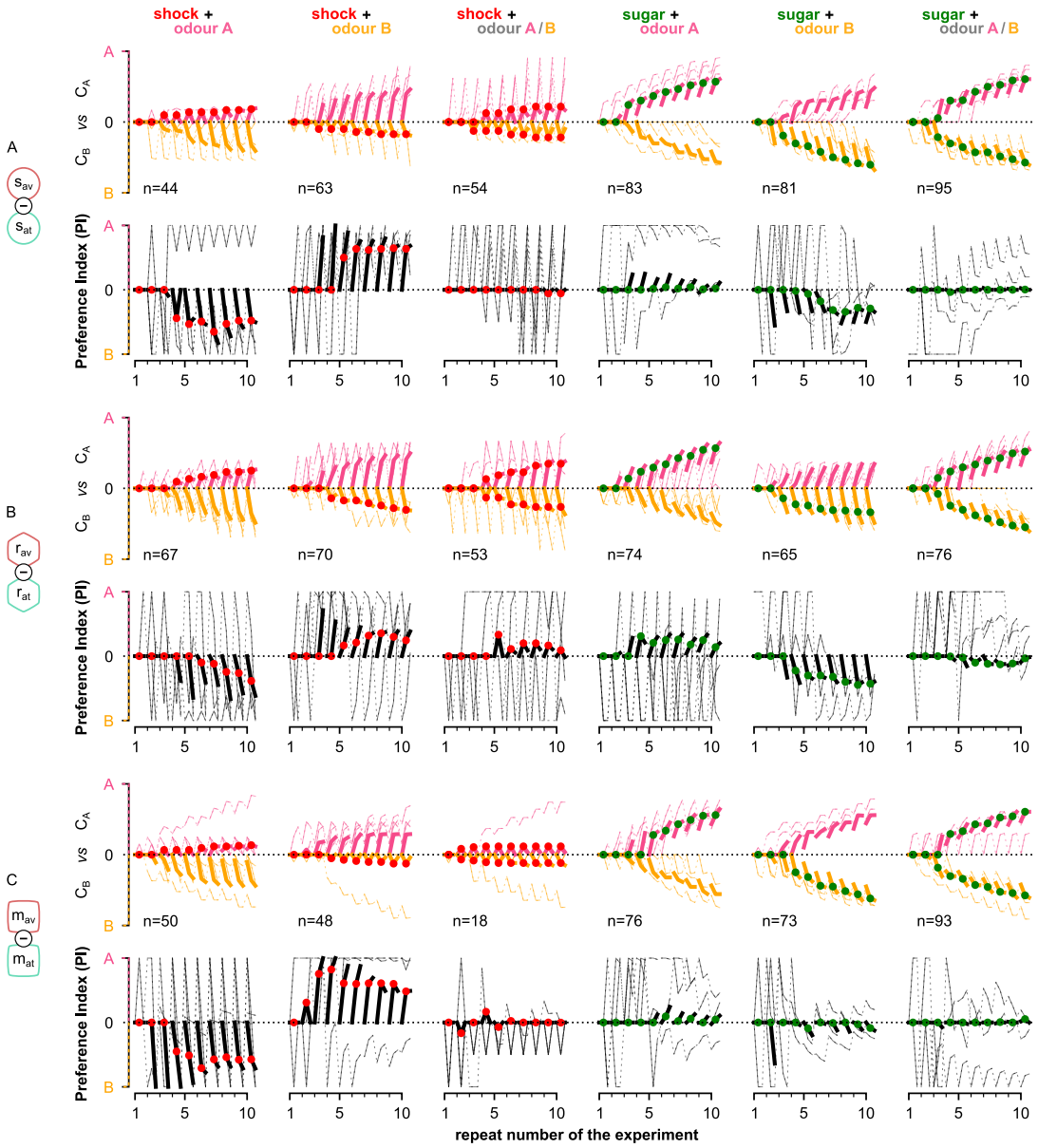

Figure E.21: Behavioural summary of simulated flies when controlled by: (A) the susceptible, (B) restrained or (C) *long-term memory* (LTM) *mushroom body output neurons* (MBON) separately. In each panel, columns show the different conditions and the population for each group. Top row: the normalised cumulative time spent exposed in odour A (pink lines) or odour B (yellow lines—note this line is reversed). For each repeat, 3 values are presented (averaged over all the pre-training, training and post-training time-steps respectively) where the values associated with the training phase are marked with red or green dots when punishment or reward has been delivered to that odour respectively. Thin lines show 3 representative samples of individual flies. Thick lines show the median over the simulated flies that visited both odours. Bottom row: the preference index (PI) to each odour extracted by the above cumulative times.

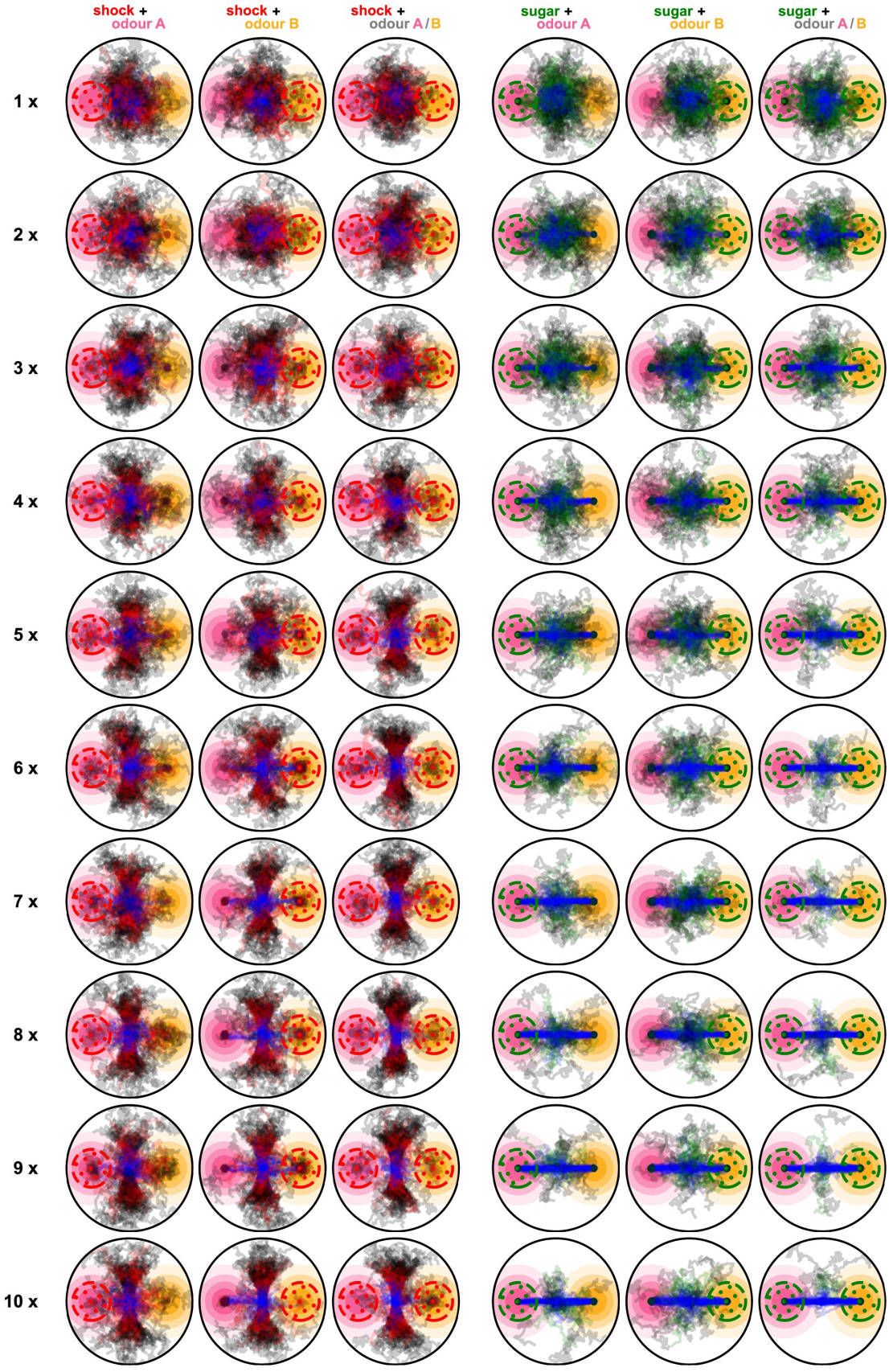

Figure E.22: Paths of the flies when using the *dopaminergic plasticity rule* (DPR) and during all 10 repeats of the experiment. Blue segments show the paths of the flies during the pre-training, red and green segments show the paths during the training phase (for punishment and reward conditions respectively), and black segments show the paths during the post-training phase. Rows are for the different repeats and columns are for the different conditions.

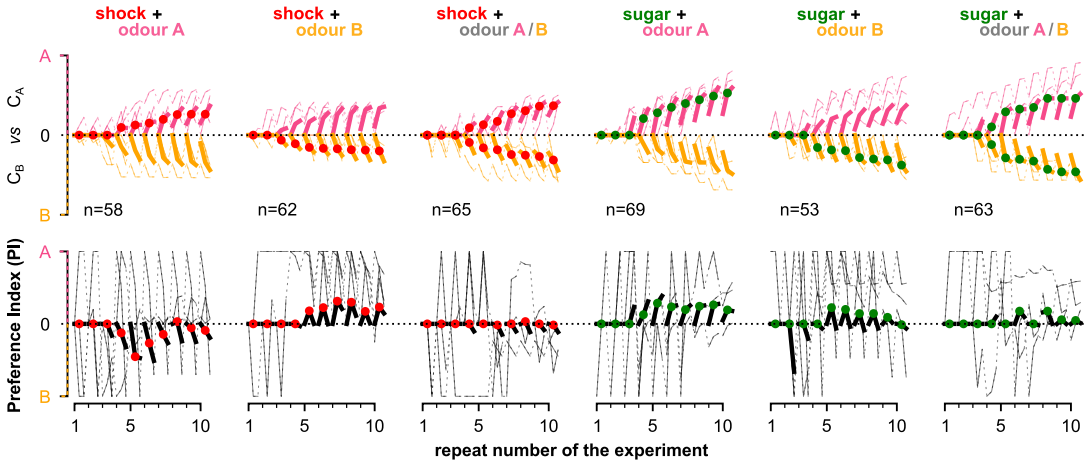

Figure E.23: Behavioural summary of a subset of simulated flies that visited both odours at any time during the 10 repeats and the plasticity rule of their neurons was replaced by the *reward prediction error* (RPE) plasticity rule. Columns show the different conditions and the population that was recorded visiting both odours. Top row: the normalised cumulative time spent exposed in odour A (pink lines) or odour B (yellow lines—note this line is reversed). For each repeat, 3 values are presented (averaged over all the pre-training, training and post-training time-steps respectively) where the values associated with the training phase are marked with red or green dots when punishment or reward has been delivered to that odour respectively. Thin lines show 3 representative samples of individual flies. Thick lines show the median over the simulated flies that visited both odours. Bottom row: the *preference index* (PI) to each odour extracted by the above cumulative times.

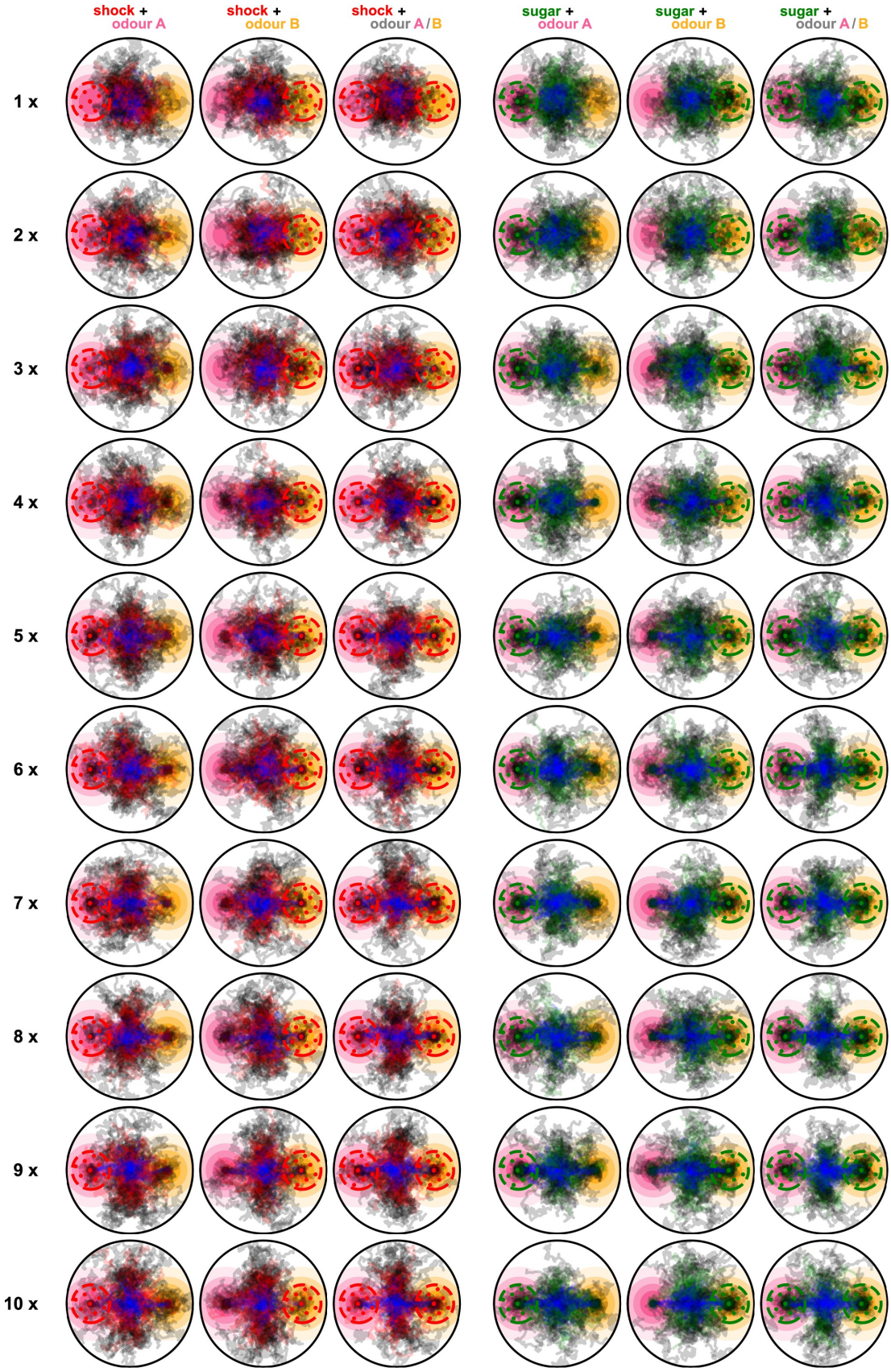

Figure E.24: Paths of 100 simulated flies when using the *reward prediction error* (RPE) plasticity rule and during all 10 repeats of the experiment. Blue segments show the paths of the flies during the pre-training, red and green segments show the paths during the training phase for punishment and reward conditions respectively, and black segments show the paths during the post-training phase. Rows are for the different repeats and columns are for the different conditions.

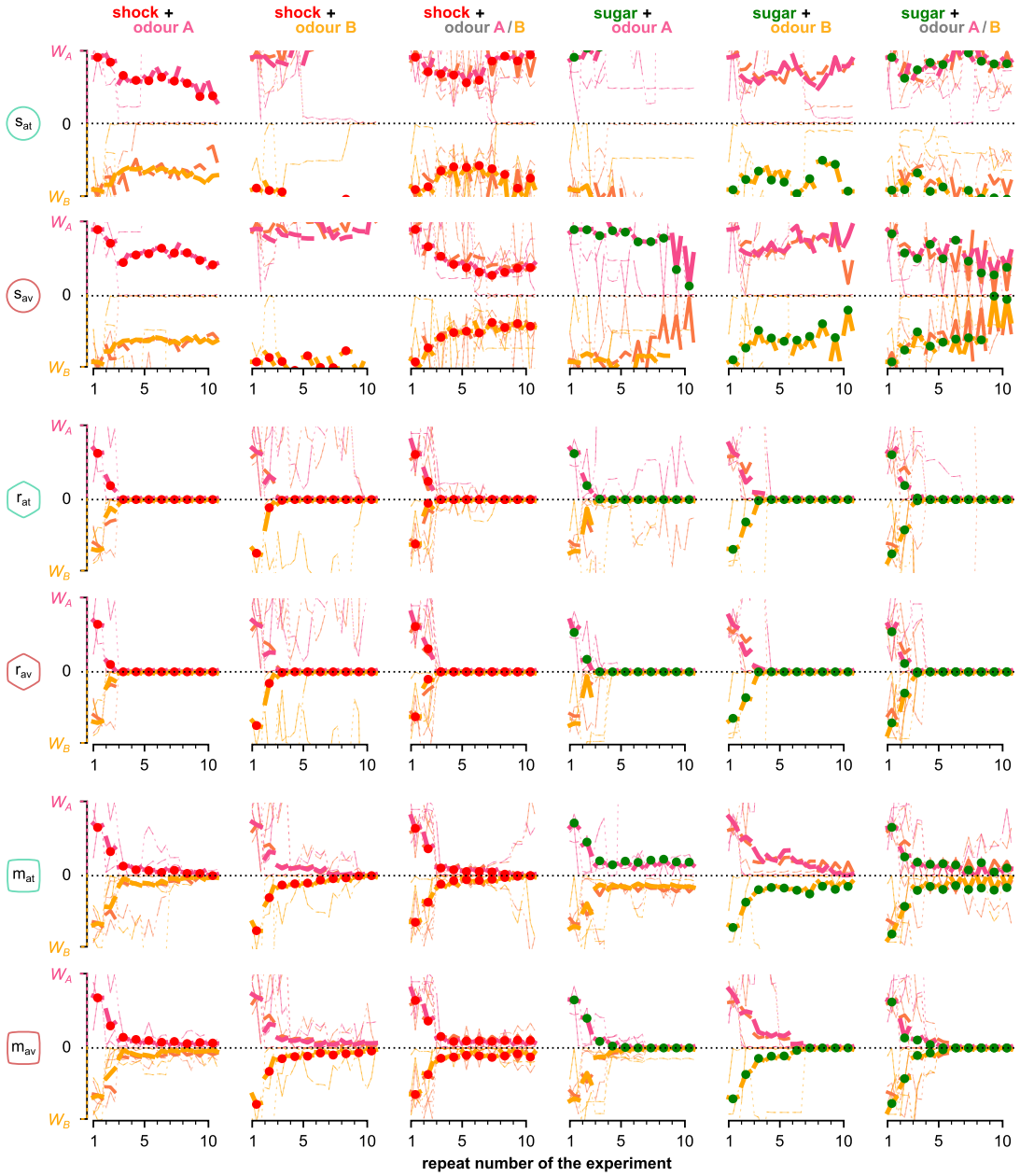

Figure E.25: The mean KC → MBON synaptic weights when using the *reward prediction error* (RPE) plasticity rule over the simulated flies that visited both odours and for each neuron, phase and repeat of the experiment. Different rows correspond to KC → MBON weights associated with a different type of *mushroom body output neuron* (MBON), from top to bottom: attraction- and avoidance driving susceptible, restrained and *long-term memory* (LTM) MBONs. Pink and yellow lines show synaptic weights associated with odours A and B respectively, and dashed lines show weights associated with both odours. Thin lines show 3 representative examples of synaptic weights in single simulated flies. Thick lines show the median of the synaptic weight over all the simulated flies that visited both odours.

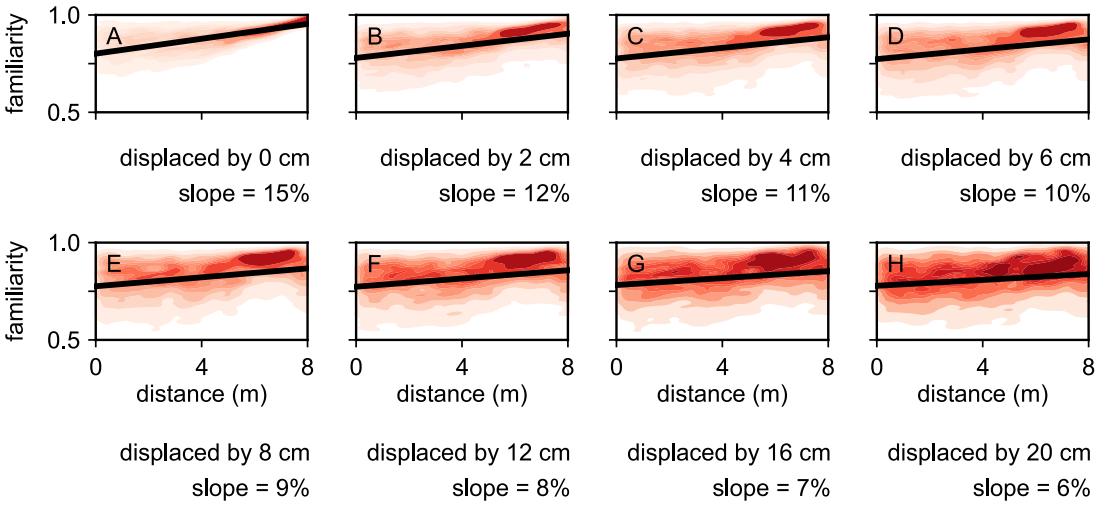

Figure E.26: The estimated normalised familiarity along the parallel displaced route when using the *incentive circuit* (IC), grouped by the displacing distance of the route. The contours show the distribution of the familiarity values along the distance. The samples of the distribution were used to fit a line with linear regression, showing that the slope of the familiarity along the routes decreases as the views are captured further away from the training route. (A) For $0\,\mathrm{cm}$ displacement (training route) the slope was $15\,\%$, (B) for $2\,\mathrm{cm}$ displacement the slope was $12\,\%$, (C) for $4\,\mathrm{cm}$ displacement the slope was $11\,\%$, (D) for $6\,\mathrm{cm}$ displacement the slope was $10\,\%$, (E) for $8\,\mathrm{cm}$ displacement the slope was $9\,\%$, (F) for $12\,\mathrm{cm}$ displacement the slope was $8\,\%$, (G) for $16\,\mathrm{cm}$ displacement the slope was $7\,\%$, and (H) for $20\,\mathrm{cm}$ displacement the slope was $6\,\%$.

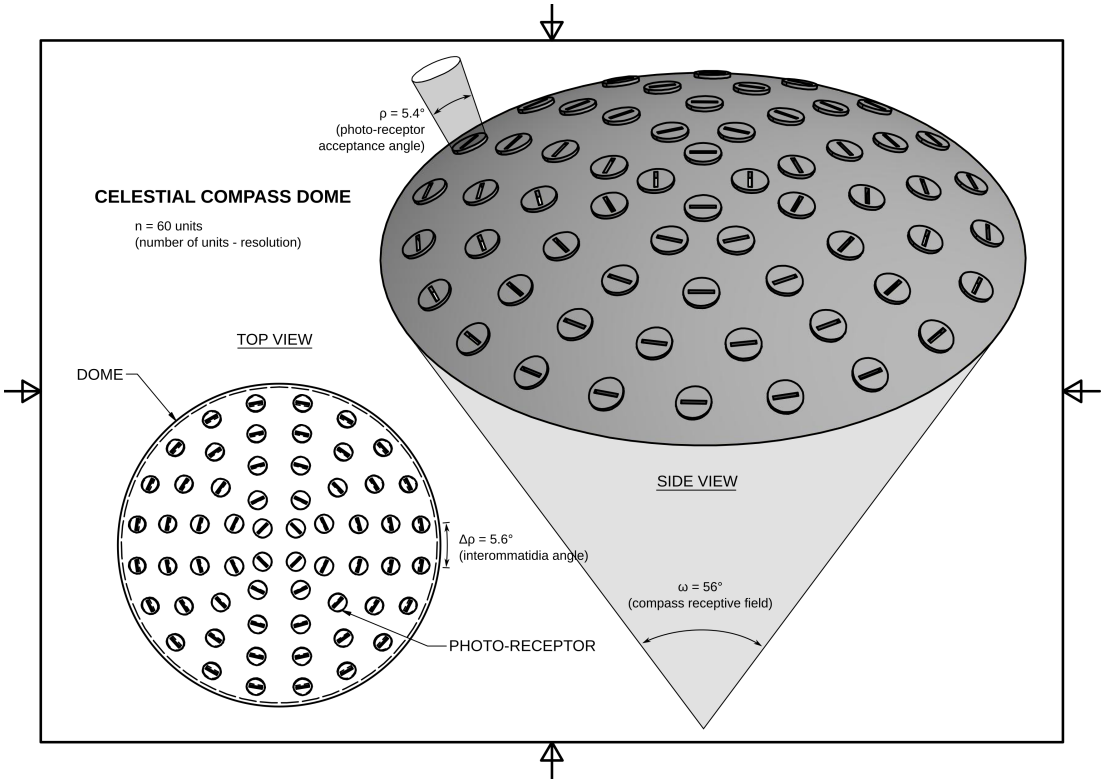

Figure E.27: Detailed view of the polarised-light compass layout. Top and side view of the sensor layout, including all the design parameters: $n = 60$ units; $\omega = 56°$ receptive field; $\rho = 5.4°$ acceptance angle for each ommatidium; $\Delta\rho = 5.6°$ interommatidia angle. The design was built using the Adobe Inventor tool and modified using Inkscape.

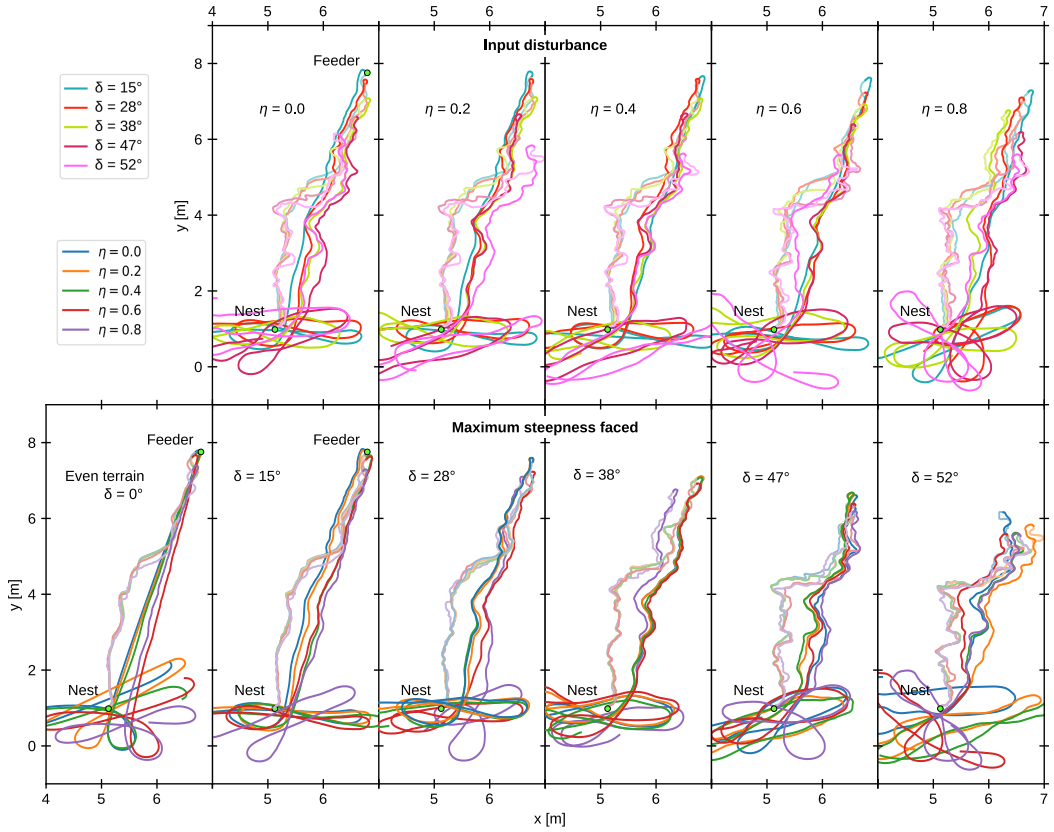

Figure E.28: Behavioural simulation for the path integration task—summarised results. Outward (away from the nest—bold lines) and inward paths (towards the nest—faded lines) of artificial ants that use the proposed compass to orient themselves in different disturbance and inclination levels. Each panel of the top row shows the route of the ant in different maximum surface steepness ($\delta \in \{15°, 28°, 38°, 47°, 52°\}$; different coloured lines in the same panel) which is associated with the respective altitude variance in the terrain shown in Fig. D.13C ($\alpha \in \{0.2\,\mathrm{m}, 0.4\,\mathrm{m}, 0.6\,\mathrm{m}, 0.8\,\mathrm{m}, 1.0\,\mathrm{m}\}$; Fig. D.13 shows results for $\alpha = 0.8\,\mathrm{m}$) for a specific sky-disturbance level (η). The panels of the bottom row show the routes of the ant in different sky-disturbance levels ($\eta \in \{0.0, 0.2, 0.4, 0.6, 0.8\}$; different coloured lines in the same panel) for a specific maximum inclination. The bottom further left panel shows the performance in case of no inclination at all (even terrain, $\delta = 0°$). The outward paths were adapted from Mangan and Webb (2012), and the inward paths are the outcome of the path integration network described in Stone, Webb, et al. (2017) using this approach to extract the TB1-neurons' responses.

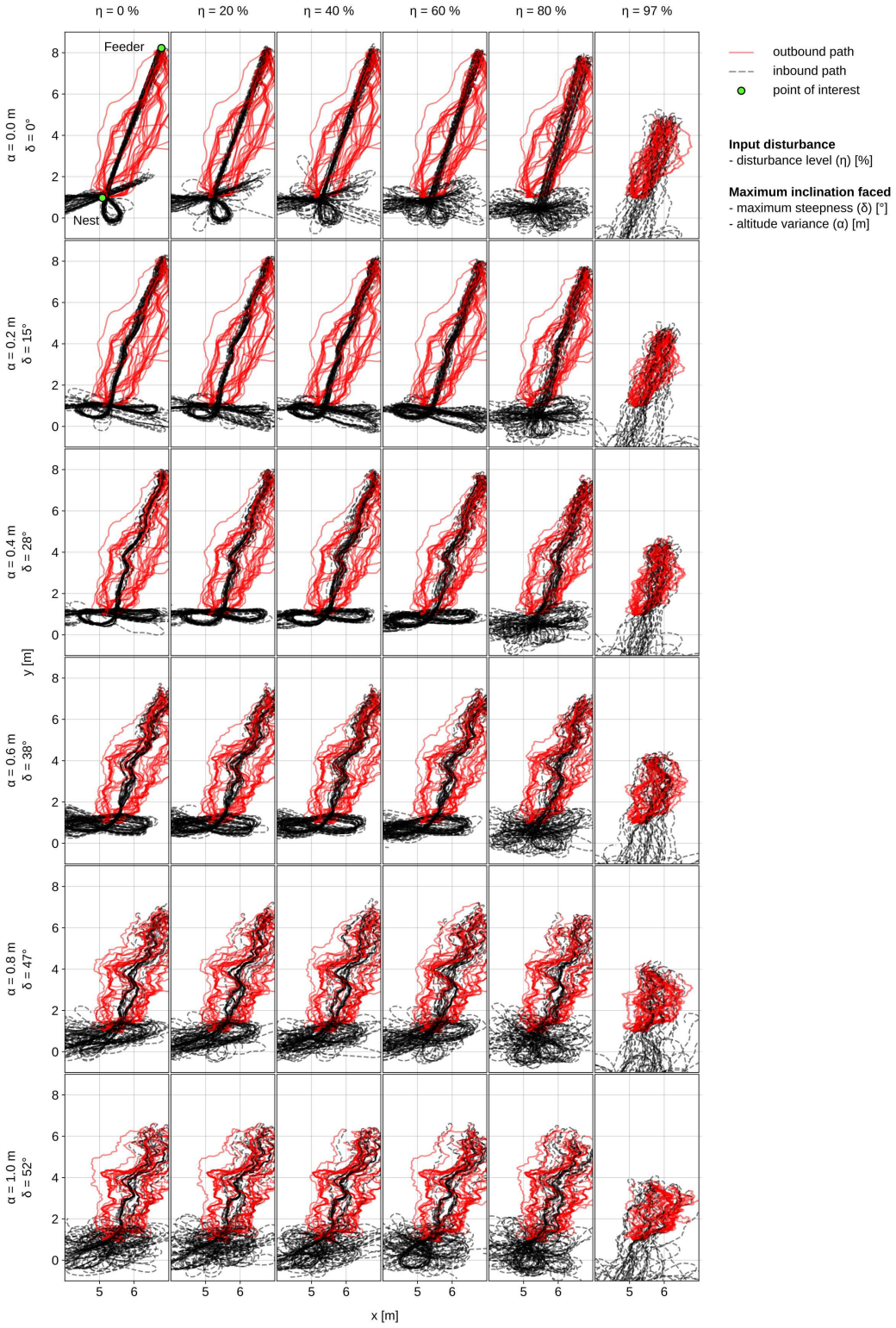

Figure E.29: Behavioural simulation for the path integration task—detailed results. Grid of panels showing the performance of the model when used for path integration for different sky-disturbance levels (different columns for different sky-disturbance levels) and maximum steepness of the terrain (different rows for different inclinations). Lines in each panel show the outward (away from the nest—red solid lines) and inward paths (towards the nest—black dashed lines) of 133 ants using the proposed compass to orient themselves. The outward paths were adapted from Mangan and Webb (2012), and the inward paths are the outcome of the path integration network described in Stone, Webb, et al. (2017) using this approach to extract the TB1-neurons' responses.

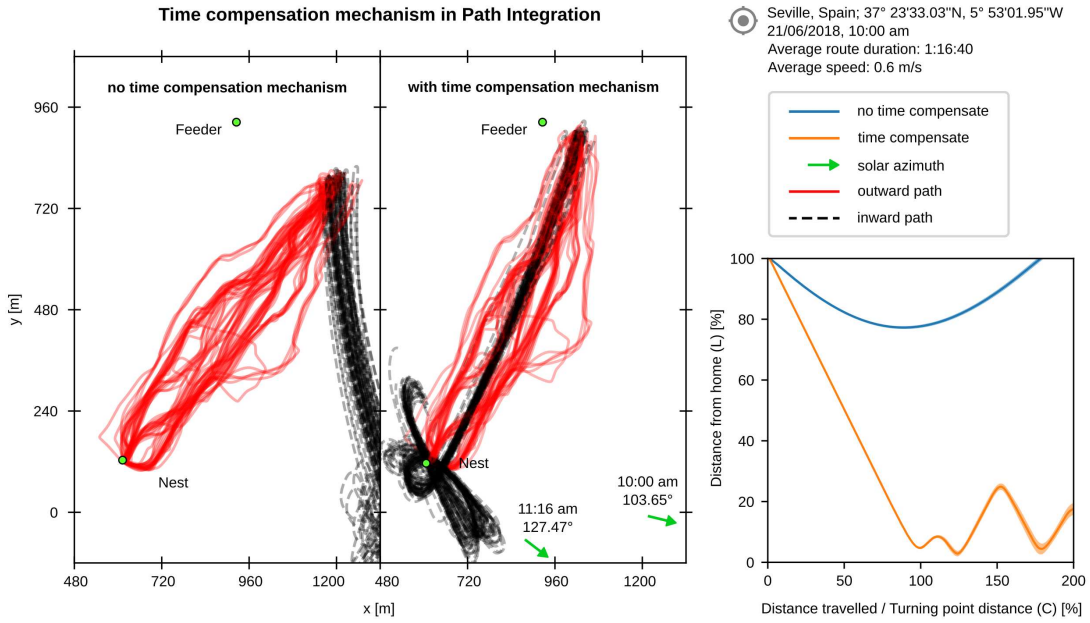

Figure E.30: Behavioural simulation for the path integration task—compensating for time. Outward (away from the nest—red solid lines) and inward paths (towards the nest—black dashed lines) of artificial ants that use the proposed compass to orient themselves during their visit to distant food sources of around 700 m away resulting in long runs. These runs take on average 1 hour and 16 minutes to complete and the solar azimuth changes for 23.82°, showing the effect of the time compensation mechanism. The outward paths were adapted from Mangan and Webb (2012), and the inward paths are the outcome of the path integration network described in Stone, Webb, et al. (2017) using this approach to extract the TB1-neurons' responses.

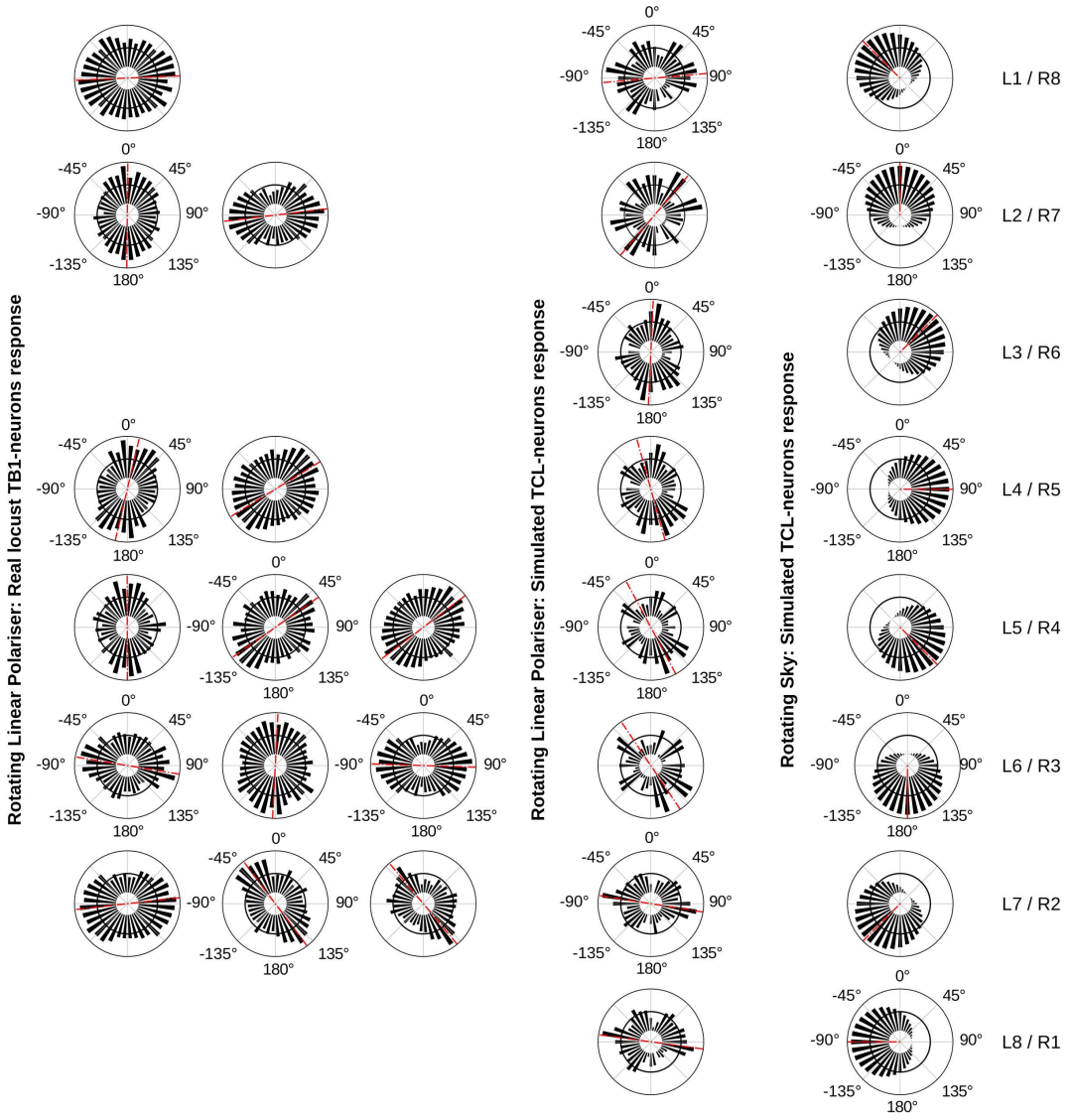

Figure E.31: Real and simulated response of compass neurons—detailed results. E-vector orientation resulting in maximum excitation, $\phi_{\max}$, of the real (desert locust) and simulated (this model) neurons in the *central complex* (CX). Each panel shows the response (black bars) and $\phi_{\max}$ (red line) of a specific TB1- (TCL-) neuron in the real (artificial) CX. The different rows of the panel grid show the location of the neurons in the *protocerebral bridge* (PB). The first three columns show the response of different TB1-neurons in the desert locust brain supplied by Stanley Heinze (Heinze and Homberg, 2007) and organised in rows with respect to their location; rows 3 and 7 are empty due to a lack of data; different panels in the same row show different neurons in the same region. The fourth column shows the response of the artificial TCL-neurons to a similar stimulus to the one used in the first three columns. The last column shows the response of the same TCL-neurons when the artificial *dorsal rim area* (DRA) is exposed to a natural skylight extracted by a simulated rotating sky.

BIBLIOGRAPHY

Aaron, C. (2006). "Graph-based normalization and whitening for non-linear data analysis." In: *Neural Networks* 19.6-7, pp. 864–876. ISSN: 0893-6080. DOI: 10.1016/j.neunet.2006.05.022.

Achacoso, T. B. and W. S. Yamamoto (1990). "Artificial ethology and computational neuroethology: a scientific discipline and its subset by sharpening and extending the definition of artificial intelligence." In: *Perspectives in Biology and Medicine* 33.3, pp. 379–390. ISSN: 1529-8795. DOI: 10.1353/pbm.1990.0020.

Adel, M. and L. C. Griffith (2021). "The role of dopamine in associative learning in Drosophila: an updated unified model." In: *Neuroscience Bulletin*, pp. 1–22. ISSN: 1673-7067. DOI: 10.1007/s12264-021-00665-0.

Åkesson, S. and A. Hedenström (2007). "How migrants get there: migratory performance and orientation." In: *AIBS Bulletin* 57.2, pp. 123–133.

Ardin, P., M. Mangan, A. Wystrach, and B. Webb (2015). "How variation in head pitch could affect image matching algorithms for ant navigation." In: *Journal of Comparative Physiology A* 201.6, pp. 585–597.

Ardin, P., F. Peng, M. Mangan, K. Lagogiannis, and B. Webb (2016). "Using an insect mushroom body circuit to encode route memory in complex natural environments." In: *PLOS Computational Biology* 12.2, e1004683. ISSN: 1553-734X. DOI: 10.1371/journal.pcbi.1004683.

Arena, P., L. Patané, V. Stornanti, P. S. Termini, B. Zäpf, and R. Strauss (2013). "Modeling the insect mushroom bodies: application to a delayed match-to-sample task." In: *Neural Networks* 41, pp. 202–211. ISSN: 0893-6080. DOI: 10.1016/j.neunet.2012.11.013.

Aso, Y., D. Hattori, et al. (2014). "The neuronal architecture of the mushroom body provides a logic for associative learning." In: *eLife* 3, e04577. DOI: 10.7554/elife.04577.

Aso, Y., A. Herb, M. Ogueta, I. Siwanowicz, T. Templier, A. B. Friedrich, K. Ito, H. Scholz, and H. Tanimoto (2012). "Three dopamine pathways induce aversive odor memories with different stability." In: *PLoS Genetics* 8.7, e1002768. ISSN: 15537390. DOI: 10.1371/journal.pgen.1002768.

Aso, Y., R. P. Ray, et al. (2019). "Nitric oxide acts as a cotransmitter in a subset of dopaminergic neurons to diversify memory dynamics." In: *eLife* 8, e49257. DOI: 10.7554/elife.49257.

Aso, Y. and G. M. Rubin (2016). "Dopaminergic neurons write and update memories with cell-type-specific rules." In: *eLife* 5, e16135. DOI: 10.7554/elife.16135.

Aso, Y., D. Sitaraman, et al. (2014). "Mushroom body output neurons encode valence and guide memory-based action selection in Drosophila." In: *eLife* 3, e04580. DOI: 10.7554/elife.04580.

Aso, Y., I. Siwanowicz, L. Bräcker, K. Ito, T. Kitamoto, and H. Tanimoto (2010). "Specific dopaminergic neurons for the formation of labile aversive memory." In: *Current Biology* 20.16, pp. 1445–1451. ISSN: 0960-9822. DOI: 10.1016/j.cub.2010.06.048.

Baddeley, B., P. Graham, P. Husbands, and A. Philippides (2012). "A model of ant route navigation driven by scene familiarity." In: *PLoS Computational Biology* 8.1, e1002336. ISSN: 1553-734X. DOI: 10.1371/journal.pcbi.1002336.

Baker, R. R. and J. G. Mather (1982). "Magnetic compass sense in the large yellow underwing moth, Noctua pronuba L." In: *Animal Behaviour* 30.2, pp. 543–548.

Balkenius, A., A. Kelber, and C. Balkenius (2006). "From animals to animats 9, 9th international conference on simulation of adaptive behavior." In: *Lecture Notes in Computer Science*, pp. 422–433. ISSN: 0302-9743. DOI: 10.1007/11840541_35.

Bargmann, C. I. (2006). "Comparative chemosensation from receptors to ecology." In: *Nature* 444.7117, pp. 295–301. ISSN: 0028-0836. DOI: 10.1038/nature05402.

Barto, A. G. (1997). "Reinforcement learning." In: *Neural systems for control*, pp. 7–30. DOI: 10.1016/B978-012526430-3/50003-9.

Batschelet, E. (1981). *Circular statistics in biology*. Vol. 111. Academic press London.

Bazhenov, M., R. Huerta, and B. H. Smith (2013). "A computational framework for understanding decision making through integration of basic learning rules." In: *The Journal of Neuroscience* 33.13, pp. 5686–5697. ISSN: 0270-6474. DOI: 10.1523/jneurosci.4145-12.2013.

Bear, D. M., J.-M. Lassance, H. E. Hoekstra, and S. R. Datta (2016). "The evolving neural and genetic architecture of vertebrate olfaction." In: *Current Biology* 26.20, R1039–R1049. ISSN: 0960-9822. DOI: 10.1016/j.cub.2016.09.011.

Bech, M., U. Homberg, and K. Pfeiffer (2014). "Receptive fields of locust brain neurons are matched to polarization patterns of the sky." In: *Current Biology* 24.18, pp. 2124–2129.

Bell, A. J. and T. J. Sejnowski (1995). "An information-maximization approach to blind separation and blind deconvolution." In: *Neural Computation* 7.6, pp. 1129–1159. ISSN: 0899-7667. DOI: 10.1162/neco.1995.7.6.1129.

Belle, J. S. de and M. Heisenberg (1994). "Associative odor learning in Drosophila abolished by chemical ablation of mushroom bodies." In: *Science* 263.5147, pp. 692–695.

Bellman, R. (1966). "Dynamic programming." In: *Science* 153.3731, pp. 34–37. ISSN: 00255572. DOI: 10.2307/3610676.

Bennett, J. E. M., A. Philippides, and T. Nowotny (2021). "Learning with reinforcement prediction errors in a model of the Drosophila mushroom body." In: *Nature Communications* 12.1, p. 2569. DOI: 10.1038/s41467-021-22592-4.

Berry, J. A., A. Phan, and R. L. Davis (2018). "Dopamine neurons mediate learning and forgetting through bidirectional modulation of a memory trace." In: *Cell Reports* 25.3, 651–662.e5. ISSN: 2211-1247. DOI: 10.1016/j.celrep.2018.09.051.

Bhandawat, V., S. R. Olsen, N. W. Gouwens, M. L. Schlief, and R. I. Wilson (2007). "Sensory processing in the Drosophila antennal lobe increases reliability and separability of ensemble odor representations." In: *Nature Neuroscience* 10.11, pp. 1474–1482. ISSN: 1097-6256. DOI: 10.1038/nn1976.

Bilz, F., B. R. Geurten, C. E. Hancock, A. Widmann, and A. Fiala (2020). "Visualization of a distributed synaptic memory code in the Drosophila brain." In: *Neuron* 106.6, 963–976.e4. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2020.03.010.

Bjorndal, M., A. Caprara, P. Cowling, F. D. Croce, H. Lourenço, F. Malucelli, A. Orman, D. Pisinger, C. Rego, and J. Salazar (1995). "Some thoughts on combinatorial optimisation." In: *European Journal of Operational Research* 83.2, pp. 253–270. ISSN: 0377-2217. DOI: 10.1016/0377-2217(95)00005-b.

Blaser, R. E., P. A. Couvillon, and M. E. Bitterman (2008). "Within-subjects experiments on blocking and facilitation in honeybees (Apis mellifera)." In: *Journal of Comparative Psychology* 122.4, p. 373. DOI: 10.1037/a0012623.WITHIN-SUBJECTS.

Blum, M. and T. Labhart (2000). "Photoreceptor visual fields, ommatidial array, and receptor axon projections in the polarisation-sensitive dorsal rim area of the cricket compound eye." In: *Journal of Comparative Physiology A* 186.2, pp. 119–128.

Borst, A. (1983). "Computation of olfactory signals in Drosophila melanogaster." In: *Journal of Comparative Physiology A* 152.3, pp. 373–383. ISSN: 14321351. DOI: 10.1007/BF00606242.

Borst, A. (2009). "Drosophila's view on insect vision." In: *Current Biology* 19.1, R36–R47. ISSN: 0960-9822. DOI: 10.1016/j.cub.2008.11.001.

Bozinovski, S. (1982). "A self-learning system using secondary reinforcement." In: *Cybernetics and Systems Research*, pp. 397–402.

Brembs, B. and M. Heisenberg (2001). "Conditioning with compound stimuli in Drosophila melanogaster in the flight simulator." In: *The Journal of experimental biology* 204.Pt 16, pp. 2849–59. ISSN: 0022-0949.

Brembs, B., F. D. Lorenzetti, F. D. Reyes, D. A. Baxter, and J. H. Byrne (2002). "Operant reward learning in Aplysia: neuronal correlates and mechanisms." In: *Science* 296.5573, pp. 1706–1709.

Brembs, B. and J. Wiener (2006). "Context and occasion setting in Drosophila visual learning." In: *Learning and Memory* 13.5, pp. 618–628. ISSN: 10720502. DOI: 10.1101/lm.318606.

Brockman, G., V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba (2016). "Openai gym." In: *arXiv preprint arXiv:1606.01540*.

Buehlmann, C., B. Wozniak, R. Goulard, B. Webb, P. Graham, and J. E. Niven (2020). "Mushroom bodies are eequired for learned visual navigation, but not for innate visual behavior, in ants." In: *Current Biology*. ISSN: 0960-9822. DOI: 10.1016/j.cub.2020.07.013.

Burke, C. J., W. Huetteroth, D. Owald, E. Perisse, M. J. Krashes, G. Das, D. Gohl, M. Silies, S. Certel, and S. Waddell (2012). "Layered reward signalling through octopamine and dopamine in Drosophila." In: *Nature* 492.7429, pp. 433–437. ISSN: 0028-0836. DOI: 10.1038/nature11614.

Busch, S., M. Selcho, K. Ito, and H. Tanimoto (2009). "A map of octopaminergic neurons in the Drosophila brain." In: *Journal of Comparative Neurology* 513.6, pp. 643–667. ISSN: 10969861. DOI: 10.1002/cne.21966.

Busto, G. U., I. Cervantes-Sandoval, and R. L. Davis (2010). "Olfactory Learning in Drosophila." In: *Physiology* 25.6, pp. 338–346. ISSN: 1548-9213. DOI: 10.1152/physiol.00026.2010.

Caron, S. J. C., V. Ruta, L. F. Abbott, and R. Axel (2013). "Random convergence of olfactory inputs in the Drosophila mushroom body." In: *Nature* 497.7447, pp. 113–117. ISSN: 0028-0836. DOI: 10.1038/nature12063.

Cervantes-Sandoval, I., A. Phan, M. Chakraborty, and R. L. Davis (2017). "Reciprocal synapses between mushroom body and dopamine neurons form a positive feedback loop required for learning." In: *eLife* 6, e23789. DOI: 10.7554/elife.23789.

Chahl, J. and A. Mizutani (2012). "Biomimetic attitude and orientation sensors." In: *IEEE Sensors Journal* 12.2, pp. 289–297.

Chakraborty, S. D., H. Chang, B. S. Hansson, and S. Sachse (2022). "Higher-order olfactory neurons in the lateral horn support odor valence and odor identity coding in Drosophila." In: *eLife* 11, e74637. DOI: 10.7554/elife.74637.

Chakraborty, S. D. and S. Sachse (2021). "Olfactory processing in the lateral horn of Drosophila." In: *Cell and Tissue Research* 383.1, pp. 113–123. ISSN: 0302-766X. DOI: 10.1007/s00441-020-03392-6.

Chapman, J. W., D. R. Reynolds, H. Mouritsen, J. K. Hill, J. R. Riley, D. Sivell, A. D. Smith, and I. P. Woiwod (2008). "Wind selection and drift compensation optimize migratory pathways in a high-flying moth." In: *Current Biology* 18.7, pp. 514–518.

Cheung, A. and R. Vickerstaff (2010). "Finding the way with a noisy brain." In: *PLoS computational biology* 6.11, e1000992.

Chiel, H. and R. Beer (2009). "Computational neuroethology." In: *Encyclopedia of Neuroscience*, pp. 23–28. DOI: 10.1016/b978-008045046-9.01953-7.

Chu, J., K. Zhao, Q. Zhang, and T. Wang (2008). "Construction and performance test of a novel polarization sensor for navigation." In: *Sensors and Actuators A: Physical* 148.1, pp. 75–82.

Claridge-Chang, A., R. D. Roorda, E. Vrontou, L. Sjulson, H. Li, J. Hirsh, and G. Miesenböck (2009). "Writing memories with light-addressable reinforcement circuitry." In: *Cell* 139.2, pp. 405–415. ISSN: 0092-8674. DOI: 10.1016/j.cell.2009.08.034.

Cleland, T. A. (1996). "Inhibitory glutamate receptor channels." In: *Molecular Neurobiology* 13.2, pp. 97–136. ISSN: 0893-7648. DOI: 10.1007/bf02740637.

Clement, L., S. Schwarz, and A. Wystrach (2022). "An intrinsic oscillator underlies visual navigation in ants." In: *bioRxiv*, p. 2022.04.22.489150. DOI: 10.1101/2022.04.22.489150.

Cohn, R., I. Morantte, and V. Ruta (2015). "Coordinated and compartmentalized neuromodulation shapes sensory processing in Drosophila." In: *Cell* 163.7, pp. 1742–1755. ISSN: 0092-8674. DOI: 10.1016/j.cell.2015.11.019.

Collett, T. (1972). "Visual neurones in the anterior optic tract of the privet hawk moth." In: *Journal of comparative physiology* 78.4, pp. 396–433. ISSN: 0340-7594. DOI: 10.1007/bf01417943.

Collett, T. S. and M. Collett (2002). "Memory use in insect visual navigation." In: *Nature Reviews Neuroscience* 3.7, pp. 542–552. ISSN: 14710048. DOI: 10.1038/nrn872.

Collett, T. S., P. Graham, and R. A. Harris (2007). "Novel landmark-guided routes in ants." In: *Journal of Experimental Biology* 210.12, pp. 2025–2032. ISSN: 0022-0949. DOI: 10.1242/jeb.000315.

Colomb, J., L. Kaiser, M. A. Chabaud, and T. Preat (2009). "Parametric and genetic analysis of Drosophila appetitive long-term memory and sugar motivation." In: *Genes, Brain and Behavior* 8.4, pp. 407–415. ISSN: 1601-183X. DOI: 10.1111/j.1601-183x.2009.00482.x.

Coulson, K. L. (1988). *Polarization and Intensity of Light in the Atmosphere*. A Deepak Pub.

Couvillon, P. A. and M. E. Bitterman (1988). "Compound-component and conditional discrimination of colors and odors by honeybees: further tests of a continuity model." In: *Animal Learning & Behavior* 16.1, pp. 67–74. ISSN: 00904996. DOI: 10.3758/BF03209045.

Couvillon, P. A., A. C. Campos, T. D. Bass, and M. E. Bitterman (2001). "Intermodal blocking in honeybees." In: *Quarterly Journal of Experimental Psychology Section B: Comparative and Physiological Psychology* 54.4, pp. 369–381. ISSN: 02724995. DOI: 10.1080/02724990143000117.

Cruse, H. and R. Wehner (2011). "No need for a cognitive map: decentralized memory for insect navigation." In: *PLoS Computational Biology* 7.3, e1002009. ISSN: 1553734X. DOI: 10.1371/journal.pcbi.1002009.

Cruz, T., T. Fujiwara, N. Varela, F. Mohammad, A. Claridge-Chang, and M. E. Chiappe (2019). "Motor context coordinates visually guided walking in Drosophila." In: *bioRxiv*, p. 572792. DOI: 10.1101/572792.

Dacke, M., E. Baird, M. Byrne, C. H. Scholtz, and E. J. Warrant (2013). "Dung beetles use the Milky Way for orientation." In: *Current Biology* 23.4, pp. 298–300.

Dacke, M., B. el Jundi, J. Smolka, M. Byrne, and E. Baird (2014). "The role of the sun in the celestial compass of dung beetles." In: *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 369.1636, p. 20130036.

Dale, H. (1935). "Pharmacology and nerve-endings (Walter Ernest Dixon memorial lecture): (section of therapeutics and pharmacology)." In: *Proceedings of the Royal Society of Medicine* 28.3, pp. 319–32. ISSN: 0035-9157.

Dalgleish, T. (2004). "The emotional brain." In: *Nature Reviews Neuroscience* 5.7, pp. 583–589. ISSN: 1471-003X. DOI: 10.1038/nrn1432.

DasGupta, S. and S. Waddell (2008). "Learned odor discrimination in Drosophila without combinatorial odor maps in the antennal lobe." In: *Current Biology* 18.21, pp. 1668–1674. ISSN: 09609822. DOI: 10.1016/j.cub.2008.08.071.

Davis, R. L. (1993). "Mushroom bodies and drosophila learning." In: *Neuron* 11.1, pp. 1–14. ISSN: 0896-6273. DOI: 10.1016/0896-6273(93)90266-t.

Deisig, N., H. Lachnit, M. Giurfa, and F. Hellstern (2001). "Configural olfactory learning in honeybees: negative and positive patterning discrimination." In: *Learning and Memory* 8.2, pp. 70–78. ISSN: 10720502. DOI: 10.1101/lm.8.2.70.

Delahunt, C. B., J. A. Riffell, and J. N. Kutz (2018). "Biological mechanisms for learning: a computational model of olfactory learning in the Manduca sexta moth, with applications to neural nets." In: *Frontiers in Computational Neuroscience* 12, p. 102. ISSN: 1662-5188. DOI: 10.3389/fncom.2018.00102.

DeVoe, R. D., W. Kaiser, J. Ohm, and L. S. Stone (1982). "Horizontal movement detectors of honeybees: directionally-selective visual neurons in the lobula and brain." In: *Journal of comparative physiology* 147.2, pp. 155–170. ISSN: 0340-7594. DOI: 10.1007/bf00609840.

Dewar, A. D., A. Wystrach, P. Graham, and A. Philippides (2015). "Navigation-specific neural coding in the visual system of Drosophila." In: *Biosystems* 136, pp. 120–127. ISSN: 0303-2647. DOI: 10.1016/j.biosystems.2015.07.008.

Dubnau, J., L. Grady, T. Kitamoto, and T. Tully (2001). "Disruption of neurotransmission in Drosophila mushroom body blocks retrieval but not acquisition of memory." In: *Nature* 411.6836, pp. 476–480. ISSN: 0028-0836. DOI: 10.1038/35078077.

Dubnau, J. and T. Tully (2001). "Functional anatomy: from molecule to memory." In: *Current Biology* 11.6, pp. 240–243.

Dudai, Y. (1977). "Properties of learning and memory in Drosophila melanogaster." In: *Journal of Comparative Physiology A* 114.1, pp. 69–89. ISSN: 03407594. DOI: 10.1007/BF00656809.

Dupeyroux, J., J. Diperi, M. Boyron, S. Viollet, and J. Serres (2017). "A novel insect-inspired optical compass sensor for a hexapod walking robot." In: *IROS 2017-IEEE/RSJ International Conference on Intelligent Robots and Systems*.

Dupeyroux, J., J. R. Serres, and S. Viollet (2019). "AntBot: A six-legged walking robot able to home like desert ants in outdoor environments." In: *Science Robotics* 4.27, eaau0307.

Dylla, K. V., G. Raiser, C. G. Galizia, and P. Szyszka (2017). "Trace conditioning in Drosophila induces associative plasticity in mushroom body Kenyon cells and

dopaminergic neurons." In: *Frontiers in Neural Circuits* 11, p. 42. DOI: 10.3389/fncir.2017.00042.
Ehmer, B. and W. Gronenberg (2002). "Segregation of visual input to the mushroom bodies in the honeybee (Apis mellifera)." In: *Journal of Comparative Neurology* 451.4, pp. 362–373. ISSN: 1096-9861. DOI: 10.1002/cne.10355.
Eichler, K. et al. (Aug. 2017). "The complete connectome of a learning and memory centre in an insect brain." In: *Nature* 548.7666, pp. 175–182. ISSN: 0028-0836. DOI: 10.1038/nature23455.
El Jundi, B., K. Pfeiffer, S. Heinze, and U. Homberg (2014). "Integration of polarization and chromatic cues in the insect sky compass." In: *Journal of Comparative Physiology A* 200.6, pp. 575–589.
Eschbach, C. et al. (2020). "Recurrent architecture for adaptive regulation of learning in the insect brain." In: *Nature Neuroscience* 23.4, pp. 544–555. ISSN: 1097-6256. DOI: 10.1038/s41593-020-0607-9.
Etheredge, J. A., S. M. Perez, O. R. Taylor, and R. Jander (1999). "Monarch butterflies (Danaus plexippus L.) use a magnetic compass for navigation." In: *Proceedings of the National Academy of Sciences* 96.24, pp. 13845–13846.
Faghihi, F., A. A. Moustafa, R. Heinrich, and F. Wörgötter (2017). "A computational model of conditioning inspired by Drosophila olfactory system." In: *Neural Networks* 87, pp. 96–108. ISSN: 0893-6080. DOI: 10.1016/j.neunet.2016.11.002.
Felsenberg, J., O. Barnstedt, P. Cognigni, S. Lin, and S. Waddell (2017). "Re-evaluation of learned information in Drosophila." In: *Nature* 544.7649, pp. 240–244. ISSN: 0028-0836. DOI: 10.1038/nature21716.
Felsenberg, J., P. F. Jacob, et al. (2018). "Integration of parallel opposing memories underlies memory extinction." In: *Cell* 175.3, 709–722.e15. ISSN: 0092-8674. DOI: 10.1016/j.cell.2018.08.021.
Fent, K. and R. Wehner (1985). "Oceili: a celestial compass in the desert ant Cataglyphis." In: *Science* 228.4696, pp. 192–194.
Fiala, A. (2007). "Olfaction and olfactory learning in Drosophila: recent progress." In: *Current Opinion in Neurobiology* 17.6, pp. 720–726. ISSN: 09594388. DOI: 10.1016/j.conb.2007.11.009.
Finelli, L. A., S. Haney, M. Bazhenov, M. Stopfer, and T. J. Sejnowski (2008). "Synaptic learning rules and sparse coding in a model sensory system." In: *PLoS Computational Biology* 4.4, e1000062. ISSN: 1553-734X. DOI: 10.1371/journal.pcbi.1000062.
Freas, C. A. and P. Schultheiss (2018). "How to navigate in different environments and situations: lessons from ants." In: *Frontiers in psychology* 9, p. 841.
Gerber, B., R. Stocker, T. Tanimura, and A. Thum (2009). "Smelling, tasting, learning: Drosophila as a study case." In: *Chemosensory Systems in Mammals, Fishes, and Insects*. Ed. by W. Meyerhof and S. Korsching. Vol. 47. Springer, pp. 139–186. ISBN: 9783540699187. DOI: 10.1007/400_2008_9.
Gerber, B. and T. Hendel (2006). "Outcome expectations drive learned behaviour in larval Drosophila." In: *Proceedings of the Royal Society B: Biological Sciences* 273.1604, pp. 2965–2968. ISSN: 0962-8452. DOI: 10.1098/rspb.2006.3673.
Gerber, B., H. Tanimoto, and M. Heisenberg (2004). "An engram found? Evaluating the evidence from fruit flies." In: *Current opinion in neurobiology* 14.6, pp. 737–744. ISSN: 0959-4388. DOI: 10.1016/j.conb.2004.10.014.
Gerber, B. and J. Ullrich (1999). "No evidence for olfactory blocking in honeybee classical conditioning." In: *The Journal of experimental biology* 202.13, pp. 1839–1854. ISSN: 1477-9145.

Gershman, S. J. and N. D. Daw (2017). "Reinforcement learning and episodic memory in humans and animals: an integrative framework." In: *Annual review of psychology* 68, pp. 101–128. ISSN: 15829146. DOI: 10.1146/annurev-psych-122414-033625.

Giurfa, M. (2007). "Behavioral and neural analysis of associative learning in the honeybee: a taste from the magic well." In: *Journal of Comparative Physiology A* 193.8, pp. 801–824. ISSN: 0340-7594. DOI: 10.1007/s00359-007-0235-9.

Giurfa, M., E. Fabre, J. Flaven-Pouchon, H. Groll, B. Oberwallner, V. Vergoz, E. Roussel, and J. C. Sandoz (2009). "Olfactory conditioning of the sting extension reflex in honeybees: memory dependence on trial number, interstimulus interval, intertrial interval, and protein synthesis." In: *Learning and Memory* 16.12, pp. 761–765. ISSN: 10720502. DOI: 10.1101/lm.1603009.

Gkanias, E., L. Y. McCurdy, M. N. Nitabach, and B. Webb (2022). "An incentive circuit for memory dynamics in the mushroom body of Drosophila melanogaster." In: *eLife* 11, e75611. DOI: 10.7554/elife.75611.

Gkanias, E., B. Risse, M. Mangan, and B. Webb (July 2019). "From skylight input to behavioural output: a computational model of the insect polarised light compass." In: *PLoS Computational Biology* 15.7. Ed. by J. Ayers, e1007123. ISSN: 1553-734X. DOI: 10.1371/journal.pcbi.1007123.

Goodman, L. J. (1970). "The structure and function of the insect dorsal ocellus." In: *Advances in insect physiology*. Vol. 7. Elsevier, pp. 97–195.

Goulard, R., C. Buehlmann, J. E. Niven, P. Graham, and B. Webb (2021). "A unified mechanism for innate and learned visual landmark guidance in the insect central complex." In: *PLOS Computational Biology* 17.9, e1009383. ISSN: 1553-734X. DOI: 10.1371/journal.pcbi.1009383.

Green, J., A. Adachi, K. K. Shah, J. D. Hirokawa, P. S. Magani, and G. Maimon (2017). "A neural circuit architecture for angular integration in Drosophila." In: *Nature* 546.7656, p. 101.

Greiner, B., T. W. Cronin, W. A. Ribi, W. T. Wcislo, and E. J. Warrant (2007). "Anatomical and physiological evidence for polarisation vision in the nocturnal bee Megalopta genalis." In: *Journal of Comparative Physiology A* 193.6, pp. 591–600.

Grover, D., J.-Y. Chen, J. Xie, J. Li, J.-P. Changeux, and R. J. Greenspan (2022). "Differential mechanisms underlie trace and delay conditioning in Drosophila." In: *Nature*, pp. 1–7. ISSN: 0028-0836. DOI: 10.1038/s41586-022-04433-6.

Guerrieri, F., H. Lachnit, B. Gerber, and M. Giurfa (2005). "Olfactory blocking and odorant similarity in the honeybee." In: *Learning and Memory* 12.2, pp. 86–95. ISSN: 10720502. DOI: 10.1101/lm.79305.

Habenstein, J., E. Amini, K. Grübel, B. e. Jundi, and W. Rössler (2020). "The brain of Cataglyphis ants: aeuronal organization and visual projections." In: *Journal of Comparative Neurology* 528.18, pp. 3479–3506. ISSN: 0021-9967. DOI: 10.1002/cne.24934.

Hammer, M. (1993). "An identified neuron mediates the unconditioned stimulus in associative olfactory learning in honeybees." In: *Nature* 366.6450, pp. 59–63. ISSN: 00280836. DOI: 10.1038/366059a0.

Handler, A., T. G. Graham, R. Cohn, I. Morantte, A. F. Siliciano, J. Zeng, Y. Li, and V. Ruta (2019). "Distinct Dopamine receptor pathways underlie the temporal sensitivity of associative learning." In: *Cell* 178.1. - Timing paper - Decrease in KC-MBON weight for forward pairing (odour-DAN) and increase in weight for

backward pairing (DAN-odour), 60–75.e19. ISSN: 0092-8674. DOI: 10.1016/j.cell.2019.05.040.
Hanesch, U., K.-F. Fischbach, and M. Heisenberg (1989). "Neuronal architecture of the central complex in Drosophila melanogaster." In: *Cell and Tissue Research1* 257.2, pp. 343–366. DOI: 10.1126/science.1251152.8..
Hausen, K. (1976). "Functional characterization and anatomical identification of motion sensitive neurons in the lobula plate of the blowfly Calliphora erythrocephala." In: *Zeitschrift für Naturforschung C* 31.9-10, pp. 629–634. ISSN: 0939-5075. DOI: 10.1515/znc-1976-9-1001.
Hausen, K. (1984). "Photoreception and vision in invertebrates." In: pp. 523–559. DOI: 10.1007/978-1-4613-2743-1_15.
Haynes, P. R., B. L. Christmann, and L. C. Griffith (2015). "A single pair of neurons links sleep to memory consolidation in Drosophila melanogaster." In: *eLife* 4, e03868. ISSN: 2050-084X. DOI: 10.7554/elife.03868.
Hebb, D. O. (1949). *The organization of behavior: a neuropsychological theory*. New York: Psychology Press. DOI: citeulike-article-id:1282862.
Heinze, S., S. Gotthardt, and U. Homberg (2009). "Transformation of polarized light information in the central complex of the locust." In: *Journal of Neuroscience* 29.38, pp. 11783–11793.
Heinze, S. and U. Homberg (2007). "Maplike representation of celestial E-vector orientations in the brain of an insect." In: *Science* 315.5814, pp. 995–997.
Heinze, S. and S. M. Reppert (2011). "Sun compass integration of skylight cues in migratory monarch butterflies." In: *Neuron* 69.2, pp. 345–358.
Heisenberg, M. (2003). "Mushroom body memoir: from maps to models." In: *Nature Reviews Neuroscience* 4.4, pp. 266–275. ISSN: 1471-003X. DOI: 10.1038/nrn1074.
Hertel, H., S. Schäfer, and U. Maronde (1987). "The physiology and morphology of visual commissures in the honeybee brain." In: *Journal of Experimental Biology* 133.1, pp. 283–300. ISSN: 0022-0949. DOI: 10.1242/jeb.133.1.283.
Herz, A. V. M., T. Gollisch, C. K. Machens, and D. Jaeger (2006). "Modeling single-neuron dynamics and computations: a balance of detail and abstraction." In: *Science* 314.5796, pp. 80–85. ISSN: 0036-8075. DOI: 10.1126/science.1127240.
Hige, T., Y. Aso, M. N. Modi, G. M. Rubin, and G. C. Turner (2015). "Heterosynaptic plasticity underlies aversive olfactory learning in Drosophila." In: *Neuron* 88.5, pp. 985–998. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2015.11.003.
Homberg, U. (2015). "Sky compass orientation in desert locusts—evidence from field and laboratory studies." In: *Frontiers in behavioral neuroscience* 9, p. 346.
Homberg, U., S. Heinze, K. Pfeiffer, M. Kinoshita, and B. El Jundi (2011). "Central neural coding of sky polarization in insects." In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 366.1565, pp. 680–687.
Horstmeyer, R., G. Euliss, and R. Athale (2009). "Flexible multimodal camera using a light field architecture." In: *Computational Photography (ICCP), 2009 IEEE International Conference on*. IEEE, pp. 1–8.
Horvath, G., A. Barta, J. Gal, B. Suhai, and O. Haiman (2002). "Ground-based full-sky imaging polarimetry of rapidly changing skies and its use for polarimetric cloud detection." In: *Applied optics* 41.3, pp. 543–559.
Hosler, J. S. and B. H. Smith (2000). "Blocking and the detection of odor components in blends." In: *The Journal of experimental biology* 203.18, pp. 2797–2806. ISSN: 0022-0949.

Huber, R. and M. Knaden (2015). "Egocentric and geocentric navigation during extremely long foraging paths of desert ants." In: *Journal of Comparative Physiology A* 201.6, pp. 609–616.

Huerta, R. and T. Nowotny (2009). "Fast and robust learning by reinforcement signals: explorations in the insect brain." In: *Neural computation* 21.8, pp. 2123–2151. ISSN: 08997667. DOI: 10.1162/neco.2009.03-08-733.

Huerta, R., T. Nowotny, M. García-Sánchez, H. D. I. Abarbanel, and M. I. Rabinovich (2004). "Learning classification in the olfactory system of insects." In: *Neural Computation* 16.8, pp. 1601–1640. ISSN: 0899-7667. DOI: 10.1162/089976604774201613.

Huetteroth, W., E. Perisse, S. Lin, M. Klappenbach, C. Burke, and S. Waddell (2015). "Sweet taste and nutrient value subdivide rewarding dopaminergic neurons in Drosophila." In: *Current Biology* 25.6, pp. 751–758. ISSN: 0960-9822. DOI: 10.1016/j.cub.2015.01.036.

Hulse, B. K. et al. (2021). "A connectome of the Drosophila central complex reveals network motifs suitable for flexible navigation and context-dependent action selection." In: *eLife* 10. DOI: 10.7554/elife.66039.

Ichinose, T., Y. Aso, N. Yamagata, A. Abe, G. M. Rubin, and H. Tanimoto (2015). "Reward signal in a recurrent circuit drives appetitive long-term memory formation." In: *eLife* 4, e10719. DOI: 10.7554/elife.10719.

Islam, M., S. Deeti, T. Murray, and K. Cheng (2022). "What view information is most important in the homeward navigation of an Australian bull ant, Myrmecia midas?" In: *Journal of Comparative Physiology A*, pp. 1–15. ISSN: 0340-7594. DOI: 10.1007/s00359-022-01565-y.

Ito, I., R. C.-y. Ong, B. Raman, and M. Stopfer (2008). "Sparse odor representation and olfactory learning." In: *Nature Neuroscience* 11.10, pp. 1177–1184. ISSN: 1097-6256. DOI: 10.1038/nn.2192.

Ito, K., K. Suzuki, P. Estes, M. Ramaswami, D. Yamamoto, and N. J. Strausfeld (1998). "The organization of extrinsic neurons and their implications in the functional roles of the mushroom bodies in Drosophila melanogaster Meigen." In: *Learning & Memory* 5.1, pp. 52–77. ISSN: 1072-0502.

Izhikevich, E. M. (2007). *Dynamical systems in neuroscience: the geometry of exhitability and bursting*. Cambridge, MA: Mit Press. ISBN: 9780262090438. DOI: 10.4324/9781315744131-2.

Jacob, P. F. and S. Waddell (2020). "Spaced training forms complementary long-term memories of opposite valence in Drosophila." In: *Neuron* 106.6, 977–991.e4. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2020.03.013.

Jiang, L. and A. Litwin-Kumar (2021). "Models of heterogeneous dopamine signaling in an insect learning and memory center." In: *PLOS Computational Biology* 17.8, e1009205. ISSN: 1553-734X. DOI: 10.1371/journal.pcbi.1009205.

Joesch, M., J. Plett, A. Borst, and D. F. Reiff (2008). "Response properties of motion-sensitive visual interneurons in the lobula plate of Drosophila melanogaster." In: *Current Biology* 18.5. ISSN: 0960-9822. DOI: 10.1016/j.cub.2008.02.022.

Jundi, B. el, E. J. Warrant, M. J. Byrne, L. Khaldy, E. Baird, J. Smolka, and M. Dacke (2015). "Neural coding underlying the cue preference for celestial orientation." In: *Proceedings of the National Academy of Sciences* 112.36, pp. 11395–11400.

Kakaria, K. S. and B. L. de Bivort (2017). "Ring attractor dynamics emerge from a spiking model of the entire protocerebral bridge." In: *Frontiers in behavioral neuroscience* 11, p. 8.

Kallman, B. R., H. Kim, and K. Scott (2015). "Excitation and inhibition onto central courtship neurons biases Drosophila mate choice." In: *eLife* 4, e11188. DOI: 10.7554/elife.11188.

Kamhi, J. F., A. B. Barron, and A. Narendra (2020). "Vertical lobes of the mushroom bodies are essential for view-based navigation in Australian Myrmecia ants." In: *Current Biology* 30.17, 3432–3437.e3. ISSN: 0960-9822. DOI: 10.1016/j.cub.2020.06.030.

Kamin, L. J. (1967). "Predictability, surprise, attention and conditioning." In: *Punishment and Aversive Behaviour*, pp. 279–296.

Karman, S. B., S. Z. M. Diah, and I. C. Gebeshuber (2012). "Bio-inspired polarized skylight-based navigation sensors: a review." In: *Sensors* 12.11, pp. 14232–14261.

Keene, A. C., M. J. Krashes, B. Leung, J. A. Bernard, and S. Waddell (2006). "Drosophila dorsal paired medial neurons provide a general mechanism for memory consolidation." In: *Current Biology* 16.15, pp. 1524–1530. ISSN: 0960-9822. DOI: 10.1016/j.cub.2006.06.022.

Keene, A. C. and S. Waddell (2007). "Drosophila olfactory memory: single genes to complex neural circuits." In: *Nature Reviews Neuroscience* 8.5, pp. 341–354. ISSN: 1471003X. DOI: 10.1038/nrn2098.

Keinert, B., M. Innmann, M. Sänger, and M. Stamminger (2015). "Spherical fibonacci mapping." In: *ACM Transactions on Graphics (TOG)* 34.6, pp. 1–7. ISSN: 0730-0301. DOI: 10.1145/2816795.2818131.

Khotanzad, A. and Y. Hong (1990). "Invariant image recognition by Zernike moments." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12.5, pp. 489–497. ISSN: 0162-8828. DOI: 10.1109/34.55109.

Kimble, G. A. (1961). "Hilgard and Marquis'" Conditioning and learning."" In: *American Psychological Association*.

Kinoshita, M., K. Pfeiffer, and U. Homberg (2007). "Spectral properties of identified polarized-light sensitive interneurons in the brain of the desert locust Schistocerca gregaria." In: *Journal of Experimental Biology* 210.8, pp. 1350–1361. ISSN: 0022-0949. DOI: 10.1242/jeb.02744.

Klopf, A. H. (1972). *Brain function and adaptive systems: a heterostatic theory*. Tech. rep. 133. Bedford, MA: Air Force Cambridge research Labolatories.

Kohler, M. and R. Wehner (2005). "Idiosyncratic route-based memories in desert ants, Melophorus bagoti: how do they interact with path-integration vectors?" In: *Neurobiology of Learning and Memory* 83.1, pp. 1–12. ISSN: 1074-7427. DOI: 10.1016/j.nlm.2004.05.011.

Komischke, B., J. C. Sandoz, H. Lachnit, and M. Giurfa (2003). "Non-elemental processing in olfactory discrimination tasks needs bilateral input in honeybees." In: *Behavioural brain research* 145.1-2, pp. 135–143. ISSN: 0166-4328. DOI: 10.1016/S0166-4328(03)00105-0.

Krashes, M. J. and S. Waddell (2011a). "Drosophila aversive olfactory conditioning." In: *Cold Spring Harbor Protocols* 2011.5, pdb.prot5608–pdb.prot5608. DOI: 10.1101/pdb.prot5608.

Krashes, M. J., S. DasGupta, A. Vreede, B. White, J. D. Armstrong, and S. Waddell (2009). "A neural circuit mechanism integrating motivational state with memory expression in Drosophila." In: *Cell* 139.2, pp. 416–427. ISSN: 0092-8674. DOI: 10.1016/j.cell.2009.08.035.

Krashes, M. J., A. C. Keene, B. Leung, J. D. Armstrong, and S. Waddell (2007). "Sequential use of mushroom body neuron subsets during Drosophila odor memory

processing." In: *Neuron* 53.1, pp. 103–115. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2006.11.021.
Krashes, M. J. and S. Waddell (2011b). "Drosophila appetitive olfactory conditioning." In: *Cold Spring Harbor Protocols* 2011.5, pdb.prot5609. ISSN: 1940-3402. DOI: 10.1101/pdb.prot5609.
Kühn-Bühlmann, S. and R. Wehner (2006). "Age-dependent and task-related volume changes in the mushroom bodies of visually guided desert ants, Cataglyphis bicolor." In: *Journal of Neurobiology* 66.6, pp. 511–521. DOI: 10.1002/neu.
Labhart, T. (1999). "How polarization-sensitive interneurones of crickets see the polarization pattern of the sky: a field study with an opto-electronic model neurone." In: *Journal of Experimental Biology* 202.7, pp. 757–770.
Labhart, T. (1986). "The electrophysiology of photoreceptors in different eye regions of the desert ant, Cataglyphis bicolor." In: *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology* 158.1, pp. 1–7.
Labhart, T. (1988). "Polarization-opponent interneurons in the insect visual system." In: *Nature* 331.6155, p. 435.
Labhart, T. (2000). "Polarization-sensitive interneurons in the optic lobe of the desert ant Cataglyphis bicolor." In: *Naturwissenschaften* 87.3, pp. 133–136.
Labhart, T. (2016). "Can invertebrates see the e-vector of polarization as a separate modality of light?" In: *Journal of Experimental Biology* 219.24, pp. 3844–3856.
Labhart, T., B. Hodel, and I. Valenzuela (1984). "The physiology of the cricket's compound eye with particular reference to the anatomically specialized dorsal rim area." In: *Journal of Comparative Physiology A* 155.3, pp. 289–296.
Labhart, T. and E. P. Meyer (1999). "Detectors for polarized skylight in insects: a survey of ommatidial specializations in the dorsal rim area of the compound eye." In: *Microscopy research and technique* 47.6, pp. 368–379.
Labhart, T., J. Petzold, and H. Helbling (2001). "Spatial integration in polarization-sensitive interneurones of crickets: a survey of evidence, mechanisms and benefits." In: *Journal of Experimental Biology* 204.14, pp. 2423–2430.
Lambrinos, D., H. Kobayashi, R. Pfeifer, M. Maris, T. Labhart, and R. Wehner (1997). "An autonomous agent navigating with a polarized light compass." In: *Adaptive behavior* 6.1, pp. 131–161.
Lambrinos, D., R. Möller, T. Labhart, R. Pfeifer, and R. Wehner (2000). "A mobile robot employing insect strategies for navigation." In: *Robotics and Autonomous systems* 30.1, pp. 39–64.
Laurent, G. (2002). "Olfactory network dynamics and the coding of multidimensional signals." In: *Nature Reviews Neuroscience* 3.11, pp. 884–895. ISSN: 1471-003X. DOI: 10.1038/nrn964.
Lee, C. and D. Kim (2018). "Visual homing navigation with Haar-Like features in the snapshot." In: *IEEE Access* 6, pp. 33666–33681. ISSN: 2169-3536. DOI: 10.1109/access.2018.2842679.
Lehrer, M. (1998). "Looking all around: honeybees use different cues in different eye regions." In: *Journal of Experimental Biology* 201.24, pp. 3275–3292. ISSN: 0022-0949. DOI: 10.1242/jeb.201.24.3275.
Li, F. et al. (2020). "The connectome of the adult Drosophila mushroom body provides insights into function." In: *eLife* 9, e62576. DOI: 10.7554/elife.62576.
Lin, S., D. Owald, V. Chandra, C. Talbot, W. Huetteroth, and S. Waddell (2014). "Neural correlates of water reward in thirsty Drosophila." In: *Nature Neuroscience* 17.11, pp. 1536–1542. ISSN: 1097-6256. DOI: 10.1038/nn.3827.

Litwin-Kumar, A., K. D. Harris, R. Axel, H. Sompolinsky, and L. F. Abbott (Mar. 2017). "Optimal degrees of synaptic connectivity." In: *Neuron* 5, 1153–1164.e7. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2017.01.030.

Liu, C., P.-Y. Plaçais, N. Yamagata, B. D. Pfeiffer, Y. Aso, A. B. Friedrich, I. Siwanowicz, G. M. Rubin, T. Preat, and H. Tanimoto (2012). "A subset of dopamine neurons signals reward for odour memory in Drosophila." In: *Nature* 488.7412, pp. 512–516. ISSN: 0028-0836. DOI: 10.1038/nature11304.

Liu, W. W. and R. I. Wilson (2013). "Glutamate is an inhibitory neurotransmitter in the Drosophila olfactory system." In: *Proceedings of the National Academy of Sciences* 110.25, pp. 10294–10299. ISSN: 0027-8424. DOI: 10.1073/pnas.1220560110.

Liu, X. and R. L. Davis (2009). "The GABAergic anterior paired lateral neuron suppresses and is suppressed by olfactory learning." In: *Nature Neuroscience* 12.1, pp. 53–59. ISSN: 1097-6256. DOI: 10.1038/nn.2235.

Loesel, R. and U. Homberg (2001). "Anatomy and physiology of neurons with processes in the accessory medulla of the cockroach Leucophaea maderae." In: *Journal of Comparative Neurology* 439.2, pp. 193–207. ISSN: 1096-9861. DOI: 10.1002/cne.1342.

Lukes, G., B. Thompson, and P. Werbos (1990). "Expectation driven learning with an associative memory." In: *International Joint Conference on Neural Networks*. 1. Washington, DC, pp. 521–524.

Lyu, C., L. F. Abbott, and G. Maimon (2022). "Building an allocentric travelling direction signal via vector computation." In: *Nature* 601.7891, pp. 92–97. ISSN: 0028-0836. DOI: 10.1038/s41586-021-04067-0.

Ma, T., X. Hu, J. Lian, and L. Zhang (2014). "Compass information extracted from a polarization sensor using a least-squares algorithm." In: *Applied Optics* 53.29, p. 6735. DOI: 10.1364/ao.53.006735.

MacLean, P. D. (1949). "Psychosomatic disease and the "visceral brain"." In: *Psychosomatic Medicine* 11.6, pp. 338–353. ISSN: 0033-3174. DOI: 10.1097/00006842-194911000-00003.

Manakov, A., J. Restrepo, O. Klehm, R. Hegedus, E. Eisemann, H.-P. Seidel, and I. Ihrke (2013). "A reconfigurable camera add-on for high dynamic range, multispectral, polarization, and light-field imaging." In: *ACM Transactions on Graphics* 32.4, pp. 47–1.

Mangan, M. and B. Webb (2012). "Spontaneous formation of multiple routes in individual desert ants (Cataglyphis velox)." In: *Behavioral Ecology* 23.5, pp. 944–954. ISSN: 1045-2249. DOI: 10.1093/beheco/ars051.

Mao, Z. and R. L. Davis (2009). "Eight different types of Dopaminergic neurons innervate the Drosophila mushroom body neuropil: anatomical and physiological heterogeneity." In: *Frontiers in Neural Circuits* 3, p. 5. DOI: 10.3389/neuro.04.005.2009.

Maronde, U. (1991). "Common projection areas of antennal and visual pathways in the honeybee brain, apis mellifera." In: *Journal of Comparative Neurology* 309.3, pp. 328–340. ISSN: 1096-9861. DOI: 10.1002/cne.903090304.

Marr, D. (1982). *Vision: a computational approach*. San Fransisco: Freeman & Co.

Masuda-Nakagawa, L. M., K. Ito, T. Awasaki, and C. J. O'Kane (2014). "A single GABAergic neuron mediates feedback of odor-evoked signals in the mushroom body of larval Drosophila." In: *Frontiers in neural circuits* 8.April, p. 35. ISSN: 1662-5110. DOI: 10.3389/fncir.2014.00035.

Matheson, A. M., A. J. Lanz, A. M. Licata, T. A. Currier, M. H. Syed, and K. I. Nagel (2021). "Organization of central circuits for wind-guided olfactory navigation." In: *bioRxiv*, p. 2021.04.21.440842. DOI: 10.1101/2021.04.21.440842.

Matsumoto, Y. and M. Mizunami (2004). "Context-dependent olfactory learning in an insect." In: *Learning and Memory* 11.3, pp. 288–293. ISSN: 1072-0502. DOI: 10.1101/lm.72504.

May, C. E., J. Rosander, J. Gottfried, E. Dennis, and M. Dus (2020). "Dietary sugar inhibits satiation by decreasing the central processing of sweet taste." In: *eLife* 9, e54530. DOI: 10.7554/elife.54530.

McCarthy, E. v., Y. Wu, T. deCarvalho, C. Brandt, G. Cao, and M. N. Nitabach (2011). "Synchronized bilateral synaptic inputs to Drosophila melanogaster neuropeptidergic rest/arousal neurons." In: *The Journal of Neuroscience* 31.22, pp. 8181–8193. ISSN: 0270-6474. DOI: 10.1523/jneurosci.2017-10.2011.

McCulloch, W. S. and W. Pitts (1943). "A logical calculus of the ideas immanent in nervous activity." In: *The Bulletin of Mathematical Biophysics* 5.4, pp. 115–133. ISSN: 00074985. DOI: 10.1007/BF02478259.

McCurdy, L. Y., P. Sareen, P. A. Davoudian, and M. N. Nitabach (Feb. 2021). "Dopaminergic mechanism underlying reward-encoding of punishment omission during reversal learning in Drosophila." In: *Nature Communications* 12.1, p. 1115. DOI: 10.1038/s41467-021-21388-w.

McGuire, S. E., P. T. Le, and R. L. Davis (2001). "The role of Drosophila mushroom body signaling in olfactory memory." In: *Science* 293.5533, pp. 1330–1333. ISSN: 0036-8075. DOI: 10.1126/science.1062622.

Menzel, R. and R. Wehner (1970). "Augenstrukturen bei verschieden gro\Arbeiterinnen von Cataglyphis bicolor Fabr. (Formicidae, hymenoptera)." In: *Zeitschrift für vergleichende Physiologie* 68.4, pp. 446–449. ISSN: 0044-362x. DOI: 10.1007/bf00297741.

Menzel, R., U. Greggers, et al. (2005). "Honey bees navigate according to a map-like spatial memory." In: *Proceedings of the National Academy of Sciences* 102.8, pp. 3040–3045. ISSN: 0027-8424. DOI: 10.1073/pnas.0408550102.

Menzel, R. and A. W. Snyder (1974). "Polarised light detection in the bee, Apis mellifera." In: *Journal of comparative physiology* 88.3, pp. 247–270.

Meyer, S., T. Nowotny, P. Graham, A. Dewar, and A. Philippides (July 2020). "Snapshot navigation in the wavelet domain." In: *9th International Conference, Living Machines 2020*. Biomimetic and Biohybrid Systems. Freiburg, Germany, pp. 245–256. ISBN: 9783030643126. DOI: 10.1007/978-3-030-64313-3_24.

Michels, B., Y.-c. Chen, T. Saumweber, D. Mishra, H. Tanimoto, B. Schmid, O. Engmann, and B. Gerber (2011). "Cellular site and molecular mode of synapsin action in associative learning." In: *Learning & Memory* 18.5, pp. 332–344. ISSN: 1072-0502. DOI: 10.1101/lm.2101411.

Millward, B., S. Maddock, and M. Mangan (2021). "CompoundRay: an open-source tool for high-speed and high-fidelity rendering of compound eyes." In: *bioRxiv*. DOI: 10.1101/2021.09.20.461066.

Milyaev, N., D. Osumi-Sutherland, S. Reeve, N. Burton, R. A. Baldock, and J. D. Armstrong (2012). "The virtual fly brain browser and query interface." In: *Bioinformatics* 28.3, pp. 411–415. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btr677.

Mnih, V. et al. (2015). "Human-level control through deep reinforcement learning." In: *Nature* 518.7540, p. 529.

Moël, F. L., T. Stone, M. Lihoreau, A. Wystrach, and B. Webb (2019). "The central complex as a potential substrate for vector based navigation." In: *Frontiers in Psychology* 10, p. 690. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2019.00690.

Möel, F. L. and A. Wystrach (2020). "Opponent processes in visual memories: a model of attraction and repulsion in navigating insects' mushroom bodies." In: *PLOS Computational Biology* 16.2, e1007631. ISSN: 1553-734X. DOI: 10.1371/journal.pcbi.1007631.

Mota, T., N. Yamagata, M. Giurfa, W. Gronenberg, and J. .-. Sandoz (2011). "Neural organization and visual processing in the anterior optic tubercle of the honeybee brain." In: *Journal of Neuroscience* 31.32, pp. 11443–11456. ISSN: 0270-6474. DOI: 10.1523/jneurosci.0995-11.2011.

Müller, M. and R. Wehner (1988). "Path integration in desert ants, Cataglyphis fortis." In: *Proceedings of the National Academy of Sciences* 85.14, pp. 5287–5290.

Murray, T., Z. Kocsi, H. Dahmen, A. Narendra, F. L. Möel, A. Wystrach, and J. Zeil (2019). "The role of attractive and repellent scene memories in ant homing (Myrmecia croslandi)." In: *Journal of Experimental Biology*, jeb.210021. ISSN: 0022-0949. DOI: 10.1242/jeb.210021.

Naumann, R. K., J. M. Ondracek, S. Reiter, M. Shein-Idelson, M. A. Tosches, T. M. Yamawaki, and G. Laurent (2015). "The reptilian brain." In: *Current Biology* 25.8, R317–R321. ISSN: 0960-9822. DOI: 10.1016/j.cub.2015.02.049.

Ng, M., R. D. Roorda, S. Q. Lima, B. V. Zemelman, P. Morcillo, and G. Miesenböck (2002). "Transmission of olfactory information between three populations of neurons in the antennal lobe of the fly." In: *Neuron* 36.3, pp. 463–474. ISSN: 0896-6273. DOI: 10.1016/s0896-6273(02)00975-3.

Nilsson, D.-E., T. Labhart, and E. Meyer (1987). "Photoreceptor design and optical properties affecting polarization sensitivity in ants and crickets." In: *Journal of Comparative Physiology A* 161.5, pp. 645–658.

Niv, Y. (2009). "Reinforcement learning in the brain." In: *Journal of Mathematical Psychology* 53.3, pp. 139–154. ISSN: 0022-2496. DOI: 10.1016/j.jmp.2008.12.005.

Nordström, K., P. D. Barnett, and D. C. O'Carroll (2006). "Insect detection of small targets moving in visual clutter." In: *PLoS Biology* 4.3, e54. ISSN: 1544-9173. DOI: 10.1371/journal.pbio.0040054.

Nowotny, T., R. Huerta, H. D. I. Abarbanel, and M. I. Rabinovich (2005). "Self-organization in the olfactory system: one shot odor recognition in insects." In: *Biological Cybernetics* 93.6, pp. 436–446. ISSN: 0340-1200. DOI: 10.1007/s00422-005-0019-7.

O'Reilly, R. C. (2001). "Generalization in interactive networks: the benefits of inhibitory competition and Hebbian learning." In: *Neural Computation* 13.6, pp. 1199–1241. DOI: 10.1162/08997660152002834.

Owald, D., J. Felsenberg, C. B. Talbot, G. Das, E. Perisse, W. Huetteroth, and S. Waddell (2015). "Activity of defined mushroom body output neurons underlies learned olfactory behavior in Drosophila." In: *Neuron* 86.2, pp. 417–427. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2015.03.025.

Page, E. B. (1963). "Ordered hypotheses for multiple treatments: a significance test for linear ranks." In: *Journal of the American Statistical Association* 58.301, p. 216. ISSN: 0162-1459. DOI: 10.2307/2282965.

Papadopoulou, M., S. Cassenaer, T. Nowotny, and G. Laurent (2011). "Normalization for Sparse Encoding of Odors by a Wide-Field Interneuron." In: *Science* 332.6030, pp. 721–725. ISSN: 0036-8075. DOI: 10.1126/science.1201835.

Papez, J. W. (1937). "A proposed mechanism of emotion." In: *Archives of Neurology & Psychiatry* 38.4, pp. 725–743. ISSN: 0096-6754. DOI: 10.1001/archneurpsyc.1937.02260220069003.

Paulk, A. C., A. M. Dacks, J. Phillips-Portillo, J. .-. Fellous, and W. Gronenberg (2009). "Visual processing in the central bee brain." In: *Journal of Neuroscience* 29.32, pp. 9987–9999. ISSN: 0270-6474. DOI: 10.1523/jneurosci.1325-09.2009.

Paulk, A. C., J. Phillips-Portillo, A. M. Dacks, J. .-. Fellous, and W. Gronenberg (2008). "The processing of color, motion, and stimulus timing are anatomically segregated in the bumblebee brain." In: *Journal of Neuroscience* 28.25, pp. 6319–6332. ISSN: 0270-6474. DOI: 10.1523/jneurosci.1196-08.2008.

Pavlov, I. P. (1949). "Readings in general psychology." In: pp. 249–267. DOI: 10.1037/11352-036.

Pavlowsky, A., J. Schor, P.-Y. Plaçais, and T. Preat (2018). "A GABAergic feedback shapes dopaminergic input on the Drosophila mushroom body to promote appetitive long-term memory." In: *Current Biology* 28.11, 1783–1793.e4. ISSN: 0960-9822. DOI: 10.1016/j.cub.2018.04.040.

Pearce, J. M. (1994). "Similarity and discrimination: a selective review and a connectionist model." In: *Psychological review* 101.4, pp. 587–607. ISSN: 0033-295X.

Pedregosa, F. et al. (2011). "Scikit-learn: Machine Learning in Python." In: *Journal of Machine Learning Research* 12, pp. 2825–2830.

Pegel, U., K. Pfeiffer, and U. Homberg (2018). "Integration of celestial compass cues in the central complex of the locust brain." In: *Journal of Experimental Biology* 221.2. ISSN: 0022-0949. DOI: 10.1242/jeb.171207. eprint: http://jeb.biologists.org/content/221/2/jeb171207.full.pdf. URL: http://jeb.biologists.org/content/221/2/jeb171207.

Peng, F. and L. Chittka (2017). "A simple computational model of the bee mushroom body can explain seemingly complex forms of olfactory learning and memory." In: *Current Biology* 27.2, pp. 224–230. ISSN: 0960-9822. DOI: 10.1016/j.cub.2016.10.054.

Pennartz, C. (1997). "Reinforcement learning by Hebbian synapses with adaptive thresholds." In: *Neuroscience* 81.2, pp. 303–319. ISSN: 0306-4522. DOI: 10.1016/s0306-4522(97)00118-8.

Perisse, E., D. Owald, O. Barnstedt, C. B. Talbot, W. Huetteroth, and S. Waddell (2016). "Aversive learning and appetitive motivation toggle feed-forward inhibition in the Drosophila mushroom body." In: *Neuron* 90.5, pp. 1086–1099. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2016.04.034.

Pfeffer, S. E. and M. Wittlinger (2016). "Optic flow odometry operates independently of stride integration in carried ants." In: *Science* 353.6304, pp. 1155–1157. ISSN: 0036-8075. DOI: 10.1126/science.aaf9754.

Pfeiffer, K. and U. Homberg (2007). "Coding of azimuthal directions via time-compensated combination of celestial compass cues." In: *Current biology* 17.11, pp. 960–965.

Pfeiffer, K., M. Kinoshita, and U. Homberg (2005). "Polarization-sensitive and light-sensitive neurons in two parallel pathways passing through the anterior optic tubercle in the locust brain." In: *Journal of Neurophysiology* 94.6, pp. 3903–3915. ISSN: 0022-3077. DOI: 10.1152/jn.00276.2005.

Pisokas, I., S. Heinze, and B. Webb (2020). "The head direction circuit of two insect species." In: *eLife* 9, e53985. DOI: 10.7554/elife.53985.

Pitman, J. L., W. Huetteroth, C. J. Burke, M. J. Krashes, S.-L. Lai, T. Lee, and S. Waddell (2011). "A pair of inhibitory neurons are required to sustain labile memory in the

Drosophila mushroom body." In: *Current Biology* 21.10, pp. 855–861. ISSN: 0960-9822. DOI: 10.1016/j.cub.2011.03.069.
Plaçais, P.-Y., S. Trannoy, A. B. Friedrich, H. Tanimoto, and T. Preat (2013). "Two pairs of mushroom body efferent neurons are required for appetitive long-term memory retrieval in Drosophila." In: *Cell Reports* 5.3, pp. 769–780. ISSN: 2211-1247. DOI: 10.1016/j.celrep.2013.09.032.
Plutchik, R. (2001). "The nature of emotions." In: *American Scientist* 89.4, pp. 344–350.
Polster, J., J. Petrasch, R. Menzel, and T. Landgraf (2018). "Reconstructing the visual perception of honey bees in complex 3-D worlds." In: *arXiv*. eprint: 1811.07560.
Posnien, N. et al. (2012). "Evolution of eye morphology and rhodopsin expression in the Drosophila melanogaster species subgroup." In: *PLoS ONE* 7.5, e37346. DOI: 10.1371/journal.pone.0037346.
Pribbenow, C. et al. (2021). "Postsynaptic plasticity of cholinergic synapses underlies the induction and expression of appetitive memories in Drosophila." In: *bioRxiv*, p. 2021.07.01.450776. DOI: 10.1101/2021.07.01.450776.
Quinn, W. G., W. A. Harris, and S. Benzer (1974). "Conditioned behavior in Drosophila melanogaster." In: *Proceedings of the National Academy of Sciences* 71.3, pp. 708–712.
Raderschall, C. A., A. Narendra, and J. Zeil (2016). "Head roll stabilisation in the nocturnal bull ant Myrmecia pyriformis: implications for visual navigation." In: *Journal of Experimental Biology*, jeb–134049.
Rayleigh, J. W. S. B. (1871). *On the scattering of light by small particles*.
Rayleigh, L. (1881). "X. On the electromagnetic theory of light." In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 12.73, pp. 81–101.
Rayleigh, L. (1899). "XXXIV. On the transmission of light through an atmosphere containing small particles in suspension, and on the origin of the blue of the sky." In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 47.287, pp. 375–384.
Rescorla, R. A. (1988). "Pavlovian conditioning: it's not what you think it is." In: *American Psychologist* 43.3, pp. 151–160.
Rescorla, R. A. and A. R. Wagner (1972). "A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement." In: *Classical conditioning II: current research and theory*. Ed. by W. F. P. A H Black. New York: Appleton-Century-Crofts, pp. 64–99.
Riemensperger, T., T. Völler, P. Stock, E. Buchner, and A. Fiala (2005). "Punishment prediction by dopaminergic neurons in Drosophila." In: *Current Biology* 15.21, pp. 1953–1960. ISSN: 09609822. DOI: 10.1016/j.cub.2005.09.042.
Robertson, H. M. (2018). "Molecular evolution of the major arthropod chemoreceptor gene families." In: *Annual Review of Entomology* 64.1, pp. 1–16. ISSN: 0066-4170. DOI: 10.1146/annurev-ento-020117-043322.
Root, C. M., J. L. Semmelhack, A. M. Wong, J. Flores, and J. W. Wang (2007). "Propagation of olfactory information in Drosophila." In: *Proceedings of the National Academy of Sciences* 104.28, pp. 11826–11831. ISSN: 0027-8424. DOI: 10.1073/pnas.0704523104.
Rossel, S. and R. Wehner (1986). "Polarization vision in bees." In: *Nature* 323.6084, p. 128.
Roxo, M. R., P. R. Franceschini, C. Zubaran, F. D. Kleber, and J. W. Sander (2011). "The limbic system conception and its historical evolution." In: *TheScientificWorldJOURNAL* 11, pp. 2427–2440. DOI: 10.1100/2011/157150.

Rummery, G. A. and M. Niranjan (1994). *On-line Q-learning using connectionist systems*. Cambridge, England: University of Cambridge.
Sakura, M., D. Lambrinos, and T. Labhart (2008). "Polarized skylight navigation in insects: model and electrophysiology of e-vector coding by neurons in the central complex." In: *Journal of neurophysiology* 99.2, pp. 667–682.
Sarkar, M., D. S. S. Bello, C. Van Hoof, and A. J. Theuwissen (2011). "Integrated polarization-analyzing CMOS image sensor for detecting the incoming light ray direction." In: *IEEE Transactions on Instrumentation and Measurement* 60.8, pp. 2759–2767.
Sato, C., Y. Matsumoto, M. Sakura, and M. Mizunami (2006). "Contextual olfactory learning in cockroaches." In: *Neuroreport* 17.5, pp. 553–557. ISSN: 0959-4965. DOI: 10.1097/01.wnr.0000209002.17610.79.
Saumweber, T. et al. (2018). "Functional architecture of reward learning in mushroom body extrinsic neurons of larval Drosophila." In: *Nature Communications* 9.1, p. 1104. DOI: 10.1038/s41467-018-03130-1.
Scargle, J. D. (1982). "Studies in astronomical time series analysis. II-Statistical aspects of spectral analysis of unevenly spaced data." In: *The Astrophysical Journal* 263, pp. 835–853.
Schleyer, M., M. Fendt, S. Schuller, and B. Gerber (2018). "Associative learning of stimuli paired and unpaired with reinforcement: evaluating evidence from maggots, flies, bees, and rats." In: *Frontiers in Psychology* 9, p. 1494. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2018.01494.
Schleyer, M., T. Saumweber, W. Nahrendorf, B. Fischer, D. v. Alpen, D. Pauls, A. Thum, and B. Gerber (2011). "A behavior-based circuit model of how outcome expectations organize learned behavior in larval Drosophila." In: *Learning & Memory* 18.10, pp. 639–653. ISSN: 1072-0502. DOI: 10.1101/lm.2163411.
Schleyer, M., A. Weiglein, et al. (2020). "Identification of dopaminergic neurons that can both establish associative memory and acutely terminate its behavioral expression." In: *Journal of Neuroscience* 40.31, JN–RM-0290-20. ISSN: 0270-6474. DOI: 10.1523/jneurosci.0290-20.2020.
Schmitt, F., S. M. Stieb, R. Wehner, and W. Rössler (2016). "Experience-related reorganization of giant synapses in the lateral complex: Potential role in plasticity of the sky-compass pathway in the desert ant Cataglyphis fortis." In: *Developmental neurobiology* 76.4, pp. 390–404.
Schmuker, M., M. d. Bruyne, M. Hähnel, and G. Schneider (2007). "Predicting olfactory receptor neuron responses from odorant structure." In: *Chemistry Central Journal* 1.1, p. 11. DOI: 10.1186/1752-153x-1-11.
Schmuker, M., T. Pfeil, and M. P. Nawrot (2014). "A neuromorphic network for generic multivariate data classification." In: *Proceedings of the National Academy of Sciences* 111.6, pp. 2081–2086. ISSN: 0027-8424. DOI: 10.1073/pnas.1303053111.
Schmuker, M. and G. Schneider (2007). "Processing and classification of chemical data inspired by insect olfaction." In: *Proceedings of the National Academy of Sciences* 104.51, pp. 20285–20289. ISSN: 0027-8424. DOI: 10.1073/pnas.0705683104.
Schroll, C. et al. (2006). "Light-induced activation of distinct modulatory neurons triggers appetitive or aversive learning in Drosophila larvae." In: *Current Biology* 16.17, pp. 1741–1747. ISSN: 0960-9822. DOI: 10.1016/j.cub.2006.07.023.
Schubert, M., H. Lachnit, S. Francucci, and M. Giurfa (2002). "Nonelemental visual learning in honeybees." In: *Animal Behaviour* 64.2, pp. 175–184. ISSN: 00033472. DOI: 10.1006/anbe.2002.3055.

Schwaerzel, M., M. Monastirioti, H. Scholz, F. Friggi-Grelin, S. Birman, and M. Heisenberg (2003). "Dopamine and octopamine differentiate between aversive and appetitive olfactory memories in Drosophila." In: *Journal of Neuroscience* 23.33, pp. 10495–10502. ISSN: 0270-6474. DOI: 10.1523/jneurosci.23-33-10495.2003.

Schwarz, S., L. Clement, E. Gkanias, and A. Wystrach (2020). "How do backward-walking ants (Cataglyphis velox) cope with navigational uncertainty?" In: *Animal Behaviour* 164, pp. 133–142. ISSN: 0003-3472. DOI: 10.1016/j.anbehav.2020.04.006.

Schwarz, S., M. Mangan, B. Webb, and A. Wystrach (2020). "Route-following ants respond to alterations of the view sequence." In: *Journal of Experimental Biology* 223.14, jeb218701. ISSN: 0022-0949. DOI: 10.1242/jeb.218701.

Schwarz, S., A. Narendra, and J. Zeil (2011). "The properties of the visual system in the Australian desert ant Melophorus bagoti." In: *Arthropod Structure & Development* 40.2, pp. 128–134. ISSN: 1467-8039. DOI: 10.1016/j.asd.2010.10.003.

Seelig, J. D. and V. Jayaraman (2015). "Neural dynamics for landmark orientation and angular path integration." In: *Nature* 521.7551, p. 186.

Seidl, T. and R. Wehner (2008). "Walking on inclines: how do desert ants monitor slope and step length." In: *Frontiers in zoology* 5.1, p. 8.

Seki, Y., H. K. M. Dweck, J. Rybak, D. Wicher, S. Sachse, and B. S. Hansson (2017). "Olfactory coding from the periphery to higher brain centers in the Drosophila brain." In: *BMC Biology* 15.1, p. 56. DOI: 10.1186/s12915-017-0389-z.

Senapati, B., C.-H. Tsao, Y.-A. Juan, T.-H. Chiu, C.-L. Wu, S. Waddell, and S. Lin (2019). "A neural mechanism for deprivation state-specific expression of relevant memories in Drosophila." In: *Nature Neuroscience* 22.12, pp. 2029–2039. ISSN: 1097-6256. DOI: 10.1038/s41593-019-0515-z.

Singh, S. P. and R. S. Sutton (1996). "Reinforcement learning with replacing eligibility traces." In: *Machine learning* 22.1-3, p. 1996.

Skinner, B. F. (1938). *The behavior of organisms: an experimental analysis*. New York: BF Skinner Foundations. ISBN: 1-58390-007-1. DOI: 10.1080/00221309.1936.9713156. arXiv: arXiv:1011.1669v3.

Smith, B. H. and S. Cobey (1994). "The olfactory memory of the honeybee Apis mellifera. II. Blocking between odorants in binary mixtures." In: *Journal of Experimental Biology* 195.1, pp. 97–108. ISSN: 00220949.

Smith, D., J. Wessnitzer, and B. Webb (2008). "A model of associative learning in the mushroom body." In: *Biological Cybernetics* 99.2, pp. 89–103. ISSN: 0340-1200. DOI: 10.1007/s00422-008-0241-1.

Smith, F. and D. Stewart (2014). "Robot and insect navigation by polarized skylight." In: *Proceedings of the International Joint Conference on Biomedical Engineering Systems and Technologies-Volume 4*. SCITEPRESS-Science and Technology Publications, Lda, pp. 183–188.

Sokolowski, M. B. (2001). "Drosophila: genetics meets behaviour." In: *Nature Reviews Genetics* 2.11, p. 879. ISSN: 0167-8396. DOI: http://dx.doi.org/10.1016/j.cagd.2005.06.004.

Sommer, S., C. von Beeren, and R. Wehner (2008). "Multiroute memories in desert ants." In: *Proceedings of the National Academy of Sciences* 105.1, pp. 317–322. ISSN: 1091-6490. DOI: 10.1073/pnas.0710157104.

Souman, J. L., I. Frissen, M. N. Sreenivasa, and M. O. Ernst (2009). "Walking straight into circles." In: *Current biology* 19.18, pp. 1538–1542.

Springer, M. and M. P. Nawrot (2021). "A mechanistic model for reward prediction and extinction learning in the fruit fly." In: *eNeuro*, ENEURO.0549–20.2021. ISSN: 2373-2822. DOI: 10.1523/eneuro.0549-20.2021.

Srinivasan, M. V. (2015). "Where paths meet and cross: navigation by path integration in the desert ant and the honeybee." In: *Journal of Comparative Physiology A* 201.6, pp. 533–546. ISSN: 14321351. DOI: 10.1007/s00359-015-1000-0.

Srivastava, N., E. Mansimov, and R. Salakhutdinov (2015). "Unsupervised learning of video representations using LSTMs." In: *Arxiv* 37. DOI: citeulike-article-id:13519737. arXiv: 1502.04681.

Stankiewicz, J. and B. Webb (2021). "Looking down: a model for visual route following in flying insects." In: *Bioinspiration & Biomimetics* 16.5, p. 055007. ISSN: 1748-3182. DOI: 10.1088/1748-3190/ac1307.

Sten, T. H., R. Li, A. Otopalik, and V. Ruta (2020). "An arousal-gated visual circuit controls pursuit during Drosophila courtship." In: *bioRxiv*, p. 2020.08.31.275883. DOI: 10.1101/2020.08.31.275883.

Stewart, D.-E. (1958). "The Emergence of Vision in the Animal World* The Lister Oration." In: *American Journal of Ophthalmology* 46.4, pp. 447–463. ISSN: 0002-9394. DOI: 10.1016/0002-9394(58)91125-5.

Stocker, R. F., M. C. Lienhard, A. Borst, and K. Fischbach (1990). "Neuronal architecture of the antennal lobe in Drosophila melanogaster." In: *Cell and tissue research* 262.1, pp. 9–34.

Stöckl, A. L., D. C. O'Carroll, and E. J. Warrant (2020). "Hawkmoth lamina monopolar cells act as dynamic spatial filters to optimize vision at different light levels." In: *Science Advances* 6.16, eaaz8645. DOI: 10.1126/sciadv.aaz8645.

Stone, T., D. Differt, M. Milford, and B. Webb (2016). "Skyline-based localisation for aggressively manoeuvring robots using UV sensors and spherical harmonics." In: *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5615–5622. DOI: 10.1109/icra.2016.7487780.

Stone, T., M. Mangan, A. Wystrach, and B. Webb (2018). "Rotation invariant visual processing for spatial memory in insects." In: *Interface Focus* 8.4, p. 20180010. ISSN: 2042-8898. DOI: 10.1098/rsfs.2018.0010.

Stone, T., B. Webb, A. Adden, N. B. Weddig, A. Honkanen, R. Templin, W. Wcislo, L. Scimeca, E. Warrant, and S. Heinze (2017). "An anatomically constrained model for path integration in the bee brain." In: *Current Biology* 27.20, 3069–3085.e11. ISSN: 0960-9822. DOI: 10.1016/j.cub.2017.08.052.

Stouraitis, T., E. Gkanias, J. M. Hemmi, and B. Webb (2017). "Predator evasion by a Robocrab." In: *6th International Conference on Biomimetic and Biohybrid Systems*. Vol. 10384. Stanford, CA: Springer, pp. 428–439. ISBN: 978-3-319-63536-1. DOI: 10.1007/978-3-319-63537-8_36. (Visited on 2017).

Strauss, R. (2002). "The central complex and the genetic dissection of locomotor behaviour." In: *Current opinion in neurobiology* 12.6, pp. 633–638. ISSN: 0959-4388. DOI: 10.1016/S0959-4388(02)00385-9.

Strutt, J. W. (1871). "XV. On the light from the sky, its polarization and colour." In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 41.271, pp. 107–120.

Stürzl, W., N. Boeddeker, L. Dittmar, and M. Egelhaaf (2010). "Mimicking honeybee eyes with a 280° field of view catadioptric imaging system." In: *Bioinspiration & Biomimetics* 5.3, p. 036002. ISSN: 1748-3190. DOI: 10.1088/1748-3182/5/3/036002.

Stürzl, W. (2017). "A lightweight single-camera polarization compass with covariance estimation." In: *Proceedings of the IEEE International Conference on Computer Vision*.
Stürzl, W. and N. Carey (2012). "A fisheye camera system for polarisation detection on UAVs." In: *European Conference on Computer Vision*. Springer, pp. 431–440.
Stürzl, W., J. Zeil, N. Boeddeker, and J. M. Hemmi (2016). "How wasps acquire and use views for homing." In: *Current Biology* 26.4, pp. 470–482. ISSN: 0960-9822. DOI: 10.1016/j.cub.2015.12.052.
Su, T.-S., W.-J. Lee, Y.-C. Huang, C.-T. Wang, and C.-C. Lo (2017). "Coupled symmetric and asymmetric circuits underlying spatial orientation in fruit flies." In: *Nature communications* 8.1, p. 139.
Sun, X., S. Yue, and M. Mangan (2020). "A decentralised neural model explaining optimal integration of navigational strategies in insects." In: *eLife* 9, e54026. DOI: 10.7554/elife.54026.
Sutton, R. S. (1988). "Learning to predict by the methods of temporal differences." In: *Machine learning* 3.1, pp. 9–44.
Sutton, R. S. and A. G. Barto (1990). "Time-derivative models of pavlovian reinforcement." In: *Learning and computational neuroscience, foundation of adaptive networks*. Ed. by M. Gabriel and J. Moore. Mit Press, pp. 497–537.
Sutton, R. S. and A. G. Barto (2018). *Reinforcement learning: an introduction*. 2nd ed. Vol. 84. Cambridge, MA: MIT press. ISBN: 9780262039246.
Szyszka, P., M. Ditzen, A. Galkin, C. G. Galizia, and R. Menzel (2005). "Sparsening and temporal sharpening of olfactory representations in the honeybee mushroom bodies." In: *Journal of Neurophysiology* 94.5, pp. 3303–3313. ISSN: 0022-3077. DOI: 10.1152/jn.00397.2005.
Szyszka, P., A. Galkin, and R. Menzel (2008). "Associative and non-Associative plasticity in Kenyon cells of the honeybee mushroom body." In: *Frontiers in Systems Neuroscience* 2, p. 3. DOI: 10.3389/neuro.06.003.2008.
Tabone, C. J. and J. S. d. Belle (2011). "Second-order conditioning in Drosophila." In: *Learning & Memory* 18.4, pp. 250–253. ISSN: 1072-0502. DOI: 10.1101/lm.2035411.
Takemura, S.-y. et al. (2017). "A connectome of a learning and memory center in the adult Drosophila brain." In: *eLife* 6, e26975. DOI: 10.7554/elife.26975.
Tanaka, N. K., H. Tanimoto, and K. Ito (2008). "Neuronal assemblies of the Drosophila mushroom body." In: *Journal of Comparative Neurology* 508.5, pp. 711–755. ISSN: 1096-9861. DOI: 10.1002/cne.21692.
Teague, M. R. (1980). "Image analysis via the general theory of moments*." In: *Journal of the Optical Society of America* 70.8, p. 920. ISSN: 0030-3941. DOI: 10.1364/josa.70.000920.
Thorndike, E. L. (1927). "The Law of Effect." In: *The American Journal of Psychology* 39.1/4, pp. 212–222.
Towne, W. F. (2008). "Honeybees can learn the relationship between the solar ephemeris and a newly-experienced landscape." In: *Journal of Experimental Biology* 211.23, pp. 3737–3743.
Tully, T. and W. G. Quinn (1985). "Classical conditioning and retention in normal and mutant Drosophila melanogaster." In: *Journal of Comparative Physiology A* 157.2, pp. 263–277. ISSN: 10534512. DOI: 10.1177/105345120103700110.
Turner-Evans, D., S. Wegener, H. Rouault, R. Franconville, T. Wolff, J. D. Seelig, S. Druckmann, and V. Jayaraman (2017). "Angular velocity integration in a fly heading circuit." In: *Elife* 6.

Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory*. New York: Springer-Verlag.
Vitzthum, H., M. Müller, and U. Homberg (2002). "Neurons of the central complex of the locust Schistocerca gregaria are sensitive to polarized light." In: *Journal of Neuroscience* 22.3, pp. 1114–1125.
Vosshall, L. B. (2007). "Into the mind of a fly." In: *Nature* 450.7167, pp. 193–197. ISSN: 1476-4687. DOI: 10.1038/nature06335.
Waddell, S. (2010). "Dopamine reveals neural circuit mechanisms of fly memory." In: *Trends in Neurosciences* 33.10, pp. 457–464. ISSN: 0166-2236. DOI: 10.1016/j.tins.2010.07.001.
Waddell, S., J. D. Armstrong, T. Kitamoto, K. Kaiser, and W. G. Quinn (2000). "The amnesiac gene product is expressed in two neurons in the Drosophila brain that are critical for memory." In: *Cell* 103.5, pp. 805–813.
Wang, Z., X. Chen, F. Becker, U. Greggers, S. Walter, M. Werner, C. Gallistel, and R. Menzel (2021). "Honey bees extract map coordinates from the dance." In: DOI: 10.21203/rs.3.rs-492542/v1.
Watkins, C. J. C. H. and P. Dayan (1992). "Q-learning." In: *Machine Learning* 8.3-4, pp. 279–292. ISSN: 0885-6125. DOI: 10.1007/BF00992698.
Webb, B. (2019). "The internal maps of insects." In: *The Journal of Experimental Biology* 222.Suppl 1, jeb188094. ISSN: 0022-0949. DOI: 10.1242/jeb.188094.
Wehner, R. and M. Müller (1993). "How do ants acquire their celestial ephemeris function?" In: *Naturwissenschaften* 80.7, pp. 331–333.
Wehner, R. (1997). "The ant's celestial compass system: spectral and polarization channels." In: pp. 145–185.
Wehner, R. (2001). "Polarization vision–a uniform sensory capacity?" In: *Journal of Experimental Biology* 204.14, pp. 2589–2596.
Wehner, R. (2003). "Desert ant navigation: how miniature brains solve complex tasks." In: *Journal of Comparative Physiology A* 189.8, pp. 579–588. DOI: 10.1007/s00359-003-0431-1.
Wehner, R., P. L. Herrling, A. Brunnert, and R. Klein (1972). "Periphere adaptation und zentralnervöse umstimmung im optischen system von Cataglyphis bicolor (Formicidae, hymenoptera)." In: *Revue suisse de Zoologie*. ISSN: 0035-418X.
Wehner, R. and T. Labhart (2006). "Polarization vision." In: *Invertebrate Vision*. Ed. by E. Warrant and D. E. Nilsson. Cambridge, UK: Cambridge University Press, pp. 291–348.
Wehner, R., B. Michel, and P. Antonsen (1996). "Visual navigation in insects: coupling of egocentric and geocentric information." In: *The Journal of experimental biology* 199.1, pp. 129–140. ISSN: 1477-9145.
Wehner, R. and F. Raber (1979). "Visual spatial memory in desert ants, Cataglyphis bicolor (hymenoptera: Formicidae)." In: *Experientia* 35.12, pp. 1569–1571.
Weiglein, A., J. Thoener, I. Feldbruegge, L. Warzog, N. Mancini, M. Schleyer, and B. Gerber (2021). "Aversive teaching signals from individual dopamine neurons in larval Drosophila show qualitative differences in their temporal "fingerprint"." In: *Journal of Comparative Neurology* 529.7, pp. 1553–1570. ISSN: 0021-9967. DOI: 10.1002/cne.25037.
Wessnitzer, J., J. M. Young, J. D. Armstrong, and B. Webb (2012). "A model of non-elemental olfactory learning in Drosophila." In: *Journal of Computational Neuroscience* 32.2, pp. 197–212. ISSN: 0929-5313. DOI: 10.1007/s10827-011-0348-6.

Wilkie, A., C. Ulbricht, R. F. Tobler, G. Zotti, and W. Purgathofer (2004). "An analytical model for skylight polarisation." In: *Rendering Techniques*, pp. 387–398.
Willshaw, D. J., O. P. Buneman, and H. C. Longuet-Higgins (1969). "Non-holographic associative memory." In: *Nature* 222.5197, pp. 960–962.
Wilson, R. I., G. C. Turner, and G. Laurent (2004). "Transformation of olfactory representations in the Drosophila antennal lobe." In: *Science* 303.5656, pp. 366–370. ISSN: 0036-8075. DOI: 10.1126/science.1090782.
Wittlinger, M., R. Wehner, and H. Wolf (2006). "The ant odometer: stepping on stilts and stumps." In: *Science* 312.5782, pp. 1965–1967. DOI: 10.1126/science.1126912.
Wolff, T., N. A. Iyer, and G. M. Rubin (2015). "Neuroarchitecture and neuroanatomy of the Drosophila central complex: A GAL4-based dissection of protocerebral bridge neurons and circuits." In: *Journal of Comparative Neurology* 523.7, pp. 997–1037.
Wu, C.-L., M.-F. M. Shih, P.-T. Lee, and A.-S. Chiang (2013). "An octopamine-mushroom body circuit modulates the formation of anesthesia-resistant memory in Drosophila." In: *Current Biology* 23.23, pp. 2346–2354. ISSN: 0960-9822. DOI: 10.1016/j.cub.2013.09.056.
Wu, Y., Q. Ren, H. Li, and A. Guo (2012). "The GABAergic anterior paired lateral neurons facilitate olfactory reversal learning in Drosophila." In: *Learning & Memory* 19.10. ISSN: 1072-0502. DOI: 10.1101/lm.025726.112.
Wu, Z. and A. Guo (2011). "A model study on the circuit mechanism underlying decision-making in Drosophila." In: *Neural Networks* 24.4, pp. 333–344. ISSN: 0893-6080. DOI: 10.1016/j.neunet.2011.01.002.
Wystrach, A. and S. Schwarz (2013). "Ants use a predictive mechanism to compensate for passive displacements by wind." In: *Current Biology* 23.24, R1083–R1085.
Wystrach, A., F. L. Moël, L. Clement, and S. Schwarz (2020). "A lateralised design for the interaction of visual memories and heading representations in navigating ants." In: *bioRxiv*. DOI: 10.1101/2020.08.13.249193.
Wystrach, A., S. Schwarz, P. Schultheiss, A. Baniel, and K. Cheng (2014). "Multiple sources of celestial compass information in the Central Australian desert ant Melophorus bagoti." In: *Journal of Comparative Physiology A* 200.6, pp. 591–601.
Xian, Z., X. Hu, J. Lian, L. Zhang, J. Cao, Y. Wang, and T. Ma (2014). "A novel angle computation and calibration algorithm of bio-inspired sky-light polarization navigation sensor." In: *Sensors* 14.9, pp. 17068–17088.
Yamagata, N., M. Hiroi, S. Kondo, A. Abe, and H. Tanimoto (2016). "Suppression of dopamine neurons mediates reward." In: *PLoS Biology* 14.12, e1002586. ISSN: 1544-9173. DOI: 10.1371/journal.pbio.1002586.
Yang, Z., J. Zhou, H. Huang, Y. Liu, and Z. Li (2017). "Measuring solar vector with polarization sensors based on polarization pattern." In: *Optik* 141, pp. 147–156. URL: http://dx.doi.org/10.1016/j.ijleo.2017.03.122.
Yarali, A., T. Hendel, and B. Gerber (2006). "Olfactory learning and behaviour are 'insulated' against visual processing in larval Drosophila." In: *Journal of Comparative Physiology A* 192.10, pp. 1133–1145. ISSN: 14321351. DOI: 10.1007/s00359-006-0140-7.
Yarali, A., M. Mayerle, C. Nawroth, and B. Gerber (2008). "No evidence for visual context-dependency of olfactory learning in Drosophila." In: *Naturwissenschaften* 95.8, pp. 767–774. ISSN: 00281042. DOI: 10.1007/s00114-008-0380-1.
Yin, Y., N. Chen, S. Zhang, and A. Guo (2009). "Choice strategies in Drosophila are based on competition between olfactory memories." In: *European Journal of Neuro-*

science 30.2, pp. 279–288. ISSN: 0953816X. DOI: 10.1111/j.1460-9568.2009.06821.x.

Young, J. M. and J. D. Armstrong (2010). "Building the central complex in Drosophila: The generation and development of distinct neural subsets." In: *Journal of Comparative Neurology* 518.9, pp. 1525–1541. ISSN: 00219967. DOI: 10.1002/cne.22285.

Young, J., J. Wessnitzer, J. Armstrong, and B. Webb (2011). "Elemental and non-elemental olfactory learning in Drosophila." In: *Neurobiology of Learning and Memory* 96.2, pp. 339–352. ISSN: 1074-7427. DOI: 10.1016/j.nlm.2011.06.009.

Yu, D., A. C. Keene, A. Srivatsan, S. Waddell, and R. L. Davis (2005). "Drosophila DPM neurons form a delayed and branch-specific memory trace after olfactory classical conditioning." In: *Cell* 123.5, pp. 945–957. ISSN: 0092-8674. DOI: 10.1016/j.cell.2005.09.037.

Zeil, J. (2012). "Visual homing: an insect perspective." In: *Current Opinion in Neurobiology* 22.2, pp. 285–293. ISSN: 09594388. DOI: 10.1016/j.conb.2011.12.008.

Zeil, J., M. I. Hofmann, and J. S. Chahl (2003). "Catchment areas of panoramic snapshots in outdoor scenes." In: *Journal of the Optical Society of America A* 20.3, p. 450. ISSN: 1084-7529. DOI: 10.1364/josaa.20.000450.

Zhang, S., H. Liang, H. Zhu, D. Wang, and B. Yu (2014). "A camera-based real-time polarization sensor and its application to mobile robot navigation." In: *Robotics and Biomimetics (ROBIO), 2014 IEEE International Conference on*. IEEE, pp. 271–276.

Zhang, W., Y. Cao, X. Zhang, and Z. Liu (2015). "Sky light polarization detection with linear polarizer triplet in light field camera inspired by insect vision." In: *Applied optics* 54.30, pp. 8962–8970.

Zhang, W., Y. Cao, X. Zhang, Y. Yang, and Y. Ning (2017). "Angle of sky light polarization derived from digital images of the sky under various conditions." In: *Applied Optics* 56.3, pp. 587–595.

Zhang, Y. V., J. Ni, and C. Montell (2013). "The molecular basis for attractive salt-taste coding in Drosophila." In: *Science* 340.6138, pp. 1334–1338. ISSN: 0036-8075. DOI: 10.1126/science.1234133.

Zhao, C., Y. F. Widmer, S. Diegelmann, M. A. Petrovici, S. G. Sprecher, and W. Senn (2021). "Predictive olfactory learning in Drosophila." In: *Scientific Reports* 11.1, p. 6795. DOI: 10.1038/s41598-021-85841-y.

Zhao, F., Y. Zeng, A. Guo, H. Su, and B. Xu (2020). "A neural algorithm for Drosophila linear and nonlinear decision-making." In: *Scientific Reports* 10.1, p. 18660. DOI: 10.1038/s41598-020-75628-y.

Zhou, M. et al. (2019). "Suppression of GABAergic neurons through D2-like receptor secures efficient conditioning in Drosophila aversive olfactory learning." In: *Proceedings of the National Academy of Sciences* 116.11, p. 201812342. ISSN: 0027-8424. DOI: 10.1073/pnas.1812342116.

Zhu, L., M. Mangan, and B. Webb (2020). "Spatio-temporal memory for navigation in a mushroom body model." In: Biomimetic and Biohybrid Systems. Freiburg, Germany, pp. 415–426. ISBN: 9783030643126. DOI: 10.1007/978-3-030-64313-3_39.

Zlotnik, G. and A. Vansintjan (2019). "Memory: an extended definition." In: *Frontiers in Psychology* 10, p. 2523. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2019.02523.

Zollikofer, C., R. Wehner, and T. Fukushi (1995). "Optical scaling in conspecific Cataglyphis ants." In: *Journal of Experimental Biology* 198.8, pp. 1637–1646. ISSN: 0022-0949. DOI: 10.1242/jeb.198.8.1637.