

NORTHWESTERN UNIVERSITY

On Music Theory Expertise: A Cognitive Framework and Galant Schema Theory Case Study

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL  
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Music

in the program of Music Theory and Cognition

By

Sarah Lindsay Gates

EVANSTON, ILLINOIS

March 2023

© Copyright by Sarah Gates 2023

All Rights Reserved

## ABSTRACT

On Music Theory Expertise: A Cognitive Framework and Galant Schema Theory Case Study

Sarah Gates

Every music theorist has their own personal reasons for engaging in the craft of theorizing about music. Many discover that there is something special about a theoretic disposition which differs from other sorts of interactions with music. It is often this special something that draws musicians into the discipline: specifically, that some form of active reflection about music—its components, its structure, and/or its meaning—greatly expands not only our conceptual understanding of music, but our experience of it. In pedagogical circles, this is referred to as the inseparable bond between thinking and listening, which functions as an iterative feedback loop whereby one activity both gains from and informs the other (Rogers 2008). In this way, active reflection about music changes how we hear, and how we hear changes our process of active reflection.

In this dissertation, I create a cognitive account of this iterative feedback loop between thinking and listening, using Galant schema theory (Gjerdingen 2007) in the context of Sonata Form analysis (Hepokoski and Darcy 2011) as a test domain. I propose that iterative analytical interactions with music are performed to foster the development of memory skill (long-term working memory, Ericsson & Kintsch 1995; Ericsson 2018), defined as expertise in situated conceptualization or simulation (Barsalou 2003a,b)—essentially, a form of expertise in the development and use of music categories. In the first chapter, I discuss music theoretic literature on music theory expertise. Here, I reinterpret the discourse on the thinking—listening dichotomy as one that centers imagery and verbalization, and their interaction, as central to developing music theoretic expertise. I situate the project within prior music cognition scholarship and

outline the primary research questions and goals. In the second chapter, I create a cognitive framework for understanding expert music concept representation and categorization by integrating Allan Paivio's Dual-Coding Theory (1986; 2007) with Barsalou's Dynamic Interpretation in Perceptual Symbols System (2003a,b). In the third chapter, I apply this framework to create an account of embodied Galant schema representation. Here I argue that experts' schema representations differ from those of nonexperts. Schema concepts acquired by music theorists are formed through repetitive interactions using various music theoretic concepts to deliberately encode a schema's multiple features and their relationships. This results in representational pools (simulators) that are modally distributed, highly structured, and multi-coded (across language and sensory systems), facilitating control over attention during online cognition. In the fourth chapter, I provide a hypothetical developmental trajectory for Galant schema acquisition by integrating historical and modern pedagogical practices with a memory expertise framework (Gates 2021). The final chapter presents results from a qualitative survey which provides a concrete demonstration of expert music-theoretic memory skill, the formation and modification of an interpretation in perception, in the context of musical ambiguous figure perception. I conclude by recontextualizing claims from chapter 1 using the framework developed through the course of the dissertation and discuss avenues for future research.

## ACKNOWLEDGMENTS

I would not have been able to complete this project without the support of many people. Firstly, I would like to thank my advisor, Richard Ashley, for all of his encouragement and support over the years. I am so grateful to have an advisor that supported me equally as a person and as a scholar. His vast knowledge of many different fields, methodologies and sources was so very motivating, and was vital to the success of this project. I am continuously inspired by his endless curiosity in engaging different ideas, and I am so thankful for his patience and enthusiasm in entertaining the *many* topics (too many?) in which I was interested throughout my degree. This process was not an easy one for me, and he has been there to support me every step of the way.

I would like to thank the other members of my committee. To Vasili Byros, who has always been a supporter of my work even when I was having the greatest of doubts, thank you. I have always admired the depth, precision, and sincerity of your work. Thank you for fostering the music theorist in me. To Robert Reinhart, the resident pedagogue on this committee: teaching in your sophomore class reaffirmed my understanding and appreciation of forming different analytical interpretations. It was a constant reminder of the value of music theory as a pedagogical tool, which helped me greatly in conceptualizing this project.

Next, I would like to thank many of the amazing music educators in my life, to whom I owe so much. I would not be here without your dedication and support over the years—I truly believe that a good teacher can make all the difference in the world. To my junior high school teacher, Dr. Tracy Fewster: thank you so much for seeing me, believing in me, and encouraging me. I felt like you were the first teacher who was really *there* for me. I knew that if you thought I could do it, then I could do it. All I needed to do was try! To my high school teachers, Barbara

Brown, Pattie Bender, Kerry Smith, and Kirsty Hunsberger: can you believe you actually made high school enjoyable? The music rooms were some of the few places I felt at home then.

To my many undergraduate professors at Wilfrid Laurier, thank you! To my very first saxophone teacher, David Wiffen, and accompanist Lorin Shalanko, who supported a first-generation university student through the trials and tribulations of applications, auditions, competitions, and the degree itself. Thank you. To my composition teachers—Glen Buhr, Peter Hatch, and Linda Caitlin-Smith—thank you for fostering my musical creativity and helping me find my musical voice. To my first-year aural skills and music theory professor, Kevin Swinden: I still remember my very first class of my undergraduate, 8am aural skills. Your precision, rigor, and dedication as a teacher, and your true joy for music theory is one of the reasons I am where I am today. You inspired me to pursue graduate training in music theory. You introduced me to musical duck-rabbits, and now I wrote a dissertation about it. Hope you still have that book! So honored that we are now colleagues and friends.

To the many other professors during my graduate studies—Wallace Halladay, Peter Schubert, William Caplan, Robert Hasagawa, Nicole Biamonte, Stephen McAdams, Jonathan Wild, Robert Gjerdingen, Mark Butler—thank you for pushing me and believing in me.

The last two educators I would like to thank are my most recent. Firstly, Susan Piagentini—thank you so much for your support, kindness, mentorship, and friendship over the years. You have been one of the core supporters for my love of music theory pedagogy, and I owe *so* much of the formation of this project to you. Secondly, Steven Demorest—though you are no longer here, the impact you had and continue to have on me as a person and a scholar cannot be overstated. You taught me how to ask questions and to find ways to solve them. I owe my passion for experimental design, data analysis and statistics to you. #stronginferenceforlife

I would also like to thank so many of my friends and colleagues. To my peers, Stefan Greenfield-Casas, Morgan Patrick, Lena Console, Stephen Hudson, Cella Westray, Aubrey Leaman, thank you for being such amazing colleagues and friends. A special thank you to Miriam Piilonen for her mentorship throughout my time at Northwestern. To my fellow Northwestern graduates and mentors, Bruno Alcalde, Rosa Abrahams, Olga Sanchez-Kisieleska, and Janet Bourne, thank you so much for showing me what is possible. Another special thank you to Vivian Luong—we first met in 8am aural skills, and we’re still here! You are such an inspiring and humble scholar. And to Adam White: thanks for making the degree truly enjoyable, our Wednesday coworking zoom sessions made completion of this dissertation possible.

To Anjni—my peer, cohort, and dearest friend—where do I even begin? I expected to finish this program with a degree in hand, and what I really got was a lifelong friendship. When I say I could not have finished this without you, I truly mean it. You are the Fawkes to my Dumbledore.

*À mon amour, Guillaume. Tu as tout vu—les rires, les larmes, les hauts et les bas. Tu m’as supporté à travers tout. Quand je pensais que c’était impossible, tu étais là, pour dire <Non! Tu es capable!> Merci pour les câlins, la nourriture, et ta confiance envers moi. Tu es la joie dans ma vie.*

This dissertation is dedicated to those living with chronic illness:

Though the journey may not go how you want it, know that you are capable. You are strong. All it takes is one step at a time.



## Table of Contents

Chapter 1 Background, Project Methodology and Scope .....	21
Music Theory Expertise as an Iterative Feedback Loop Between “Thinking” and “Listening” .....	21
‘Thinking’ and ‘Listening’ as Verbal and Imagery Processes .....	27
On Cooperative Independence: Synthesis or Dissociation? .....	31
Explanatory Approaches in Music Cognition.....	43
Effects of Music Theory Training.....	48
Project Methodology and Scope .....	52
Which “Music Theory?” .....	52
The Descriptive-Suggestive Distinction: Introspection versus Perception.....	53
On a Galant Schema-Theory Case Study: Acquiring Eighteenth Century Hearing .....	55
Guiding Research Questions and Claims.....	57
Dissertation Outline .....	59
Chapter 2 Dual-Coding Theory: Overview, Updates and Adaptation .....	62
Dual Coding Theory: An Overview.....	63
Representational Systems and Units.....	64
Logogens.....	66
Imagens .....	67
Connections and Activation Processes: A DCT Approach to Meaning .....	68
Representational activation.....	69
Associational activation .....	71
Referential activation .....	73
Dual Coding Functions .....	73
Bootstrapping and improving memory: code additivity and the conceptual peg hypothesis .....	74
Representational exchange and creation: within-system association and intermodal transfer .....	76
Paivio on Musical Expertise .....	77
Adapting and Updating DCT: Defining an Approach to Concepts and Categorization.....	80
Barsalou’s DIPSS and DCT .....	82
Reconciling DIPSS and DCT: Language, Introspection, and Abstraction .....	87
The concrete-abstract continuum: on the role of language and introspection .....	88

The structure and content of long-term memory .....	95
Long-term memory structure. ....	96
Long-term memory content. ....	99
Situated conceptualization and working memory .....	106
Re-Defining Categorization: Situated Conceptualization as Spreading Activation in DCT .....	113
Accounting for Introspection: Imagery, Feeling of Knowing, Emotional Construction and Interoception .....	122
On Sensations of Memory: Priming, Item Pre-Selection, and “Feeling of Knowing” .....	128
Emotional Construction and Interoception .....	132
DCT Updated: A Conclusion and Summary .....	140
Chapter 3 An Embodied Approach to Galant Schema Representation .....	142
When ‘Schemata’ are not Enough: Issues in Abstract Unitary Representation.....	143
Novice versus Expert Representation: Loose versus Structured Representation .....	147
What is a Galant Schema? The “Conceptual Peg” Account.....	149
Loose Holistic Representation: The Novice .....	151
On Representation: An Account of Statistical Learning.....	152
On Simulation: Activational Pathways and Limitations.....	160
Structured Distributed Representation: The Expert.....	170
On Representation: Music Theoretic Concepts as Simulators.....	171
Representing declarative knowledge in the verbal system .....	173
Property simulators .....	176
Scale degrees.....	177
Harmony. ....	183
Relation simulators .....	187
Counterpoint. ....	188
Tonality and harmonic function.....	190
Forms and formal function.....	194
On Simulation: Flexibility in Retrieval and Maintenance .....	196
Summary and Conclusions .....	206
Chapter 4 Galant Schema Acquisition as Memory Expertise: Acquiring Eighteenth Century Hearing.....	207
Memory Expertise Defined: Ericsson and Kintsch (1995) Long-Term Working Memory....	207

Dual-Coding and Situated Simulation as Memory Expertise: Insights from Gates (2021)....	210
The “Loop” as LTWM Acquisition .....	211
Memory Skill and Category Learning: Historical and Modern Pedagogical Approaches Compared.....	213
Schemata Learning as Memory Skill Acquisition: A Historical Perspective .....	214
Training Types and their Functions .....	215
Solfeggio. ....	217
Partimenti, counterpoint, and composition. ....	220
Gradual Memory Elaboration Through Progressive Variation.....	224
Modern North American Music Theory Training: A General Account of Contemporary Galant Schemata Acquisition.....	228
Re-aligning Simulators through Schemata ‘Prototypes’ and Gradual Memory Elaboration .....	232
Acquiring LTWM Through Analysis .....	252
Summary and Conclusions .....	267
Chapter 5 Expertise in Action: LTWM as Control Over Simulation .....	269
Issues in Top-Down Effects in Interpretation and Perception: Hazy Boundaries Between Perception and Cognition.....	269
Control over Interpretation in Perception: Musical Ambiguous Figures and the Case of the Modulating Prinner .....	271
Modulating Prinner and Sonata Form: An Overview .....	272
A Mozart Case Study: Two Transitional Schema Options.....	276
Modulating Prinner .....	277
Piano Sonata no. 2, K. 280, iii. ....	278
Piano Sonata no. 10, K. 309, i.....	281
Step-Descent Fauxbourdon Romanesca .....	283
Piano Sonata no. 15, K. 576, i.....	285
Piano Sonata no. 14, K. 310, iii. ....	288
An Experimental Verification: Effects of Attention, Memory, and Expertise on Bistable Perception in Mozart’s Piano Sonata no. 2, K. 280, iii.....	290
Guiding Questions and Hypotheses .....	298
Method .....	309
Participants.....	310
Stimuli.....	311
Procedure. ....	313

Data cleaning and pre-processing .....	315
Results.....	317
Ease of hearing ratings (DV1). .....	318
Ease of change of interpretation ratings (DV2). .....	324
Excerpt familiarity and expertise. ....	326
Discussion .....	328
Summary and Conclusions .....	334
Chapter 6 A Project Recapitulation and Postscript.....	335
Operationalizing Claims from Chapter 1 .....	339
Future Research .....	344
Works Cited .....	347

## List of Figures

Figure 1.1. Proposed iterative relationship between thinking and listening (modified from Rogers 2004, 8). .....	23
Figure 2.1. Representations in each symbolic system by modality (Paivio 2007, 36) .....	66
Figure 2.2. Diagram of The Subsystems and their Relations (Paivio 1986, 67).....	69
Figure 2.3. Dominance Order of Simulators (Barsalou 2005).....	86
Figure 2.4. Approach to Storage and Retrieval (Simulation) from Barsalou (1999).....	100
Figure 2.5. Example of Mental Model for Phrase 'Royal Wedding' (Sadoski and Paivio 2013, 58) .....	108
Figure 2.6. New Connector Types for LTM.....	109
Figure 2.7. Baddeley's Working Memory Model (2000).....	111
Figure 2.8. New DCT Layout Showing Input (top), LTM (middle) and WM (bottom) .....	112
Figure 2.9. LTM Trace Colours Showing Activation Types .....	113
Figure 2.10. Recognition, Identification, and Categorization as Spreading Activation in DCT	115

Figure 2.11. Initial Recognition of a Visual Stimulus through Representational Activation. ....	116
Figure 2.12. Identification of Apple through Referential Activation .....	117
Figure 2.13. Availability of Categorization through Referential and Associational Activation	118
Figure 2.14. Two Types of Simulation. Categorization Decision for "Fruit" (a) and Imagined Taste (b) .....	119
Figure 2.15. Recognition (a), Identification (b) and Categorization Availability (b) from Verbal Representational Activation.....	121
Figure 2.16. Vividness Scale Added to WM .....	127
Figure 2.17. Control Scale Added to WM Showing Ease of Change of a Simulation .....	127
Figure 2.18. Relationship Between Various Metacognition Judgements and Learning Processes (Neilson and Narens 1990, 129) .....	129
Figure 2.19. Feeling-of-Knowing (FoK) During a Tip-of-the-Tongue State .....	130
Figure 2.20. Sparse Simulator with Low FoK .....	131
Figure 2.21. Elaborated Simulator with High FoK.....	132
Figure 2.22. Simulation of Emotional Construction "Hungry" .....	136
Figure 2.23. Simulation of Emotional Construction "Anxious" .....	136
Figure 2.24. Misinterpretation of Anxiety for Hunger .....	138
Figure 2.25. Elaborated Simulators for Hunger and Anxiety .....	139
Figure 3.1. Schema Stages for Prinner Prototype (graphic from Gjerdingen 2007, 455).....	150
Figure 3.2. Prinner Schema Stages as Conceptual Pegs in DCT .....	151
Figure 3.3. Auditory Imagen Organization of K. 545 for a Novice Listener .....	155
Figure 3.4. Auditory Imagen Information of K. 545 for an Encultured Listener .....	156
Figure 3.5. Modally Distributed Simulator for K. 545 .....	157

Figure 3.6. Modally Distributed Simulator for K. 545 (Live performance) .....	158
Figure 3.7. Simulator for K. 545 Including Referential Connections to Verbal System .....	159
Figure 3.8. Direct Representational Activation of Auditory Trace (a) and Motor Trace (b) .....	161
Figure 3.9. Chunk Access During Simulation (Imagery). Unable to jump (a), Sequential Associational Activation (b), Direct Representational Activation (c) and Sequential Maintenance (d).....	164
Figure 3.10. Category-Level Activation of Exemplars in Exemplar Pools Through Representational Activation.....	167
Figure 3.11. Goodness of Fit Ratings for Prinner Schemata and Variations .....	169
Figure 3.12. Imagery Completion for Original Prinner and Alternate .....	170
Figure 3.13. Verbal Associations for Prinner Concept .....	175
Figure 3.14. Prototypical Features of Prinner Schema from Gjerdingen (2007).....	175
Figure 3.15. Encoding of Visual Imagen for Bassline Simulator .....	179
Figure 3.16. Encoding of Associated Auditory Imagen for Bassline .....	181
Figure 3.17. Encoding Auditory Imagen with Auditory Attention to Bassline .....	181
Figure 3.18. Encoding Process Repeated for the Soprano Line .....	182
Figure 3.19. Figured Bass Harmony Simulator and Associations to Scale Degree Simulators .	184
Figure 3.20. Harmony Simulator Including Roman Numerals.....	186
Figure 3.21. Elaborated Harmony Simulator with Additional Interactions and Traces .....	186
Figure 3.22. Scale Degree Property Simulator and Counterpoint Relation Simulator .....	189
Figure 3.23. Harmonic Function added as Relation Simulator.....	192
Figure 3.24. Rule of the Octave Simulator with Harmonic Function.....	193
Figure 3.25. Prinner Simulators for K. 545 .....	195

Figure 3.26. Form and Formal Function (Verbal System) as Relation Simulators for Retrieval on Nonverbal Side.....	196
Figure 3.27. Simulation of Middle Chunk of K. 545 Through Visual Representational Activation (a) and Auditory Association (b) .....	198
Figure 3.28. Verbal Representational Activation (a) and Referential Activation (b).....	199
Figure 3.29. Simulation of Soprano Line in Context Using Prinner Simulator.....	200
Figure 3.30. Access to End of Auditory Imagen Through Referential Activation (a) and Auditory Scanning (b) .....	201
Figure 3.31. Representational Exchange Between Verbal and Nonverbal Units .....	202
Figure 3.32. Abstract Schema Shapes.....	203
Figure 3.33. Visual Priming and Referential Activation in Prinner Simulators .....	204
Figure 3.34. Visual Priming and Referential Activation in Romanesca Simulators .....	204
Figure 3.35. Associational Activation and Auditory Simulation of Prinner Bassline .....	205
Figure 3.36. Selection and Simulation of Familiar Prinner Exemplar.....	205
Figure 4.1. Sample Retrieval Structure (from Gates 2021) .....	209
Figure 4.2. Aural Skills Activities as LTWM Development (from Gates 2021).....	211
Figure 4.3. LTWM Development Updated for Music Theory Expertise in the Current Framework .....	213
Figure 4.4. Traditional Training Activities as LTWM Acquisition.....	216
Figure 4.5. Solfeggio Training Types from Baragwanath (2020) as LTWM Acquisition .....	218
Figure 4.6. Partimenti and Counterpoint Training Types as LTWM Acquisition.....	221
Figure 4.7. Three similar introductory solfeggio exercises from Levesque and Bèche (1779)..	226

Figure 4.8. Durante <i>Partimenti diminuti</i> , three ways of embellishing the ascending half step in the bass from the leading tone to its tonic (Durante and Gjerdingen, “Partimenti Diminuti.” From Monuments of Partimenti, <a href="https://partimenti.org/">https://partimenti.org/</a> ) .....	227
Figure 4.9. Sketch and refinement of two eyes by Bernard Julien (Gjerdingen 2020, 277).....	228
Figure 4.10. Episodic Memory Blurring (“Generalization”) Over the Course of Several Memory Episodes .....	232
Figure 4.11. Example 6.2 From Laitz (2016, 189) showing Stages of Embellishment.....	233
Figure 4.12. Score and Reduction of Beethoven’s Piano Sonata in D minor, “Tempest,” op 31, no.2, Allegretto, from Laitz (2016, 192-193) .....	234
Figure 4.13. Galant Schemata Prototypes from Open Music Theory .....	235
Figure 4.14. Romanesca Prototype from the Prototype Appendix (Gjerdingen 2007, 454). .....	235
Figure 4.15. Three Learning Interactions with Music in the Galant Style (2007) in Memory Episode 1.....	238
Figure 4.16. Visual and Verbal Interaction with Prinner Prototype from Gjerdingen (2007)....	239
Figure 4.17. Interaction with Example 3.1 from Gjerdingen (2007) .....	240
Figure 4.18. Interaction with Example 3.2 from Gjerdingen (2007) .....	241
Figure 4.19. Visual Fixation on Prinner Prototype Stages and Association with Prototypical Prinner Logogen (La-Sol-Fa-Mi).....	243
Figure 4.20. Addition of La-Sol-Fa-Mi Logogen and Visual Fixation Activates Similar Interactions in Previous Memory Chunks.....	243
Figure 4.21. Memory Episode 2 Involving Review and Application of Previously Learned Material .....	244
Figure 4.22. Use of Stored Prinner Logogens to Direct Writing Out of Prinner Schema .....	245



Figure 4.23. Memory Trace for Newly Created Prinner in Auditory, Visual and Motor Modalities (Piano Performance) .....	246
Figure 4.24. Memory Episode 4 Involving Recall and Creation of New Traces.....	247
Figure 4.25. Schema Recall Practice in the Auditory Modality .....	248
Figure 4.26. Episodic Memory Blurring Across the Four Memory Episodes .....	248
Figure 4.27. Category-Level Prinner Traces for Novice (a), Intermediate (b) and Expert (c) levels .....	251
Figure 4.28. Category Level Representation for Prinner Subtypes .....	252
Figure 4.29. Four Analysis Sessions for K. 545 .....	254
Figure 4.30. Review of K. 545 Schema Chart from Gjerdingen (2007).....	255
Figure 4.31. Review of K. 545 Score Analysis in Gjerdingen (2007).....	256
Figure 4.32. Formal Score Analysis of K. 545 (by hand).....	257
Figure 4.33. Resulting Memory Traces from K. 545 Score Analysis.....	258
Figure 4.34. Addition of Schema Labels onto Formal Analysis of K. 545 .....	259
Figure 4.35. Resulting LTM Trace from Addition of Prinner Analyses in K. 545 .....	259
Figure 4.36. Activation of Schema Pools Over Time in the Transition of K. 545 Showing Prinner (a), Indugio (b) and Ponte(c) Activation.....	262
Figure 4.37. Lack of Representational Activation for Sequential Prinner in K. 545.....	263
Figure 4.38. Recoding Prinner Simulator Through Addition of Soprano Voice Trace .....	264
Figure 4.39. Activation of Prinner Traces Through Solmization After Recoding.....	265
Figure 4.40. Prinner Subtype Exemplars and Associations with Formal Sections and Functions .....	266
Figure 5.1. Figure 27.1 from Gerjdingen (2007, 372) .....	275

Figure 5.2. Embedded Prinner and Converging in an Indugio Schema from Byros (2011).....	276
Figure 5.3. Expositonal Transition of K. 280, ii, 13-37.....	279
Figure 5.4. Recapitulatory Transition of K. 280, iii, 120-148 .....	280
Figure 5.5. Expositonal Transition from K. 309, i, 18-32 .....	282
Figure 5.6. Recapitulatory Transition of K. 309, i, m 113-126 .....	283
Figure 5.7. Expositonal Transition of K. 576, i, 16-27.....	286
Figure 5.8. Recapitulatory Transition of K. 576, i, 101-121 .....	287
Figure 5.9. Expositonal Trimodular Block (TMB) in K. 310, iii, 19-63.....	288
Figure 5.10. Recapitulatory Transition of K. 310, iii, 190-221 .....	290
Figure 5.11. Modulating Prinner Interpretations in the Exposition (a) and Recapitulation (b) of K. 280, iii .....	294
Figure 5.12. Step-Descent Romanesca Interpretations for the Exposition (a) and Recapitulation (b) of K. 280, iii, m 120-148.....	295
Figure 5.13. Monotonal Prinner Schemata Interpretation for the Recapitulatory Transition of K. 280, iii, m 120-148.....	297
Figure 5.14. Direct Representational Activation of Prinner Traces and Partial Priming of Romanesca Traces in K. 280, iii, Exposition.....	301
Figure 5.15. Direct Representation Activation of Prinner Traces and Increased Activation of Romanesca Traces in K. 280, iii, Recapitulation.....	302
Figure 5.16. Formation of Modulating (a) and change to Nonmodulating (b) Bass Line Interpretations in K. 280, iii, Exposition.....	304
Figure 5.17. Formation of Modulating Prinner (a) and Step-Descent Romanesca (b) Schema Interpretations in K. 280, iii, Exposition.....	306

Figure 5.18. Direct Representational Activation of Modulating Prinner Traces of a Hypothetical Expert.....	308
Figure 5.19. Suppression of Modulating Prinner Traces and Attempted Indirect Activation of Step-Descent Romanesca Traces of a Hypothetical Expert.....	309
Figure 5.20. Expositional Transition of K. 280, iii.....	311
Figure 5.21. Recapitulatory Transition of K. 280, iii .....	312
Figure 5.22. Visually Presented Interpretations for the Bass Line (a), Soprano Line (b) and Schemata (c) Interpretations of the Exposition.....	313
Figure 5.23. Duck-Rabbit Ambiguous Figure from Jastrow (1899).....	314
Figure 5.24. Bass Line Interpretations for K. 309 used in the Survey Introduction.....	314
Figure 5.25. Average Schema Hearing and Scale Degree Hearing Ratings for Each Expertise Group .....	317
Figure 5.26. Average Ease of Hearing Ratings (DV1) by Expertise Group.....	319
Figure 5.27. Average Ease of Hearing Ratings (DV1) by Attended Feature .....	319
Figure 5.28. Average Ease of Hearing Ratings (DV1) for Attended Features by Sonata Section .....	321
Figure 5.29. Average Ease of Hearing Ratings (DV1) for Modulation Type by Attended Feature .....	321
Figure 5.30. Average Ease of Hearing Ratings (DV1) for Attended Features by Expertise Group .....	323
Figure 5.31. Average Ease of Hearing Ratings (DV1) for Modulation Type by Expertise Group .....	323
Figure 5.32. Average Ease of Change Ratings (DV2) by Expertise Group .....	325

Figure 5.33. Soprano Voice Imagery During Simulation for Intermediate and Expert Groups . 331

Figure 5.34. Average Ease of Hearing Ratings (DV1) for Modulation Type by Attended Feature  
for Schemata Experts ..... 332

## **List of Tables**

Table 2.1. Updated Sensorimotor Systems in DCT ..... 135

Table 5.1. Excerpt Familiarity and Analysis Responses by Expertise Group ..... 327

Table 5.2. Excerpt Familiarity Categories by Expertise Group..... 327

Table 5.3. Interpretation Formation Types by Expertise Group..... 334

## Chapter 1

### Background, Project Methodology and Scope

In this chapter, I will discuss the background, methodology and scope of this dissertation. In the first portion of this chapter, I discuss the idea that motivated the project: music theory expertise positioned as an iterative loop between thinking and listening. I outline selected scholarship that discusses this phenomenon, and then propose that the thinking-listening dichotomy be understood as interactions between verbalization and imagery processes, respectively. I then discuss the ways in which verbalization and imagery may either be synthesized into a new form of perceptual ability, or be separated, leading to dissociation and fragmentation. I conclude this section by discussing scholarship in music cognition that offers some explanatory mechanisms for music theory expertise. In the second half of the chapter, I outline this project's methodology and scope, which is in a similar vein to clarify previous music theory/cognition frameworks (e.g., Zbikowski 2002). Building on existing research in mental representation, imagery, categorization, memory, and expertise, I set forth a cognitive theoretical framework capable of elucidating the iterative loop between thinking and listening in music theory expertise. The project uses expertise in Galant schemas as a focal domain and concludes with a quantitative study demonstrating the effects of expertise on schemas which can be understood as musically ambiguous figures.

#### Music Theory Expertise as an Iterative Feedback Loop Between “Thinking” and “Listening”

Music theory expertise is often viewed as an iterative process involving thinking and

listening. Thinking or contemplation is done to inform, clarify, and/or modify listening activities and experiences, and listening is done to inform, clarify, and/or modify thinking activities. One of the clearest expositions of this position in the literature of music theory pedagogy is found in Rogers (2004), who explicitly positions the practice and pedagogy of music theory as a loop between thinking and listening (see Figure 1.1). Rogers proposes that the goal of music theory pedagogy should be the integration or fusion of thinking and listening, which provides a foundation for a lifetime process or set of attitudes for responding to music (Rogers 2004, 8). This process is continual and iterative, spurring growth in knowledge, understanding, and music experience. Rogers states that "...each part of this thinking/listening duality feeds into the other: the more thinking that takes place, the more there is to hear; the more listening that takes place, the more there is to ponder" (ibid.). He also notes that developing this expertise is a "life-long process," one akin to other time- and practice-intensive musical disciplines:

Music theory, in my opinion, is not a subject like pharmacy with labels to learn and prescriptions to fill but is an activity—more like composition or performance. The activity is theorizing: i.e., thinking about what we hear and hearing what we think about—and I would include even thinking about what we think (Rogers 2004, 7-8).

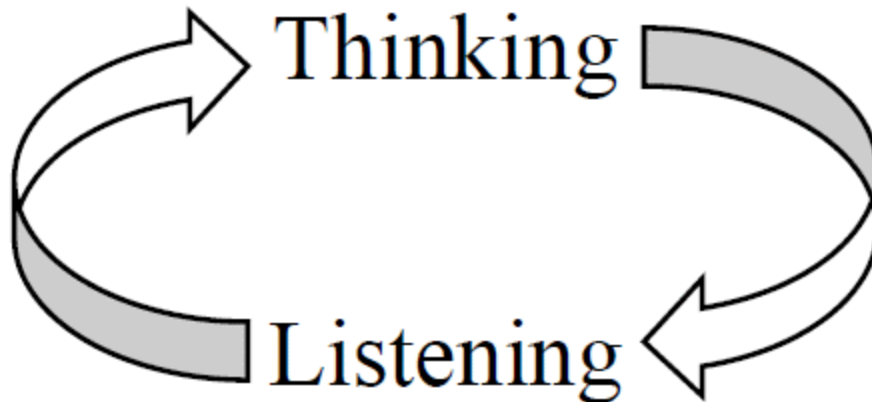


Figure 1.1. Proposed iterative relationship between thinking and listening (modified from Rogers 2004, 8).

Other writers offer related but different perspectives. Gordon's (2012) theory of developing *audiation* (the process of assigning meaning to sound, Gordon 2012, 3) notes the reciprocal and iterative nature of musical this musical skill:

In other words, when you are audiating as you are listening to music, you are summarizing and generalizing content of music patterns in the context you just heard as a way to anticipate or predict what will follow. Every action becomes an interaction. What you are audiating depends on what you have audiated. As audiation develops, it becomes broader and deeper, and thus, reflects more on itself (Gordon 2012, 5).

Meyer (1973) notes the iterative and exhaustive nature of critical analysis:

But no matter how inclusive and detailed a critical analysis is, it is seldom exhaustive, and it is never definitive. It is seldom exhaustive because most pieces

of serious music are complex. Consequently, it is almost always possible to discover relationships not previously observed. The critical analysis of a particular work is never definitive because the theory of music and those of related disciplines such as psychology are likely to change. And because it is partly dependent upon such theories, analytic criticism will probably change too...[and] Though in all probability they will subsequently be revised, or even rejected, such works and theories endure because they are exciting and seminal: they lead to new discoveries and further formulations, and thereby continue to affect language, thought and behavior. (p 24-25)

Schenker (1979) discusses the cultivation of musical understanding as an iterative process, referencing the long and slow acquisition of this skill. While his goal is not solely limited to acquiring theoretic skill but also skill in composition, Schenker nevertheless positions mastery of organic coherence at the center of musical creation and comprehension (p. xxiv). Mastery of organic coherence is positioned to be a slow and deeply iterative process and is placed in direct opposition to modern condensed courses in language and music (p. xxiii). Through the creations of reductional graphic representations, which were considered by Schenker to be a distilled form of the true nature of the composition, one is able to comprehend the points of structural significance of a given work, and thus have understanding of the piece in a similar way to that of the (genius) composer who created it. While this approach is much more concerned on its face with musical comprehension and understanding from a synchronic, atemporal, third person perspective, it is clear that proponents of a Schenkerian approach believe that study of organic coherence can indeed change diachronic, in-the-moment musical experience. This is seen in Salzer (1962), who refers to Schenkerian theory as *musical*



*understanding of music*, as opposed to those theories which provide purely “descriptive” explanations of music (Salzer 1962, 30). Any theoretical observations made by the analyst must therefore be informed by and have implications for listening.<sup>1</sup>

Cone (1977) proposes that the analytic-iterative process involves contemplation, repeated listening, and a conscious mode of musical engagement, albeit in a more “naturalistic” form than a Schenkerian study of organic coherence. Through a process which Cone calls *identification*, the listener gains understanding and appreciation of a piece of music, engaging in “active participation...by following its progress” (Cone 1977, 563). Through multiple hearings or interactions with a piece of music, a listener progresses through three stages of interpretation. The first hearing is purely experiential, while the second is synoptic as the listener actively comprehends structure, which usually involves some form of atemporal study (*ibid.*, 564).<sup>2</sup> The third hearing, called an “ideal” first hearing by Cone (*ibid.*, 559) is one in which the experiential and synoptic readings blend together to provide informed or enhanced experience; one in which the listener’s thought moves between active, conscious comprehension and a sort of enlightened visceral experience of the music informed by this contemplation of structure.

Lewin also discusses the relationship between the analytic study of music and its effects on listening. According to him, transformational theory, or rather the attainment of a *transformational attitude*, provides a means for experiencing structures uncovered in analysis. In this attitude, the analyst takes a position *inside* the music, and can learn to experience the movement of pitch structures as movement through musical space (Lewin 2007a, 159). Creating

---

<sup>1</sup> Salzer (1962) discusses the need for theory and aural skills rudiments before undertaking his Schenkerian approach. Rather than relying solely on theoretical abstractions, the student must have extensive listening experience (familiarity) with the music, and subsequently have the ability to identify by eye and by ear important musical structures (35-6).

<sup>2</sup> While not explicitly stated, it is strongly implied that ‘atemporal study’ refers to score analysis. The reader works “...at his own speed—even in his own order, for he may have to separate adjacent ideas and juxtapose distant ones in order to uncover all of the relationships governing the musical structure.” (Cone 1977, 564).

spatialized networks of pitch-class sets is a method often employed by Lewin but is not solely created through “pure logic” or universal rules. Instead, the spatializations are reflective of consistency in his “...hearing and thinking about that particular piece, or rather [his] *having heard* and thought about it.” (Lewin 2007b, 22). Lewin also discusses the importance of a conscious mode of musical attending which actively engages meaning. In an earlier publication, he refers to this as a specific “mode of response” (Lewin 1986, 381) in which perception is an inherently creative act; the musician perceives *by* creating. This positions “perception” as a form of action, an iterative process cultivated through musical experience. Acts of musical expression (performing, composition, listening, etc.) are both the objects of perception and the means of perceiving these objects.

Taking a more cognitivist approach, albeit in a naturalistic manner, Meyer (1956) discusses the relationship between thinking and listening, including the importance of familiarity with a musical style, musical expectations, and listening habits. He posits that “habit responses,” tendencies to respond in specific manners to music, are formed in response to “sound terms,” a sort of proto-schema or commonly occurring musical phrase which prompt a habituated response from a learned listener (Meyer 1956, 5). The learning of sound terms and their related habit responses occurs in a form like that of language acquisition (*ibid.*, 32). Meyer also proposes that habit responses can either be conscious or unconscious. When a habit response is conscious, this typically indicates that a learned response has been inhibited (i.e., expectation violation), resulting in a tendency to respond in an explicit or conscious manner. Unconscious habit responses are those which take their normal course stylistically, and therefore do not constitute or arise from an expectation violation. Meyer also references that conscious intellectual responses to habit responses may occur certain individuals, at certain times, whether due to

training or disposition, may be more predisposed to conscious or intellectual experience (Meyer 1956, 39). This suggests that an affective response may be prior to an intellectual one, where a conscious affective response can initiate a process of conscious rationalization for the effect (ibid., 32).

Taken together, the accounts put forth by these diverse scholars serve to support the notion that not only can repeated interaction with music change our experience of it, but that slower and more deliberate forms of contemplation, such as the study of organic coherence in the case of Schenker, or synoptic “atemporal” readings in the case of Cone, can inform both our experience and conscious comprehension of music. There are several common themes in the various descriptions provided by theorists discussed above: a long acquisitional trajectory, iterative interactions, conscious contemplation of meaning, etc. However, there are two which are explicitly referenced consistently and explicitly attached to the processes of thinking and listening respectively—verbalization and imagery. I will now examine the positioning of verbalization and imagery as ‘thinking’ and ‘listening,’ and as distinct yet ultimately cooperative processes within music theoretic expertise.

### **‘Thinking’ and ‘Listening’ as Verbal and Imagery Processes**

In this section, I demonstrate that the separation of thinking and listening prevalent theoretical discourse is demonstrative of two distinct types of processes and/or interactions that occur in music theory expertise—verbalization and imagery. For example, DeBellis highlights the prevalence of ‘conceptual’ (‘cognitivist’) and ‘nonconceptual’ (‘perceptual’) viewpoints in theoretical discourse (DeBellis 2005, 46). In Hanninen’s work on a theory of music analysis, she discusses a separation between two types of concepts, those that are sonic in nature, and those that are structural (or categorical) in nature (Hanninen 2001, 423). Similarly, Marion Guck

separates ‘experiences with music’ from ‘conceptual and verbal invention’ as two distinct types of music analytical interactions (Guck 2006, 201-203). Theorists who discuss these activities in detail often refer to verbalization and categorization processes as taking place more out-of-time, while imagery is a term referring to a wide range of phenomena related to in-time musical experiences and perceptions.

Rogers (2004) describes the iterative feedback loop as consisting of two complimentary domains and activities that involve verbalization and imagery: first, ‘thinking’ which is developed through what he calls mind training or the learning of music theory fundamentals, and second, ‘listening’ which is developed through aural skills training. These two activities are combined or linked together through the process of analysis. For Rogers, mind training is used to develop musical thinking through the acquisition of theoretical concepts, ideas, and terminology (Rogers 2004, 33), whereas traditional ear training, defined as dictation and sight singing (Rogers 2004, 100), develops listening skills by developing the ability to hear musical relationships accurately and with understanding (*ibid.*). The goal of dictation is not to be good at transcription, nor is the goal of sight singing to produce competent vocalists. These activities are designed to produce listeners who can hear sound (either heard externally or imaged internally from reading a score) as meaningful patterns. An important facet of this is the ability to recognize and identify sound fragments by symbol or label by ear without the use of written notation which helps to break down the listening process into clear parts (Rogers 2004, 104). In this way, while not stating it in these terms, Rogers emphasizes the acquisition of relevant musical chunks which become accessible in hearing (through dictation) and in music reading or in aural imagery (through sight singing). At a higher level, ear training is designed to teach students the how the listening process itself operates, in terms of how the brain groups, stores,

simplifies, manipulates, and constructs incoming sensory input. Ear training produces a kind of analytical listening that engages learned concepts and provides a form of conscious understanding by binding “...individual surface traits together through short- and long-term memory, and aural imagery” (ibid., 102).

Similarly, Gordon’s (2012) approach to pre-college music training as a learning sequence emphasizes the importance of both imagery and verbalization for developing what he calls “audiation”—the process of assigning meaning to sound, ostensibly transforming sound into music (Gordon 2012, 3), much in the same way one thinks, predicts, and understands in language comprehension (ibid., 4). Gordon distinguishes audiation from passive listening, which he calls aural perception (i.e., sound being heard), and notes that while audiation involves imagery, it is not equal to it because it is a “more profound process” (ibid.). As will be outlined shortly, within his learning sequence for developing audiation, verbalization in the form of solmization plays a vital role (ibid., 98). Audiation then can be understood as a foundational skill essential to many musical activities, particularly theorizing, which is given as the final stage of audiation learning.

Cone’s (1977) *identification* approach is one in which the listener progresses through different stages of hearings or interpretations to come to a greater understanding of a piece of music. This process, analogous to the one experienced in multiple readings of literature, involves a shift of conscious understanding informed by both imagery and verbalization processes. While it is explicitly presented as a threefold process, one can conceptualize that it continues to be active through subsequent interactions with a piece of music or literature. The First Hearing of a piece in Cone’s stages emphasizes what could be understood as ‘listening.’<sup>3</sup> It is purely experiential, one where “The trajectory of the [listener]’s thought is one-dimensional, moving

---

<sup>3</sup> Cone often refers to the development of sensory memory, which strongly suggests an imagery component here.

along the path laid out by the author” (Cone 1977, 558). Active participation, or *identification*, is not possible in this purely experiential first hearing: “...except in the case of a simple and unproblematic composition in a well-known style...” (Cone 1977, 564). Thus, the goal of this First Hearing is to find out what happens in the music, which may take more than a single listening of a piece. Once the listener is “reasonably confident” of what is coming (so that they can follow the course of the composition), they are ready to move to a Second Hearing (Cone 1977, 564).

In the Second Hearing, synoptic in nature, the listener actively works to comprehend structure, emphasizing ‘thinking’ through verbalization and categorization. The goal of this hearing “...is to arrive at a spatially oriented view of the compositional as a whole, and its method is atemporal study...” (Cone 1977, 564). Here, the reader examines the composition at their own speed, and in their own order (presumably through score study), and “...separate(s) adjacent ideas and juxtapose(s) distant ones in order to uncover all the relationships governing the musical structure” (ibid., 564). Cone also offers a description of the listener’s/reader’s mental activity as a kind of zig-zag that shifts between planes of memory and experience (ibid., 558), or is “...more accurately described as thinking about the story [composition in our instance] while using the text [score] as a means of ensuring accuracy. In a word, the Second Reading aims at an analysis...” (ibid., 557). This Second Hearing therefore is not experiential in that it follows the temporal flow of the music, but that it becomes “...an object abstracted or inferred from the work of art, a static art-object that can be contemplated timelessly” (ibid., 558). The goal of this second reading or hearing is complete according to Cone when “... it ceases to be a reading at all—when it becomes pure contemplation of structure” (ibid.). A few pages later, Cone explicitly references memory formation, and in particular, the difference in the structure or content of

memory from the Second Hearing compared to the temporal memory of the First Hearing. He states that “...our memories of the Second Hearing are more likely to be verbal, diagrammatic, or even mathematical...” and that “[t]hese concepts can direct, or modify, or occasionally correct what we actually hear...” (ibid., 565). This synoptic reading can be explicit—as in the case of an analysis completed by a musician—or implicit as completed by an encultured listener (i.e., without explicit diagramming, use of theoretical categories, etc.). While Cone does not describe this implicit process, presumably it is one that involves contemplation of structure without the aid of a score. The Third Hearing, in which *identification* becomes possible, will be discussed in the following section.

What Cone’s proposed system suggests is that the relation between “thinking” and “listening” occurs through memory and attention: repeated exposure to the same stimuli using different modes of attending (i.e., temporal and static) provides the necessary means to achieve fusion of the two in a Third Hearing. His general emphasis on memory and repetition, as well as the importance of objectification and abstraction in out of time study in the second hearing suggests an important role for categorization, conceptualization, and verbalization (“thinking” activities).

### **On Cooperative Independence: Synthesis or Dissociation?**

Scholars note how ‘thinking’ and ‘listening’ function as separate processes that work in an independent but cooperative manner, suggesting that together they provide additive benefits compared to those obtained from either alone. Such benefits are sometimes described as a new type of perception or hearing where listening is imbued with conscious recognition or understanding. This is reflected by the terms such as ‘hearing-as’ (Guck 2017; Dubiel 2017;

Larson 1993; 2012),<sup>4</sup> and ‘audiation’ (Gordon 2012). Researchers in music cognition propose that such effects arise from processes of embodied categorization which combine aspects of verbalization with embodied responses, such as image schemata (Zbikowski 2002), analogy (Bourne 2015), metaphor (Larson 1993; 2012) and embodied imitative processes (Cox 2011; 2016). However, other scholars, including pedagogues, caution that such synthesis between thinking and listening is not a certainty, but comes from specific conditions being met. Some writers suggest that this synthesis arises from a particular analytical disposition or sequencing of activities (Meyer 1973; Cone 1977; Rogers 2004; Gordon 2012), while others believe that the synthesis only occurs when particular music-theoretic concepts are used (Hansberry 2017). In order to begin untangling these matters, I will first discuss music-theoretic scholarship relating to achieving a synthesis between thinking and listening. This will be followed by a brief discussion of music cognition literature which attempts to explain some of the cognitive mechanisms behind this phenomenon.

Rogers (2004) discusses several necessary conditions for the synthesis of thinking and listening to occur. He notes that automation of ‘thinking’ activities is required to connect them to listening. He also emphasizes the necessity for music theory fundamentals (e.g., keys, scales, meters, roman numerals, etc.) to be learned to the point of automation, not as a goal in of itself, but in order to ensure continuity in reading and hearing music, resulting in the coalescing of “...separate bits of conceptual and perceptual information into groups...” (Rogers 2004, 39). Without such fluency in fundamentals, Rogers states that pattern recognition is likely not to occur, and individual musical events or elements (presumably conceptual and perceptual

---

<sup>4</sup> Guck (2017) suggests that ‘hearing-as’ melds perception with thought: “Hearing music involves interpretation of the musical sounds as instances of familiar musical entities whose meanings have been learned through extensive experience. It seems like direct perception but is more specialized, interpretation-infused “music perception.” Hearing-as melds perception with thought” (Guck 2017, 242).



information) will never be held together long enough to form “something bigger” (ibid.). This implies that once a fundamental concept is deeply learned, the automation afforded by that learning is what helps to build more knowledge and connect knowledge to the listening experience within the iterative feedback loop. In discussing the acquisition of roman numeral labels, Rogers notes that attaching function (PD, D, T) to those labels is a necessary link for connecting “inert symbols” to the dynamics of musical tension important in the listening experience (ibid., 46). Similarly, within the mind training chapter, he discusses the importance of keyboard training for the reinforcement of conceptual skills (ibid., 70). “The tangible act of pushing down keys, seeing distances and hearing resultant sounds can cement a concept to the mind and ears in a way that no amount of paperwork or talking can ever accomplish.” (ibid.). These activities, namely the connection of roman numerals to function and the reinforcement of theoretical concepts using the keyboard, are ways of opening this knowledge to connections with perceptual information. In a sense, these are “proto-analytic” activities, building on which analysis proper acts as the link between the mind and ear. The attachment of learned concepts via verbalization is also required to both grasp and enhance perceptual phenomena. Rogers states:

“But unless we are consciously aware of [perceptual states], they are only vaguely felt, and their nuances and structural importance escape us. It is essential that these sensations be heard through the reinforcement of visual and intellectual experience—through mind training” (Rogers 2004, 46).

Rogers also discusses *music analysis*, which he posits to be the activity that links mind training to ear training, effectively connecting thinking to listening. He states that analysis is not simply a description of the structure of a piece (e.g., ABA), but is a position or mental posture taken to demonstrate a specific point of view (Rogers 2004, 75). Through analysis, the theorist

attempts to answer several multifaceted questions, none of which alone can capture the totality of the piece, but rather require multiple angles to continue to deepen understanding. The analyst must engage with “how” and “why” questions, seek to understand causal logic and relationships between sections or events, and understand patterns and their place in the hierarchy of the piece, among many other analytic activities (ibid., 75-76). Rogers also notes that analysis can take many forms, such as working from the “inside “out from a more static overview or working from the beginning to the end of a composition to explain the dynamic experience of the listener (ibid., 78). Rogers also lists several dangers of this process, notably the attempt to connect musical events that are incompatible or on disparate levels (ibid., 76), or fragmenting the work too deeply, leading to the separation of details, and ultimately the separation of “...listening from thinking, and finally feeling from thinking...” (ibid., 79). He notes that some dissection and fragmentation is necessary; however, this process cannot be completed without recourse to “listening” as the “re-creative act” that can synthesize what talking and thinking have separated (ibid.).

Gordon’s (2012) learning sequence approach emphasizes the importance of the ordering of activities to ensure what he sees as the proper connection of thinking and listening. In particular, he emphasizes that any intellectual understanding of music must be preceded by deeply learned musical knowledge.<sup>5</sup> As previously discussed, Gordon’s model in general is centered around the acquisition of audiation, a multimodal skill used during many different musical activities, including performance, composition, and theorizing. Audiation is the process of assigning meaning to sound, thereby transforming sound into music (Gordon 2012, 3) much in

---

<sup>5</sup> Musical knowledge for Gordon specifically refers to music as sounded and experienced. His sequencing of activities involves the solidifying of sounded representations before others, such as notation and solfège, are introduced. In this way, other forms of knowledge, including theoretical knowledge, are layered on top of firmly established musical experience. See the description of his Skill Learning Sequence below.

the same way one thinks, predicts, and understands in language comprehension (ibid., 4). Gordon distinguishes audiation from passive listening, which he calls aural perception (i.e., sound being heard), and notes that while audiation involves imagery, it is not equal to it because it is a “more profound process” (ibid.). Audiation then can be understood as a foundational skill essential to many musical activities, especially theorizing.

Gordon’s model does not provide an *explanation* of how or why audiation works. Instead, he proposes a *description* of its development, acquisition, and relationship to various musical activities (listening, reading, and writing in various musical contexts; see Figure 1.2). Gordon proposes that the ability to audiate is grounded in the acquisition of familiar tonal and rhythm patterns. Through this proposed learning sequence, students move from what Gordon calls *discrimination learning*—the ability to recognize, produce and audiate familiar patterns—to *inference learning*—the ability to anticipate familiar patterns, understand unfamiliar patterns, and make predictions about what might occur, even in unfamiliar contexts. Students move through various modalities and types of musical activities at each level of learning (discrimination then inference), and from activities that are more production based (listening and singing) to those that are more static (reading/writing, and ultimately theoretical understanding, which is the highest level of inference learning, Gordon 2012, 118). Similarly, when moving through discrimination learning to inference learning, students progress through activities that are primarily based in imitation to those which are generative (e.g., improvisation/composition). When paired with Gordon’s “Stages and types of audiation,” the learning sequences model provides an in-depth description of musical skill acquisition which highlights how audiation

develops and adapts in different musical contexts, and for different musical purposes.<sup>6</sup> I will now go through his sequence (summarized in Fig. 1.2) in some modest detail.

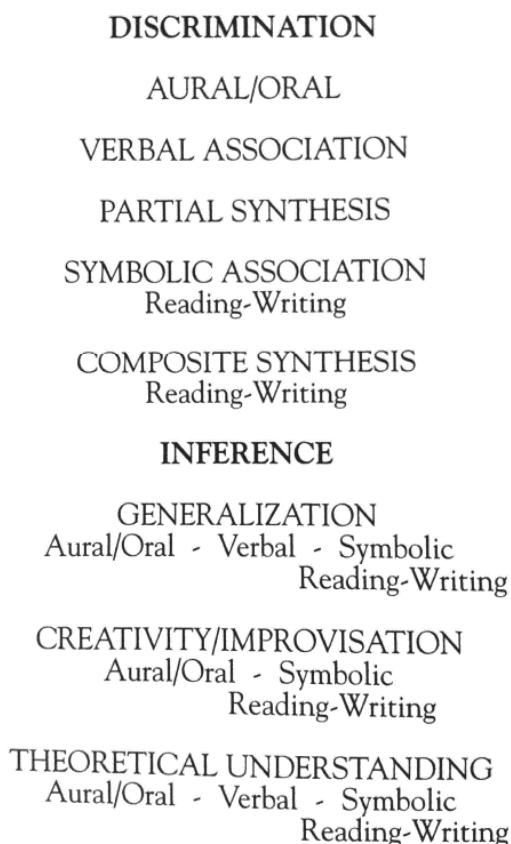


Figure 1.2: Gordon's Levels and Sublevels of Skill Learning Sequence (Gordon 2012, 98)

Gordon's first stage of skill sequence learning is discrimination learning, in which students progress from primarily aural to more notation-based skill sets. The aim of discrimination learning in general is to build memory for the foundational building blocks of music (musical patterns), so that they may be recognized (distinguished from one another),

---

<sup>6</sup> See Gordon (2012) 13-23. Here Gordon describes how the audiation ability changes through the stages of discrimination and inference learning in the various musical contexts and modalities (aural/oral, reading/writing, performance/composition/improvisation). In general, audiation progresses through basic processes of imitation (e.g., momentary retention and production of familiar patterns), to the ability to recall, interrogate and "maintain" patterns in memory (e.g. ability to explicitly identify patterns, verbalize them, etc.), and finally audiation progresses through to the final stage of being able to anticipate in familiar music and predict in unfamiliar music (p. 23).

identified (solmized), and used (produced) in various musical contexts. All stages of discrimination learning must be mastered before moving onto the next, and before moving onto inference learning. Students begin discrimination training by learning tonal and rhythm patterns in listening (aural, identification) and production (oral, typically singing on neutral syllables), after which they learn to associate these tonal and rhythm patterns with tonal and rhythmic syllables (solmization systems) at the verbal association level.<sup>7</sup> Gordon is quick to emphasize that these syllable systems are merely a means to an end, and not an end in themselves.

Therefore, solfège verbalization is only a technique for comprehension. Eventually syllable use, according to Gordon, becomes unconscious; the listener is unaware of syllables during audiation until the syllables are needed for “momentary clarification” at which point the student can bring them into consciousness to aid audiation/compression (Gordon 2012, 113).<sup>8</sup> At the next stage of learning, the partial synthesis level, the mappings between familiar patterns and syllables become stable, at which point the student is ready to progress to learning notation at the symbolic association level (both the reading and writing of familiar patterns).<sup>9</sup> At the final level of discrimination learning, students reach the composite synthesis stage where all prior stages (aural/oral, verbal association and symbolic association) are stable so that students can successfully audiate at the same time as they are reading and writing notation. Parallel to the dissociation that can occur between sound and syllable, Gordon is quick to caution here that

---

<sup>7</sup> Gordon is quite strict about which solfège systems are appropriate, specifically that the solfège system used must reflect the internal logic of tonal and rhythm patterns (Gordon 2012, 60). To this end, he proposes that moveable do solfège with la based minor should be used (see page 72).

<sup>8</sup> Similarly, Gordon states that syllables (both tonal and rhythmic) do not always need to be taught, as students will eventually be able to use them in new and different ways in both familiar and unfamiliar contexts. Essentially, solmization also becomes iterative—a thinking tool that builds on itself (Gordon 2012, 113).

<sup>9</sup> He notes that in dictation, the teacher should use neutral syllable because students can figure out pattern context theoretically. It is important for students to generate their own context at this level: the ability to generate context *is* audiation. Essentially, the fusion of sound context inflecting the heard pattern. Solmization therefore gets placed on “on top” of the heard pattern, after it has been audiated. I translate this to mean that the sound, once imagined in context, has a specific quale onto which students map the appropriate syllable. See page 113 on partial synthesis.

notation is simply a cue for what has already been audiated; sound to symbol mappings occur in that order with sound taking precedence. Notation is therefore a cue for what has already been learned; it is a form of imitation.

Inference learning is the fulfillment of audiation (Gordon 2012, 96) and involves the ability to predict and anticipate in both familiar and unfamiliar contexts. In the first stage of inference learning, called “generalization,” students begin to be able to identify and sing (aural/oral), solmize (verbal association) and read and write (symbolic association) familiar patterns in unfamiliar orders, and unfamiliar patterns among familiar patterns in various orders. It is at this generalization stage that students begin using audiation abilities gained from learned patterns to make inferences in unfamiliar situations, be that figuring out the syllables for an unfamiliar pattern or converting written notation of an unfamiliar pattern into sound (real or imagined, called notational audiation). Once this stage is mastered, students are able to progress to the second stage of imitation learning which involves generative rather than imitative activities, such as improvisation and composition.

It is only after this second stage that students should, according to Gordon, be introduced to the highest level of inference learning, theoretical understanding (Gordon 2012, 118), which “...explains to students through intellectual understanding how and why they audiate as they do when they engage in musical thought and performance” (ibid., 140). While Gordon’s perspective on what music theory may be quite limited and somewhat problematic,<sup>10</sup> it is clear from the ordering of the stages in inference learning that any conceptual understanding in Gordon’s

---

<sup>10</sup> What Gordon includes in the “theoretical understanding” stage of inference learning is primarily what many pedagogues and theorists would include in music theory fundamentals rather than more advanced theorizing, including individual analyses of pieces. The bulk of what is included are definitions of and differences between fundamental musical features such as a “definition of tonality” and “differences between modal and mode” (p. 140). Gordon includes aesthetics and some basic stylistic knowledge in this category, as well as some Schenkerian notions of deep structure, background, middle ground, and foreground. See pages 140-141 for a complete list.

system *must* be preceded by *musical* understanding. This is contrary to Rogers' (2004) position which places conceptual understanding as occurring at the same time as musical (i.e., thinking and listening training at once), although Rogers does make the claim that conceptual understanding develops faster than aural skills in general (Rogers 2004, 103). This difference may stem from the differences in the context for each method;<sup>11</sup> however, one can also understand the difference as one of approach. Gordon's is one where musical understanding (i.e., more listening/experienced based) is gained first, which aides conceptual understanding as grounded in musical experience. In Rogers' approach, conceptual understanding is thought to aid musical understanding (and vice versa), but a binding agent (musical analysis) is needed to connect the two. In Gordon's system therefore, a binding mechanism between thought and listening is never needed because musical experience is always presumed to be active in audiation, and therefore he attempts to ensure that "sound experience" is never separated from conceptual thinking to begin with.

In summary, audiation is developed primarily through the acquisition and subsequent use of tonal and rhythm patterns, for which verbal cues (syllables, labels, etc.) are simply tools to reinforce, guide and strengthen the audiation ability. Eventually, these tools "disappear" (become unconscious) and are only used when new learning takes place, for example when learning an unfamiliar pattern (Gordon 2012, 89). As each stage in discrimination and inference learning relies on the material and skills learned in the previous stage, the sequence learning process can be viewed as an iterative one where audiation becomes more profound at each stage. While more conceptual forms of thinking, such as music theoretic understanding, are found in only the final

---

<sup>11</sup> Gordon's sequence model describes hypothetical process for younger learners including children and teens, whereas Rogers' approach is specifically for post-secondary learning where the time frame is more limited to undergraduate core.

stage of the entire process, the interaction between thinking and listening is embedded throughout audiation development.

Cone (1977) also discusses the synthesis of thinking and listening as the culmination of both perceptual and conceptual interactions with a piece of music. Recall that the first two Hearings developed perceptual (experiential) and conceptual (synoptic) readings respectively. The Third Hearing, called an “ideal” First Hearing by Cone (1977, 559) is one in which the experiential and synoptic readings culminate to provide informed or enhanced experience; one in which the listener’s thought moves between active, conscious comprehension and a sort of enlightened visceral experience of the music informed by this contemplation of structure. Therefore, this Third Hearing is similar to the first in that it is temporally oriented. Now, however, it is informed through synoptic comprehension which “...replace(s) the naïve pleasure [of the first hearing] with intelligent and informed appreciation” (ibid., 558). How the listener achieves this state is not spelled out; but, according to Cone, it requires a few key elements. Firstly, the listener must bring forth (through memory) the synoptic understanding gained from the Second Hearing, and they must also be fully engaged in the music’s temporal flow as gained from the First Hearing. Secondly, the listener needs to engage in some intentional forgetting of the synoptic analysis in order to actively and fully engage with the temporal flow. Cone describes this process as a double trajectory: “...thought simultaneously moves on two levels: one fully conscious, the other at least partially suppressed” (ibid., 558). He notes that “...it is difficult to concentrate on the perceptual present when memories keep interrupting consciousness,” (ibid., 564), but also “...the immediacy of what we are hearing makes it difficult or impossible to entertain other *musical ideas* at the same time” (ibid., 565). What Cone is saying here is that while we may attempt to remember musical ideas (likely through imagery) gathered



from the previous Hearings, that they cannot be fully realized because the incoming input overrides the internally generated content. Therefore, while synoptic analysis provides concepts that "...can direct, or modify, or occasionally correct what we actually hear...they can never replace or contradict it—not, that is, so long as we are fully participating in the actual music itself" (ibid., 565). Therefore, as long as one is attending to the temporal present—what is being heard—concepts gained from the synoptic analysis can affect or inflect, but not replace, what is being perceived. To make this more concrete, Cone provides as an example the opening harmony of Beethoven's Ninth symphony, which is a dominant chord in D minor. Cone notes that it only *becomes* a dominant chord in a Third Hearing after one has developed both temporal (First) and synoptic (spatial/bird's eye view, Second) Hearings (ibid., 565).

What remains unclear is how one achieves an ideal Third Hearing: is a specific binding mechanism required to relate the temporal and the static, or does this simply occur "naturally" over time? Cone's suggestion of intentional forgetting seems to imply, but not require, the latter. The memories constructed from prior hearings will affect future experience "on their own" after study and contemplation, but they will not necessarily exert an influence without some effort on the part of the listener. One appears to need to focus on the perceptual present, while engaging with prior learning only to the extent that it does not fully overwhelm one's attention. This way, memories from the synoptic analysis inform or inflect incoming temporal experience, but do not direct consciousness away from it.

To engage one last scholar on these matters, we turn to Hansberry (2017), who posits that different types of music theoretic concepts have different affordances for cognition during analysis, such that some concepts are more purely conceptual, some are more experiential, and some are a blend of the two. For Hansberry, concepts fall into two broad types. The first type is a

*phenomenal concept* which is formed from the conceptualization of what it is like to experience the object being considered. Such concepts have specific *qualia* and refer to elements of felt musical experience. The second type is the *theoretical concept* which is an abstract object that does not include reference to experiential information. Theoretical concepts "...refer to abstract objects [which] broadens the scope of music theory beyond only perceptual phenomena, allowing us to possess theoretical concepts for inaudible aspects of a composition that we may still want to invoke in analysis." (Hansberry 2017, 20). Using this distinction between phenomenal and theoretical concepts, Hansberry is able to explore how these different conceptual types are used in analytical frameworks. He subsequently outlines two types of theories: *simple theories* and *complex theories*. *Simple theories* rely either on theoretical or phenomenal concepts, but typically not both. An example is corpus study research (using computers), an approach that abides primarily by theoretical concepts (ibid., 68), while Zbikowski's theory based in conceptual metaphor uses primarily phenomenal concepts (ibid., 124). *Complex theories* are those that use both theoretical and phenomenal concepts and are separated into two types: segregated approaches (which use theoretical and phenomenal concepts separately), and mixed approaches (which use mixed concepts). Segregated complex theories are those like transformational and neo-Riemannian theories where analytical motivations are quite experiential in nature, but also use theoretical concepts as a means to provide analytical consistency (ibid., 148). Complex theories which use mixed types of concepts, such as Schenkerian approaches, which afford a synthesis between the conceptual and experiential that other theoretical frameworks may not afford.

## Explanatory Approaches in Music Cognition

Several different approaches within the music cognition literature attempt to explain or at least situate some of these aspects of music-theoretic expertise within a cognitive framework. For the first of these I return to Leonard Meyer's *Emotion and Meaning in Music* (1956) and *Explaining Music* (1973). His work encapsulates several features of the loop discussed by his music theoretic and pedagogical peers outlined previously, with a particular emphasis on the need for categorization and conceptualization, both naturalistically through repetition (acquisition of "habit responses" to "sound terms" in familiar music, see Meyer 1956, 5) or more manually through analysis (discovery of statistically regular schemata in style analysis, see Meyer 1973, 7). Meyer views both forms of categorization, naturalistic and cultivated, as extensions of the same cognitive phenomenon, in that they require a specific attentional set or mode of contemplation. In this way, an analytical disposition is a natural one, and one where conscious awareness itself is what compels conceptualization which "...takes place whenever anyone attends intelligently to the world" (Meyer 1973, 5). In the naturalistic setting of listening and attending one's habit responses give rise to an affective response which may in fact be prior to an intellectual response as used in an analytical setting, insofar that a conscious affective response can initiate a process of conscious rationalization for the effect (Meyer 1956, 32).

Despite Meyer's view that an analytical disposition is merely an extension of a naturalistic one, he still highlights the potential problem of apparent separation between "thinking" and "listening," particularly in critical analysis where the goal is causal explanation of idiosyncratic features of a piece of music (Meyer 1973, 6). He refers to this as the "disparity effect" whereby the speed and ease with which music is experienced is at odds with the complexity and time involved in the completion of an analysis, and its explanation (Meyer 1973,

14-15). In this way, critical analysis can appear to be purely intellectual and detached from experience. Poorly executed critical analysis can fragment and conceptualize what should be felt, whereas good criticism "...separates where separation is warranted by the musical structure and unites where the musical organization permits" (Meyer 1973, 6). For Meyer then, the process of binding "thinking" and "listening" together is one which occurs naturally, as long as the critic remains true to their musical experience. The following passage emphasizes this process perfectly:

Repeated playing and listening may be required. Because the several parameters do not necessarily move in congruent fashion (with the result that harmony, melody, rhythm and so on may each yield a different pattern of organization), it will at time be helpful to analyze the parameters separately in order to study their interrelationships. Often it is illuminating to "normalize" a passage—rewrite it in a simpler, archetypal form—in order to understand how the composer modified a traditional schema. Always it is important to discover which tones or harmonies are structurally essential and which are ornamental. When employing such techniques—which are not modes of explanation, but methods for disclosing how a musical event functions—the critics "ear," his musicality, must guide analysis. It must accept or reject a linear abstraction, an harmonic reduction, or a rhythmic analysis. **His ear keeps the critic honest** [emphasis added]. Without its control, theory or style analysis tends to become a Procrustean bed to which the practice of composers is made to conform (Meyer 1973, 17-18).

Meyer cites that the reason this binding occurs naturally, as long as one engages with musical experience, is that emotional responses to the world are undoubtedly linked to cognitive patternings. Therefore, "...the musical processes and structures explicitly conceptualized in criticism are those which evoke affective responses in sensitive and experienced listeners" (ibid., 6).

Meyer also raises an important point that is not explicitly mentioned in the theoretical and pedagogical literature outlined earlier: that analysis, particularly analysis as completed and taught in an academic context, is subject to certain restrictions. Firstly, there are aspects of musical experience that more easily lend themselves to abstraction, and are therefore easier to

analyze, discuss and teach (ibid., 5). The more elusive forces that shape musical experience are less frequently studied because they are less easily transformed into abstract concepts. Secondly, and related to this, only certain types of music lend themselves to reflective study and analysis. To this issue, Meyer cites the need for hierarchy in music, particularly for the integration of thinking and listening practices: “Because music is hierarchic—tones form motives, motive phrases, and so on—what is separated at one level becomes unified on the next” (ibid., 6). He notes that the music of certain composers, such as in the transcendental music of Cage, cannot be analyzed, but only described, because there is no hierarchy to its structure (ibid.). Meyer therefore suggests, prefiguring Hansberry (2017), that the relationship between reflective analysis (thinking) and listening may only be possible with certain types of musical concepts (i.e., those that can be easily abstracted), and with certain types of music in which these musical parameters are organized hierarchically.

Zbikowski (2002) places categorization, concepts, and metaphoric mappings at the center of human abilities to conceptualize highly complex music upon a first encounter. While music theory may be a specific disciplinary manifestation of conceptualizations of music, Zbikowski positions any musical sense making, music theoretic claims included, within the realm of ecological cognition. Any conceptualization of music formed by a listener or analyst therefore relies on the same cognitive mechanisms used to make sense of the world on a daily basis (Zbikowski 2002, 5). On the whole, any conceptualization of music relies on the formation of categories, which then facilitate higher level cognition. Musical concepts, according to Zbikowski, are those which are a product of a process of categorization, are essential for guiding future action, and can be related to other concepts, including those related to perceptual categories, bodily states, and linguistic constructs (ibid., 61). These categories can be formed

without recourse to language, but in the case of music theoretic observations are typically provided with some form of linguistic label. These concepts, once formed, facilitate higher cognition, including cross-domain metaphoric mappings and the development of conceptual models and theories. This is what is claimed to underlie the ability to conceptualize spatial aspects of pitch; one learns to associate change of pitch with the concept of up and down, aided in the embodied experience of bodily movement upwards and downwards. These cross-domain mappings are what facilitate the formation of larger theories and conceptual models, which Zbikowski demonstrates by examining the work of several theorists, including Lewin, Schenker, Rameau and others (ibid., 96-97). Importantly, Zbikowski positions musical categories as flexible and open to update as thought unfolds; these categories are not *a priori* or fixed.

In a similar fashion to Zbikowski, Cox (2016) places categorization and metaphor at the center of musical comprehension, but with an even greater emphasis on the role of the body in musical comprehension which he terms the *mimetic hypothesis* (see also Cox 2011). This hypothesis states that much of how music is comprehended by a listener is through bodily imitation, which can be either covert or overt. Imitation on the part of the listener is facilitated partly by the sound-producing actions of a performer, or, in the instance where the performer is not directly visible, on sound-source identification (Cox 2016, 12-13). Cox supports this hypothesis by citing cognitive research on various forms of imitation within and outside of the discipline of music, such as developmental infant studies, subvocalization in music imagery, and mirror neuron studies. Contrary to Zbikowski, Cox more narrowly focuses on comprehension of a listener, rather than extending the work to the theorist or analyst. He does, however, consider musical expertise as a form of specialization in which a listener gains more embodied knowledge which can then be used to make conceptualizations about music.

Cox also discusses the implications of mimesis for musical conceptualization, namely how different aspects of musical motion can be conceived of and experienced by an active listener. Cox posits that several types of listening are important for affording different kinds of musical experience. He proposes a tripartite perspective in musical listening, in which a first-, second-, or third-person perspective can be taken to conceptualize musical space and time. Along with this, musical motion can be conceptualized from either an interior perspective (*within* the music), or an exterior perspective (from *outside* the music) (Cox 2016, 139). These different positions afford different musical experiences and different types of conceptualizations. For example, the exterior perspective on a moving observer is the most common position taken by music analysts according to Cox, which allows for a single view of the musical landscape (from a birds-eye view), but because of this removed perspective, the listener (or analyst in this case) is more attentive than actively listening, resulting in a reduction in the immediacy of present musical experience (ibid., 139).

Finally, other scholars have also proposed that metaphor and analogy offer a means to combine embodied experiences with verbalization and categorization processes. Larson (2012) argues that the experience of physical motion shapes experience of musical motion such that conceptualization of gravity, magnetism and inertia in music are based on memories of the physical experience of these sensations (Larson 2012, 1-2). These metaphors shape both thinking *about* music and thinking *in* music, which combine together to create musical meaning (ibid.). Similarly, Bourne (2015, 2016) applies Structure-Mapping Theory as developed by Dedre Gentner (2016a) to explain how theorists are able to perceive musical irony. This framework relies on the notion of categorization (for schema, themes, and forms) as the foundation for analogical thinking which facilitates the perception of irony. Here a theme is a category (or

musical schema) or an exemplar of such a category, which is formed through experience. Properties or elements of this theme (e.g., scale degree contour, rhythmic contour) can be mapped onto another element (e.g., up-down, short-long), which affords analogical thinking through the use of metaphor. For listeners experienced in a particular style, the perception of musical irony is afforded by violations within a common set of musical vocabulary (i.e., schema). These violations include generalized expectation violations and a flouting of Gricean Maxims, or the commonly understood parts of conversation (in a musical context, formal or schematic violations, see Bourne 2016, 3.2 and 3.3). While Bourne's project deals with a specific kind of theoretical activity applied to a specific body of repertoire, it is clear that here, as in the work of Zbikowski and Cox, categorization and metaphor play an integral role in her theoretical apparatus to conceptualize music and make claims about its structure and content.

### Effects of Music Theory Training

The manner in which theorists and pedagogues discuss the interaction of thinking and listening suggests that the effects are indeed real and tangible. While the majority of these effects are positioned within a positive light, some are indicative of negatives or trade-offs.

*Enhanced or modified perception.* The most commonly cited claim regarding the effects of the thinking-listening dichotomy deals with enhanced or modified perception. This was referenced above with regard to the acquisition of 'hearing-as' and 'audiation.' While such terms reflect a change to conscious perception (i.e., that what may change is comprehension, not necessarily 'hearing'), some do refer to expertise as affecting hearing more directly. This seems most typically claimed in discussions of increased sensitivity to particular features or patterns, as a kind of enriched perception. Cone (1977) notes that the ideal Third Hearing is a true



combination of experiential and synoptic hearings. In Cone's view, "(l)ike the First Hearing, the Third is temporal and experiential; but it is characterized by an *enriched perception* [emphasis added] of the temporal flow, and it realizes a controlled and appreciated experience" (Cone 1977, 364). Meyer (1973) explicitly discusses the effect of musical education on the enhancement of listening, specifically, that music theoretic training can affect listening "(b)y calling attention to patterns and relationships which might otherwise have been missed..." which in turn "...refines the aural imagination and increases the sensitivity of the cognitive ear" (Meyer 1973, 17).

*Language benefits the precision of thought.* The notion that language has a profound effect on the way someone is able to think is a pervasive one indicative of the Whorfian hypothesis (that language influences thought, see Hunt and Agnoli 1991). Such influences of language on thought have been observed in studies which show that differences in language for color can provide benefits to perceptual color categorization (Martinovic, Paramei, and McInnes 2020), which can lead to reaction time advantages for languages that provide more granularity in color shades (Saunders and van Brakel, 1997; Witzel and Gegenfurtner 2015; 2016). A similar argument is made in scholarship that discusses the loop between thinking and listening. Gordon (2012) discusses the important role that verbalization plays in bootstrapping musical memory. While verbal labels may eventually be discarded (or they at least fall into subconscious awareness), they and their uses are central to knowledge growth. Some quotations can highlight this point, notably one which indicates the importance for verbalization as a springboard for further development:

The aural/oral and verbal association level of learning naturally occur together in speaking because we make use of and develop our language instinct as a result of our desire to name objects we sense. *Thought engenders language and language*

*gives precision to thought.* When we hear tonal patterns and rhythm patterns performed, our immediate desire is to listen and distinguish patterns (perhaps as a way to make what is unfamiliar appear familiar), and only later apply names to what we hear. Thus, in music, an additional level of learning, verbal association, is needed to teach syllable names along with names of tonalities and meters.

(Gordon 2012, 103)

And, in a similar vein regarding verbalization providing explicit benefits to memory:

Verbal association allows students to retain and recall patterns, tonalities, and meter for use in higher levels of discrimination learning and in all levels of inference learning. That is particularly true in creativity/improvisation because without verbal association, even patterns once audiated at the aural/oral level may be lost. Because a neutral syllable is used at the aural/oral level, students rarely retain no more than ten tonal patterns in a given tonality and ten rhythm patterns in a given meter. When pitches are given syllable names with internal logic at the verbal association level, however, patterns will not be easily forgotten. Students make free use of syllables for purposes of listening and performing as well as expanding pattern vocabularies. (Gordon 2012, 106)

*Biased perception and 'inattentional deafness.'* While verbalization is understood to add precision to thinking and/or perceptual discrimination, scholars also suggest that extensive theoretical training can bias the ways in which we hear and experience music. Temperley (2001) notes that while theorists' listening may be enhanced, it may also be 'contaminated': "Surely the hearing of music theorists has been influenced (enhanced, contaminated, or just changed) by very specialized and unusual training...I do not wish to claim that music theorists hear things like

harmony, key and so on exactly the same way as untrained listeners; surely they do not” (Temperley 2001, 7). This is indicative of a sort of inattentional deafness—where music theorists may possess enhanced awareness or perception for particular features or patterns, this comes with a trade off—the inability to hear other features due to limited attentional resources. Inattentional deafness, like inattentional blindness, is found when features of an event are not observed or remembered by participants because their attention is directed elsewhere, even when that event should stand out (e.g., the random appearance of a man in an ape suit at a basketball game, see Simons and Chabris 1999). The presence of inattentional deafness has been confirmed in musician populations (Koreimann, Gula, and Vitouch 2014), and can even occur cross-modally where visual attention overload can increase the likelihood of inattentional deafness (MacDonald and Lavie 2011, Raveh and Lavie 2015). Such findings have strong implications for music theory and analysis as these activities heavily involve visual score analysis, which may cause inattentional deafness for features and patterns outside the purview of the analytical system or tools being used.

*Intuition in music theory as automation and introspective availability.* The reference to automation and increased introspective availability is much more implicitly discussed in the literature. The general notion here is that over time with expertise, the ‘distance’ or separation between thinking and listening becomes smaller such that the two activities or processes are freely available to the theorist, can occur at the same time or can be done in any order. This is again indicative of the terms ‘hearing-as’ and ‘audiation’ discussed above, in which thinking and listening are fused into a new type of consciousness-laden perception. Several pedagogues refer to increased automation either more explicitly (Rogers 2004, 39), or more implicitly through

discussion of an increased availability of skill in multitasking, such as in Gordon's (2012) learning sequence approach.

## Project Methodology and Scope

Having surveyed the music theoretic literature, I will now outline the methodology and scope of this project. This project synthesizes research in expertise, categorization, concepts, mental representation, and imagery to develop an explanatory cognitive framework for the iterative feedback loop between listening and thinking in music theoretic expertise. I employ an approach that combines theory-creation with experimental application. The next four chapters of the dissertation develop a cognitive framework through synthesis of cognitive and music theoretic scholarship, while the remaining chapters apply the framework to explain results from a quantitative survey on musical ambiguous figure perception.

### **Which "Music Theory?"**

Not all music theories share the same ontological perspectives, and therefore may not share the same goal of 'deepening understanding' through iterative thinking-and-listening activities. The goal of the current project is not to examine different analytical traditions and assess the extent to which they align with this perspective. Rather, the goal is to create a concise cognitive framework capable of explanatory power for theories in which this may be true. To this end, the project will narrowly focus on Galant schema theory in the context of sonata form analysis. To understand this, I begin with the so-called descriptive-suggestive dichotomy of music theories.

## The Descriptive-Suggestive Distinction: Introspection versus Perception

Temperley (1999) discusses two contrasting types of music theories: ‘suggestive’ and ‘descriptive’. A suggestive theory is one where the goal of analysis is to find and present a new way of hearing a piece or to enhance the experience of it (Temperley 1999, 70). A descriptive music theory is, by way of contrast, a form of psychological music theory related to the discipline of cognitive science which aims to describe listeners’ unconscious mental representations of music (Temperley 1999, 68). Temperley suggests that descriptive theories are akin to approaches in cognitive linguistics which seek to reveal, describe, and explain cognitive processes through introspection. From this perspective, descriptive theories are capable of uncovering underlying, pre-existing mental representations through introspective means, which can then be confirmed using experimental methods (or corpus studies, in the case of Temperley’s research, e.g., de Clercq and Temperley 2011).

DeBellis (2009) criticizes Temperley’s descriptive-introspectionist method on two grounds: firstly, it is unclear what introspective mechanism is responsible for accessing namely unconscious mental representations (DeBellis 2009, 122), and secondly, it remains unclear as to how music-theoretic terminology used by scholars could convey theoretical intuitions in a consistent manner from reader to reader (*ibid.*, 123). As an alternative, DeBellis proposes what he terms a perceptual approach: instead of bringing conscious awareness to existing mental representations, an analytical attitude toward musical contemplation forms the object to be contemplated, essentially forming a new mental representation (*ibid.*, 127). In this way, the primary tool used by the theorist is not introspection, but a sort of perception-laden theoretical judgement. DeBellis’s skepticism in language’s ability to communicate musical intuitions seems to be grounded in an argument against consistency and knowability in qualia, in that there is no

way to confirm that the *feeling* of a qualitative state is consistent between people. Ultimately, DeBellis disagrees with an introspectionist approach, and favors one in which potential mental representations of musical objects discussed by theorists are more likely created than discovered.

Temperley (2009) defends his introspectionist position against DeBellis, citing a wealth of research (corpus studies included) which support the notion that introspective music theory has indeed been capable of revealing existing mental representations. Temperley relies primarily on analyses of rhythmic and metric phenomena in which the introspective or subjective experience of meter strongly corresponds to existing structures and mental representations (Temperley 2009, 134). He does, however, concede that a perceptual approach, where representations are created rather than discovered, is certainly a possibility, and aligns well with a suggestive type of music theory (ibid., 136-137).

I argue that it is possible to view the suggestive/descriptive divide as two sides of the same coin: suggestive music theory can be viewed as the thinking→listening portion of the loop pictured in Figure 1.1, while descriptive music theory can be viewed as the listening→thinking portion. In this way, suggestive music theory can be thought to build and/or change representations of music used in listening through analysis (i.e., thinking affects listening, or a more ‘perceptual’ approach). However, where mental representations of music already exist, descriptive music theory serves to describe these representations; or more pointedly, to provide more direct conscious access to the ones already present (i.e., listening affects thinking). Therefore, both descriptive and suggestive theory and analysis can be seen to operate within the listening/listening loop; descriptive engagement is a means to consciously highlight or access mental representations that already exist, and suggestive engagement is a means to either create new representations or update existing ones.

## On a Galant Schema-Theory Case Study: Acquiring Eighteenth Century Hearing

The approach to Galant schema as outlined in *Music in the Galant Style* (Gjerdingen 2007) is one very much grounded in a descriptive music theory akin to cognitive linguistics. These meaningful patterns have been ‘discovered’ using corpus analysis methodologies (see Gjerdingen 1988) and verified as historically relevant through examination of pedagogical source materials (see Gjerdingen 2007; 2020; Baragwanath 2020). Scholars explicitly understand Galant schemata to be categories of the natural kind, i.e., that they can be learned through statistical learning (or encultured listening) alone. Gjerdingen notes that “...these schemata were designed to be noticed by anyone who listened to enough of this music. For modern devotees of classical music, every schema in this volume may sound quite familiar” (Gjerdingen 2007, 15). Uncovering these patterns therefore relies equally on corpus creation through score analysis and on introspection involving imagery: “Meyer’s entry into the schema concept was initially through introspection: one imagines that his meditations on style and habit responses in the 1950s were accompanied by 1–7, 4–3s sounding in his inner ear” (Byros 2012a, 282).

Despite such descriptive roots, Galant schema theory has also been understood and used as a type of suggestive music theory. Learning these patterns offers a way to hear eighteenth-century repertoire as eighteenth-century listeners and composers may have: “To hear this music more as a Mozart might have heard it, to imagine musical behaviors more consonant with the premises and goals of those who lived at Galant courts, and to seek a more realistic account of how Galant musical craftsmen fashioned raw tones into finished art has been the aim of this book” (Gjerdingen 2007, 452). Similarly, learning schemata, even for those already enculturated in the repertoire, is seen as a way to come to a greater understanding and appreciation of the art: “In learning to recognize the schemata of Galant music, one becomes better able to appreciate

the art of the Galant composer. And in learning to judge the manner in which the schemata are presented in a particular composition, one becomes better able to understand the equally important art of the Galant listener and patron” (Gjerdingen 2007, 7).

Byros (2009a,b; 2012a,b) has argued similarly that expertise with Galant schemata categories provides a means to access eighteenth century modes of hearing, particularly as it relates to tonality perception. His examination of the reception history of the first movement of Beethoven’s *Eroica* reveals three primary strains of tonal interpretation: one which perceives a modulation to G Minor, a “Cloud” strain which proposes a form of tonal ambiguity, and one which reads the whole passage within the key of E-flat major. Byros demonstrates that the tendency to perceive key change in these opening bars operates on a historical axis, indicating that the historical situation or context is a contributing factor in the variation of key perception. The G-minor hearing therefore appears to be a historically and stylistically appropriate one, one which Byros proposes arises from the perception of the *le-sol-fi-sol* schema.

Galant schema theory therefore offers an ideal domain or case study for the current project, as the categories of these schemas appear amenable to implicit learning through exposure as well as explicit learning through direct training with the categories. These can be available introspectively through imagery, i.e., ‘descriptive’), and can in addition both imbue a modern listener with eighteenth-century hearing and provide, even for enculturated listeners, modifications to listening in the form of increased understanding and appreciation (i.e., ‘suggestive’). Thus, engagement with Galant schemas captures both directions of the ‘loop’ previously discussed, where encultured exposure (‘listening’) can inform the category structures (‘thinking’) available to a theorist, and explicit study of the categories (‘thinking’) can modify existing nonverbal (auditory) representations (‘listening’) such that a change to hearing can take



place. Similarly, the Galant schema categories offer a strong case study for experimental verification as what it means to ‘hear’ a schema is relatively well defined; these categories are amenable to immediate recognition and identification, either in parts or as wholes, by listeners and experts alike. Contrasted with Schenkerian approaches in which the analyst attempts to ‘distance’ themselves from the musical object, Schwab-Felisch notes that:

The Galant listener, in contrast, is a contemporary in the emphatic sense: he experiences the music as a participant in a collective social, cultural, aesthetic, and performative event. It is true that he might, as a “connoisseur,” be theoretically informed, or, as a composer, might have a knowledge of composition at his disposal. But it is not the savant who is the prototype for the Galant listener of Gjerdingen’s kind, but rather the dilettante, in the original sense of an amateur who owes his formation primarily to the practice of listening (Schwab-Felisch 2014, 109).

### **Guiding Research Questions and Claims**

Much of the scholarship within music theory suggests that the loop between thinking and listening is a central and pervasive phenomenon within the discipline. However, this research does not provide explanatory power for the concrete effects of such expertise. Therefore, the current project endeavors to create a more comprehensive framework capable of explanatory power for both claims made in music theory scholarship and results from experimental research. The project revolves around the following guiding questions:

1. What mechanisms facilitate the iterative feedback loop between thinking and listening?  
More specifically, by what activities, and using what cognitive mechanisms?
2. What role do theoretical concepts play in this loop?

3. How does this loop permit change to one's hearing?
4. Similarly, what kinds of cognitive processes/tools are afforded through this kind of theoretical expertise that would otherwise be unavailable to the average, enculturated listener? In what ways does expertise provide concrete changes to cognition?

Through the course of the dissertation, I will make and support several claims that address these questions:

1. The iterative growth of knowledge proposed by the 'loop' between thinking and listening stems from the co-operative independence between verbal and nonverbal systems (multi-coding in representation, see Dual-Coding Theory below).
2. Music theory concepts are central to music theory expertise, including expertise in Galant schemata. Music theory concepts function as cognitive tools which facilitate interactions with schema categories—in a sense, they function both as a means to direct attention, and as a kind of 'container' in memory for representational information.
3. Music theoretic concepts, including Galant schemata, are constructed out of different types of representations stemming from the various ways in which theorists interact with these categories: listening, score analysis, verbalization, singing, writing, piano playing, etc. The representational make up of concepts will differ based on these interactions, and they will therefore serve different functions in expertise.
4. Schema representation is radically different between those explicitly trained in schemata categories (music theorists) and those whose interaction with the categories is from 'natural' exposure (i.e., statistical learning) alone. This is due to the distributed type of representations (verbal, nonverbal, multimodal) that experts have as a result of their more diverse interactions with schemata categories outside the realm of listening alone.

5. The effect of expertise is ‘real’ and observable. Specifically, the acquisition of ‘eighteenth century hearing’ and its impact on cognition should be observable in experimental contexts. The learning of Galant schemata therefore does concretely affect one’s hearing through memory priming, imagery, and online categorization processes.

## **Dissertation Outline**

In the next chapter, I synthesize and update Paivio’s Dual-Coding Theory (Paivio 2007) and Barsalou’s Dynamic Interpretation in Perceptual Symbols Systems (Barsalou 2003b) in order to create a cognitive framework capable of handling multi-modal and multi-system codes in long-term memory representation. The framework embraces the multi-coding view of mental representation proposed by DCT, and the more unitary and embodied approach to categorization (called simulation) in DIPSS. The framework I develop is one in which long-term memory is structurally unitary, but functionally modular: what distinguishes acts of categorization from acts of remembering particular episodic memories is a function of memory access (recall), not one defined by category abstractions which are stored separately from lived, fully embodied experiences in memory. In order to account for introspectively available states during online categorization, I also integrate this framework with auditory imagery research (Halpern 2015), metamemory (Dunlosky and Bjork 2008), and interoceptive representation in emotional cognition (Craig 2015; Barrett 2017a,b). This allows for several types of introspectively available cognitive states during categorization: imagery vividness, imagery control, feeling-of-knowing, and interoception.

In the third chapter, I use this framework to develop an embodied approach to Galant schema representation. Here I argue that current versions of schema theory are unable to account for individual or group differences in representation that stem from expertise. Using the

framework developed in chapter 2, I demonstrate that the difference in schema representation between the encultured listener and schema expertise is one grounded in centralization or distribution of representation. For the encultured listener, the representations of schemata are represented primarily in the auditory and interoceptive modalities. Contrastingly for an expert, representations of schemata are much more distributed across modality (vision, audition, haptic), and system (verbal, nonverbal). I emphasize the important role that music theory concepts play in acquiring such distributed representations as they help to promote particular types of interactions (e.g., attention to particular features or processes), explicitly encoding them into different, distributed memory traces. I then demonstrate that this distributed kind of representational structure affords much more flexibility in category simulation as experts can switch between representations of different types (auditory, motor, verbal) to more easily meet different task demands.

In the fourth chapter, I create an account of Galant schema acquisition by integrating the current framework with insights discussed in Gates (2021). I demonstrate that schema acquisition is a type of memory expertise, or long-term working-memory (LTWM). I show that traditional training practices align with a developmental model for memory expertise, with each training domain (solfège, partimenti, counterpoint) functioning within the LTWM framework. Early on in traditional training, learners focused primarily on encoding. As these learners progressed, their pedagogy encouraged more focus on skilled retrieval of information acquired in prior activities. I then contrast this by detailing a learning trajectory for a hypothetical modern learner of Galant schemata using contemporary texts. For the modern music theorist, the target domain of memory expertise is analysis. The goal of analysis is to use thinking and listening activities iteratively in order to be able to bring online particular memory representations

(simulators) at the correct time and in the correct order during listening to facilitate the hearing of Galant schema (simulation).

In the fifth chapter, I provide a concrete example of Galant schemata LTWM in action. Here I argue that the formation and modification of an interpretation in perception represents LTWM in action. As an analytical case study, I discuss schemata usage in sonata form transitions in Mozart piano sonatas. Here, I demonstrate that both the modulating Prinner and step-descent Romanesca (Fauxbourdon variant) are used in transitional spaces, providing a case for some perceptual ambiguity in these passages. I then discuss the development and results of a quantitative survey examining interpretation formation and modification in the transition of Mozart's Piano Sonata no. 2, K. 280, iii *Presto*. The results confirm most of the hypotheses garnered using the cognitive framework developed in this dissertation, demonstrating that the effects of Galant schema expertise are in fact real, and that they likely stem from the representational differences posited in chapters 3 and 4. In the concluding chapter, I recontextualize the claims from chapter 1, and propose areas for future research including application to other music theories, and future experimental research.

## Chapter 2

### **Dual-Coding Theory: Overview, Updates and Adaptation**

In this chapter, I will provide updates and adaptations to dual-coding theory (hereafter DCT) as needed for the current framework and application to music theory. I will begin by overviewing the original version of DCT, focusing on the particular aspects of the theory important for applying it to music theoretic expertise. These include the functional aspects that the theory posits are important for improved memory and cognition. I will also detail Paivio's underdeveloped views on musical expertise, noting which facets of the theory require updating for music. These include solidifying an approach to categorization, including adapting for multimodality and time (sequential organization and processing), and accounting for introspection and affect. I address each of these in turn in the last two sections of the chapter.

In the second section, I update and adapt DCT to provide an approach to concepts and categorization by integrating it with Barsalou's dynamic interpretation of perceptual symbols system (hereafter DIPSS), I seek to reconcile the two theories, noting the importance of introspection for representing abstract concepts. With regards to long-term memory (hereafter LTM), I retain a unitary but functionally modular approach to LTM structure, and a modified exemplar approach for LTM representation. I also maintain the modular distinction between LTM and working memory (hereafter WM) to account for active, online categorization. This section is concluded by outlining the integrated DCT-DIPSS theory being used for the current framework, which can account for both multimodal and sequential (time-course) categorization.

In the last section of the chapter, I account for introspection by discussing three types of subjectively available phenomena, their relationship to memory and categorization, and how they

will be integrated into the current framework. Firstly, I discuss subjective assessment of imagery through qualitative questionnaires, adapting the commonly used vividness and control assessment scales to account for different aspects of active simulation. Secondly, I account for metacognitive awareness in memory function through priming and ‘feeling-of-knowing’ judgements. And lastly, I update the framework to better account for affect through the addition of interoceptive representation as a separate representational code, which is vital for abstract concepts such as emotional constructions.

### Dual Coding Theory: An Overview

Dual coding theory (hereafter DCT) was developed by Allan Paivio as a reaction to the imagery debates in cognitive science during the last quarter of the twentieth century, in which a large number of researchers rejected the notion of analog imagistic representations in favor of a position which held that all representations were amodal, abstract, and propositional (see Pearson and Kosslyn 2015 for a summary). Paivio (1978, 1986, 2007) develops a theory that posits roles for two primary representational systems that specialize in storing and processing different forms of information: one is verbal (linguistic), and the other is nonverbal (or imagistic). Contrary to purely propositional theories of representation, DCT specifies that these different subsystems are developed through interactions in the environment, and therefore retain perceptual, motor, and affective features of these interactions (Paivio 1986, 55). Therefore, internal representations are built through knowledge derived from perceptual, behavioral, and affective experiences with the world, which retain aspects of these interactions resulting in representations that are modality-specific and thereby multimodal (Paivio 2007, 25). This approach is in accordance with more

recent embodied theories of mental representation,<sup>12</sup> which posit important roles for sensory information (motor movement, affective response, etc.) in cognition. It is important to emphasize that in DCT there exist no abstract representations; abstract functions (such as verbal categorization) may be carried out by modality-specific representations and processes, but the representations themselves are concrete because they retain their original mode of interaction in memory.<sup>13</sup>

Similarly, Paivio notes that the dual coding approach helps account for iterative growth in several types of knowledge; knowledge about the nature of our own memory and how we think (or metamemory), knowledge that can be used as a sort of feedforward mechanism to accelerate its own growth, and lastly, knowledge that draws increasingly on the cooperative activity between language and nonverbal cognitive systems (Paivio 2007, 26). DCT was designed to offer explanatory power for a wide range of findings in memory research. As will be shown, DCT offers a grounded yet powerful way to conceptualize and explain individual differences and effects of contextual factors. It is important to note that the theory was built primarily around findings from word-imagery pairs that explicitly involved visual imagery, and not auditory imagery or imagery in other modalities. This gives rise to challenges and difficulties in applying the theory to music theoretic expertise, some of which I will address in this chapter.

### **Representational Systems and Units**

DCT posits the existence of two primary representational systems; an evolutionarily older sensory or “imagery” system specializing in processing nonverbal information, and an

---

<sup>12</sup> See Barsalou (1999, 2003a), to be discussed shortly.

<sup>13</sup> For example, a ‘proposition’ as discussed in abstract theories of representation is considered to be a multimodal verbal representation stored in the verbal system in DCT.



evolutionarily newer system, specializing in processing verbal information. These two systems are functionally independent but are interconnected: they can operate independently or in cooperation depending on what is most useful in any given situation. Such “cooperative independence” may in some situations result in additive benefits from using both verbal and nonverbal systems; however, in other situations it may lead to interference effects and a decrease in overall processing efficiency (Paivio 2007, 58). Therefore, Paivio proposes that the systems evolved to allow for cooperative interaction, for selective reliance on one system or the other (particularly when one is more relevant to a given task), and for the capacity to switch back and forth between systems according to changing task demands.

Recall that DCT asserts that mental representations are acquired through interaction with the world. Paivio calls the mental representations arising from these interactions *logogens* and *imagens*, which refer to the underlying representations that are either linguistic or imagistic in nature, respectively (Paivio 1986, 59). Retaining their experientially derived characteristics, these representational structures and processes are modality specific rather than amodal as presumed in propositional theories (see Figure 2.1). DCT therefore proposes a continuity between perception and memory, as well as between behavioral and cognitive skills. Importantly, the contents and structure of representational units are not predetermined; they can change over time with experience, and will differ between individuals and groups on the basis of their differing experiences. Each representational unit (*imagen*, *logogen*) is assumed to be biologically instantiated as a distributed population of neurons (Sadoski and Paivio 2014, 52).

Orthogonal Relation Between Symbolic Systems and Sensorimotor Systems of Dual Coding Theory With Examples of Modality-Specific Information Represented in Each System.

<i>Sensorimotor Systems</i>	<i>Symbolic Systems</i>	
	Verbal	Nonverbal
Visual	Visual language	Visual Objects
Auditory	Auditory language	Environmental sounds
Haptic	Braille, handwriting	“Feel” of objects
Gustation	—	Taste memories
Olfaction	—	Smell memories
Emotion	—	Felt emotions

*Note.* Empty cells indicate absence of verbal representations in these modalities.

Figure 2.1. Representations in each symbolic system by modality (Paivio 2007, 36)

## Logogens

The term ‘*logogen*’ is used as a catch-all term for “verbal representation.” Logogens are encoded as separate representations in each modality for which they exist: auditory, visual, motor, and haptic (Paivio 2007, 37). They are activated and used in all linguistic phenomena, including recognition, memory, and production. In DCT, the logogen reflects the internal organization and size of language units as perceived and produced; this includes units such as letters, morphemes, and words, as well as stock phrases, idioms, and sequences such as long memorized poems, plays or stories (Paivio 2007, 38). When stored in long-term memory, logogens are structured in sequential, hierarchic structures in which larger units (e.g., memorized poem) are composed of different combinations of smaller units (e.g., idioms, phrases, words). Thus, a memorized poem is understood in DCT as an extended motor logogen in which already established word-level motor logogens are bound together through extended motor rehearsal into a sequentially integrated structure. As such, logogens are sequentially constrained, with their

ordering of storage and retrieval restricted into a fixed temporal sequence.<sup>14</sup> DCT does not assume that lexical representations are inherently semantically meaningful. It instead presumes that logogens derive their meaning from their connections and associations to other verbal and nonverbal representations; meaning is therefore contextual, and can differ between individuals (Paivio 2007, 38). I will return to the issue of meaning shortly.

## Imagens

*Imagens* are sensorimotor representational units that can give rise to conscious (reportable) imagery when activated (Paivio 2007, 39). They are used in many different cognitive processes involving nonverbal objects, including perceptual recognition, memory, and drawing. They exist in different modalities: there are visual, auditory (environmental sounds), haptic, motor, and olfactory (smell) *imagens*. Like logogens, *imagens* are organized hierarchically in nested sets; however, unlike logogens, individual *imagens* in the visual system are organized synchronously rather than sequentially (Paivio 2007, 39). For example, the visual *imagen* of a face is composed of smaller *imagen* representations: eyes, ears, nose, lips, etc.<sup>15</sup> The organization of a facial *imagen* is synchronous in that all of the parts are available simultaneously for processing, although may not be accessible all at once.<sup>16</sup> The mental scanning an *imagen*'s different parts is not sequentially constrained (i.e., can go in any order), unlike a lengthy verbal logogen which is constrained by sequential organization, (i.e., can only go in one

---

<sup>14</sup> For example, it is nearly impossible to recite a memorized poem backwards without re-memorizing and rehearsing it in such a manner. See Paivio (2007), 38.

<sup>15</sup> It is also important to note that all *imagens* have some motor components. For example, visual *imagens* are formed through sensory interactions with the eyes, which involve motor movement (see Paivio 2007, 39). In this sense, any visual *imagen* learned through the eyes are inherently multimodal as they also are stored in association with motor movements of the eyes.

<sup>16</sup> In visual imagery one can imagine a face with all its complete parts available at once, however if one were to inspect one part of the facial image, say for example the eyes, the other parts of the facial *imagen* would no longer be available.

order), constrained by the structure of language in listening, reading and speaking. Importantly, imagens have properties which are analogous to those of their external perceptual referents, whereas logogens do not. Thus, imagens correspond to natural objects or groupings of objects, and retain both static and dynamic properties of these objects perceived through the senses; these properties can vary continuously in shape, size, color, and other dimensions (Paivio 1986, 59). Logogens, by way of contrast, are structurally discrete and arbitrary (Paivio 2007, 40).

### **Connections and Activation Processes: A DCT Approach to Meaning**

DCT takes a functional approach to meaning: the “meaning” of a given term or concept is equivalent to the pattern of activation of representations in response to an internal or external prompt. This activation pattern can be within or across systems (recall that imagery and verbal systems can operate independently or in cooperation). Thus, activity in one system can initiate activity in the other, and vice versa. Activation patterns are conditioned based on the strength of the interconnections within and between the systems, developed through prior experience. Crosstalk between systems is developed the co-occurrence of different activities and behaviors (Paivio 1986, 63). The functional connections between logogens and imagens are multimodal, can operate bidirectionally, and can be activated either automatically or through conscious effort (for a summary, see Figure 2.2). For example, names of presented objects may or may not be generated automatically but such automaticity is highly likely under certain conditions. If one is asked to imagine and describe their dining room table, they can consciously imagine the table, then provide verbal descriptions of it (Paivio 1986, 62). Therefore, the probability of activation of various representations is a function of the combined effects of stimulus attributes, individual differences, and other contextual factors.

Paivio defines three types of activation in DCT: a direct form of activation called representational, and two forms of indirect activation called associative (within-system) and referential (across-system). These will now be addressed in turn.

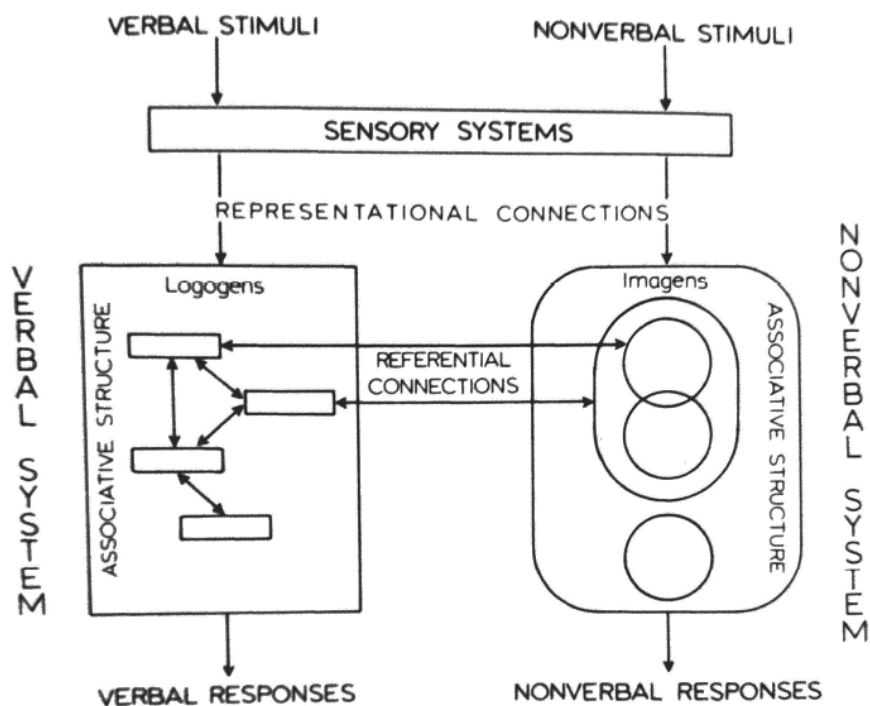


Figure 2.2. Diagram of The Subsystems and their Relations (Paivio 1986, 67)

### Representational activation

Representational processes involve the relatively direct activation of verbal representations by linguistic stimuli and nonverbal representations by nonlinguistic stimuli, as in, the activation of the visual logogen *cat* through the reading the word *cat* (i.e., logogen in the visual mode), or the activation of the imagen of a cat by seeing a cat. Representational “meaning” therefore implies that an imagen or logogen corresponding to a verbal stimulus, or an object is available for further processing (Paivio 2007, 41). This level of activation therefore corresponds to meaningful perception whereby a representation (logogen or imagen) is initiated

directly by a verbal or nonverbal stimulus. Paivio notes that this type of activation is commonly measured by stimulus familiarity and is DCT's best predictor of how easily a stimulus can be recognized when presented very briefly or in a degraded format (e.g., a fuzzy or blurred image) (Paivio 2007, 43). Recognition is probabilistic and depends on a similarity match between the stimulus and an internal representation selected from multiple likely candidates. When an object or word is viewed, the sensory pattern "homes in" on the most similar imagen or logogen (ibid.). Another test of the strength of representations in direct representational activation is figural closure where an incomplete word or picture is presented, and the perceiver is easily able to complete the word or picture. Importantly, Paivio notes that representational activation can be greatly improved by training procedures (e.g., discrimination training) so that very different stimuli can become functionally equivalent with regards to activation, or it can be narrowed so that a recognition response only occurs to stimuli that do not differ by more than a certain amount from each other (Paivio 2007, 43).

Representational activation may also be indirect if logogens and imagens are activated internally without an external prompt. Given the substantial overlap between perception and imagery in terms of cognitive processing, it is likely that similar though not identical representations are activated in perception and imagery.<sup>17</sup> DCT attributes this to different activational pathways for representations activated directly through a stimulus and those activated via internal pathways (Paivio 2007, 49). However, there is evidence to support that perception and imagery draw from a similar subset of representations, particularly given that

---

<sup>17</sup> This is supported by recent findings in musical imagery research that shows somewhat distinct yet partially overlapping neural pathways for music perception and imagery (Zatorre and Halpern 1993; Zatorre et al., 1996; Halpern and Zatorre 1999; Herholtz et al., 2012; Zatorre et al., 2009).

perception and imagery have been shown to interfere with one another.<sup>18</sup> For example, Paivio suggests that when we see a ship, the ship seen “homes in” on a relatively small set of similar ship imagens (or exemplars), with a particular imagen being directly activated when the ship is close enough to be identified. When asked to generate the picture of a ship in the mind’s eye however, we draw from a larger pool of imagens, and so it takes more time before a particular imagen is activated to form a conscious image (Paivio 2007, 43). Importantly, the size of the representational pool (i.e., number of representations) and the availability of certain representations (i.e., what gets activated) depends on prior experience: the breadth, depth, and recency of experience in the sensory or perceptual domain.

#### Associational activation

Associative processing refers to activation of representations within either system by other representations within the same system. Associative activation is always indirect and internal: it occurs when one already activated representation activates another. For verbal phenomena, this is seen in word-pair associations, commonly measured through word association tasks. For example, the logogen for “cat” may be tightly coupled with “dog;” once the logogen “cat” is activated, it may activate and thus make available the logogen “dog.” For nonverbal phenomena, the associations are often synchronous (at least in vision) so that parts bind into integrated wholes (e.g., nose, eyes and mouth bind together to form a face) (Paivio 2007, 50). Nonverbal association can also involve imagen wholes; for example, imagining the male, red-colored bird “cardinal” may prime or activate the female equivalent (mostly brown, with tinges of red features). Associative activation on the verbal side is typically more systematic

---

<sup>18</sup> Research suggests that these interference effects are modality specific, so that visual signals are interfered with more by visual imagery than by concurrent auditory or motor imagery (Paivio 2007, 49).

as determined by conventions in language use, whereas association on the nonverbal imagery side is much less systematic.<sup>19</sup>

Lastly, any form of multimodal or cross-modal activation within a representational system is associational activation in DCT. For example, on the verbal side, an activated visual logogen “dog” can activate, through association, the auditory logogen “dog.” Therefore, through improved associative processing, it becomes possible to listen to a story spoken out loud and track the same story in visual text. Likewise, the auditory imagen of a phone ringing may activate, through association, the visual imagen of a phone, or motor imagens of interacting with the phone. DCT therefore posits rapid intramodal transfer through associational processing, rather than involving some form of amodal system which becomes instantiated modally following activation. Associative processing can be improved with training or experience so that the activational patterns become more seemingly “automated” over time. (Paivio 2007, 51). The form which associative processing takes also varies depending on the representational system. As previously noted, imagens within the verbal system are constrained sequentially, and therefore association is constrained sequentially. Visual imagens in the imagery system are not constrained sequentially but are synchronous and are therefore available simultaneously for selection. They are not, however, simultaneously available for processing; for example, one cannot imagine a room from all angles at once or focus on a certain area of an imaged room at the same time. One can, however, scan the imagined room in many different orders; therefore, associational processing in the visual imagery system is less constrained than in the verbal system. Paivio briefly notes that in real world, multimodal situations, both synchronous and

---

<sup>19</sup> From a DCT perspective, associational processes on the non-verbal side can help integrate separate parts into higher order memory or perceptual units, i.e., form chunks in long-term memory. A chunk in long-term memory is therefore formed through repeated associative pairing in which a pair functions as an integrated unit.



sequential processing occur because other forms of imagery (motor, auditory) are sequentially constrained (Paivio 2007, 52). He does not elaborate on this as DCT focuses almost exclusively on visual imagery. I will address this issue in the following chapter.

### Referential activation

Referential processing is a form of cross-system, indirect activation of the nonverbal system by the verbal system and vice versa. For example, the presented word “dog” will first directly activate the logogen “dog,” which then can activate one or more dog imagens in the nonverbal system. Referential meaning is therefore derived from the relations between words and their referents internalized as associations between logogens and imagens. Such activation is usually measured by the nameability of presented objects and the image-evoking potential of a verbal stimulus (Paivio 2007, 43). This kind of processing is optional and probabilistic, not automatic, and obligatory meaning that the word “dog” need not activate the image “dog” in every instance. The probable pattern of activation depends on long-term memory, recent experiences, and the context in which activation occurs. This form of activation is therefore quite dependent on prior learning and experience, which can drastically alter the types of referential connections, their probability, and their speed of activation.

### Dual Coding Functions

Dual coding in the human perceptual and cognitive system serves several broad functions that provide various benefits to cognition. The most notable of effects, which DCT was developed to explain (Paivio 2007, Chapter 4), are for memory: mnemonic functions, multiple codes for improved memory (code additivity) and benefits of concrete imagery (what Paivio calls the “conceptual peg” hypothesis, to be described below). Other functions, all of which

derive from interactions of verbal and nonverbal codes in memory, include emotion and motivation, communication and meaning, recognition memory, prediction and anticipation, and creativity (extended from iterative growth in knowledge) (see Sadoski and Paivio 2014, 45-48).

One view particularly unique to DCT is that dual coding is directly responsible for evaluative and cognitive control functions. According to this view, there exists no separate abstract central executive system for controlling one's own mental processes; this is merely a function of dual coding interactions within and across systems:

Any structuring or monitoring functions occur within and between the systems themselves, and not from separate, disembodied sources such as abstract schemata, propositions, or central executives. For example, we can regulate our cognitive behavior through self-regulatory inner speech and/or through imagining alternative scenarios from different points of view (Sadoski and Paivio 2014, 29).

The implications of this view will be discussed and updated shortly. I will turn now to the primary function for memory bootstrapping for which DCT was developed, namely code additivity and the conceptual peg hypothesis, as well as representational exchange within and across systems.

Bootstrapping and improving memory: code additivity and the conceptual peg hypothesis

In DCT, improvements to memory are thought to occur through code additivity. Representations in DCT are made of multiple traces or codes; each 'word representation,' for example, is stored in multiple traces distributed across modalities: vision, spoken, auditory, haptic. In the case of concrete language there will also be referential connections to related

nonverbal representations, distributed across modes. Each of these traces can operate independently or simultaneously during recall; distributed across systems (verbal/nonverbal) and modes (auditory, visual, haptic, etc.), there are multiple traces through which to recall information rather than relying on a single trace (Paivio 2007, 72). However, Paivio's research has shown that the nature of the code distribution also has an effect on memory, with concrete imagery playing a vital role in improved memory and recall. This led to the development of the "conceptual peg" hypothesis.

The conceptual peg hypothesis was developed through the course of Paivio's work on the effects of language (concrete/abstract) and imagery on memory and recall of stimuli pairs (e.g., word-to-word, word-to-image). Paivio (1965) found that pairs of concrete words, such as *coffee-pencil*, are remembered better than abstract word pairs such as *virtue-fact*. This effect was not attributed to word association, but to the imagery value of concrete words, which facilitated integration of the two words into an integrated unit (for example, imagining a pencil inside a cup of coffee). Similarly, the effect was found to boost memorability for concrete-abstract word pairs, such that *coffee-fact* was remembered better compared to pairs of abstract words, indicating that it was the concrete word which acted as a strong retrieval cue for the pair.<sup>20</sup> From this, Paivio hypothesized that mental images play a central role in the organization of information and its retrieval from memory, serving as 'conceptual pegs' onto which memory episodes are 'hung,' providing benefits to later retrieval. Paivio expanded this work throughout his career to investigate other variables on the conceptual peg effect, including examining the effects of pictures versus words and the effects of imagery instructions on tasks (Paivio 2007, 60,

---

<sup>20</sup> This effect was also only found for word pairs in which concrete words came first, indicating that the high imagery word serves as a better retrieval cue for information following the peg, not before. The pair in order *abstract-concrete* was still be found to outperform *abstract-abstract* pair but had poorer performance than the *concrete-abstract* ordering (Paivio 2007, 65).

67). This large body of work found a *picture superiority effect*, showing that verbal recall is consistently better for picture-pairs than word-pairs, and that this benefit extended when pictures and words were combined (e.g., *picture-concrete word* had better recall than *concrete word-concrete word*). From the DCT perspective, this is because pictures directly activate imagens (through representational activation) in the nonverbal system, which prompts a better imagery response pairing for the formation of a conceptual peg, whereas two concrete words only activate imagens *indirectly* through referential processing across system. Similarly, explicitly instructing participants to use imagery in these types of tasks results in better performance (Paivio 2007, 68).

The conceptual peg hypothesis has since been extended to explain verbal/nonverbal interactions during the reading process (Sadoski and Paivio 2014). Here, Sadoski and Paivio suggest that concrete language helps to form conceptual pegs around which temporary reading episodes are stored and comprehended. In this way, mental imagery plays a central role for chunking in memory. These chunks, defined by concrete conceptual pegs (images generated in the nonverbal system) are vital for linking, integrating, and unifying reading episodes stored in episodic memory (Sadoski and Paivio 2014, 78), facilitating text cohesion and comprehension through time (*ibid.*, 103-104).

Representational exchange and creation: within-system association and intermodal transfer

In addition to the memory improvements found by distributing codes across systems (verbal/nonverbal) and modes (visual/auditory/haptic), a central feature of memory bootstrapping and iterative growth in knowledge in DCT is that representational units can be ‘exchanged’ for those that are ‘equivalent’ (Sadoski and Paivio 2014, 62). Within each system, this exchange can take the form of intermodal transfer: ‘exchanging’ one representational unit stored in one

modality (e.g., vision) for a similar one stored in a different mode (e.g., vision). Within the verbal system, this intermodal transfer is relatively instantaneous and is completed often during language use. For example, the visual imagen for the word *cat* can be exchanged for the auditory-motor imagen */kat/*, and so on. Because verbal representations exist in a large network of associations, word-level logogens can also be exchanged for other equivalent words (acquired through experience and association in memory), as in the case of synonyms. Nonverbal imagens can also be exchanged within the nonverbal system, such as the visual imagen of a *cat* and the auditory imagen of a *meow*. Through referential connections across systems, each of these units (logogens *cat* and */kat/*, imagens for *cat* and *meow*) can also be exchanged for one another. It is precisely the mind's capacity for rapid intermodal and cross-system transfer that affords iterative growth in knowledge, memory bootstrapping, and even creativity.<sup>21</sup>

### **Paivio on Musical Expertise**

In *Mind and its Evolution*, Paivio extends DCT to discuss theoretical application to expertise acquisition, including musical expertise (see Paivio 2007, chapters 14 and 18). Understandably, his conception of musical expertise is decidedly limited as he primarily discusses instrumental performance and composition. He correctly identifies these forms of expertise as primarily sensory and nonverbal as they do not involve substantial verbalization; however, he notes that because musical performance specifically involves fine motor skills, it shares characteristics with language and verbal skills, notably the sequentially organized nature of the representations (Paivio 2007, 330-331). However, he goes on to say:

---

<sup>21</sup> For example, each representational unit has slightly different associative and referential connections within and across systems (i.e., visual logogen *cat* may have different connections in memory than does the visual imagen *cat* or even the auditory-motor logogen */kat/*). Being able to switch between these different representational units therefore provides near instantaneous access to a slightly different pool of representations.

Beyond that...the defining characteristics are fundamentally different...music segments and their temporal patterning are not meaningful in the same sense as language. Music has aesthetic and emotional meaning, but it has no referential meaning. The changing patterns in music are like prosodic changes in speech unaccompanied by meaningful phonemic changes (Paivio 2007, 331).

Similarly, when outlining DCT, Paivio notes that

Music and dance are also important nonverbal symbolic forms because they are ancient in origins...and universal across societies although not across individuals. They are not representational in the same sense as imagery (drawings) or language. We perform, compose, and we think about music or dance, but not with them. There are parallels between the sequential patterning of language and the written notational systems for music and dance choreography, but they do not stand for anything outside of themselves. Of course, here too we can point to ambiguous cases such as the imitative acoustic patterning of Rimsky-Korsakov's "The flight of the bumble bee" or a sexy dance, but these are representational only in the limited sense of onomatopoeia in language (Paivio 2007, 35).

This conception of musical expertise is problematic for several reasons. Firstly, it assumes that only those sensory representations that are inherently tied to language through referential connections are "meaningful," which is somewhat contrary to Paivio's view of functional meaning. Paivio is therefore implying that sensory representations which are only activated by sensory stimuli and are only connected to other sensory representations through association do not have meaning, or at least do not have the same sort of meaning provided by connections with

logogens.<sup>22</sup> Similarly, Paivio offers a rather uninformed and underdeveloped understanding of musical thought, noting that “[w]e perform, compose, and we think about music or dance, but not with them” (Paivio 2007, 35), suggesting that thinking with sensory representations such as imagined actions and sounds is somehow not thinking “with music” but merely about it.<sup>23</sup>

While these ideas may be quite uninformed, it is clear that part of music theoretic expertise is necessarily about adding referential connections between language and sensory representations, providing what Paivio refers to as a form of enhanced representational meaning afforded by interplay between language and sensory systems. Paivio seems unaware of such musical training in which verbalizations do in fact play an important role, like music theoretic training of the North American, collegiate-level kind. He does recognize the importance of some verbalization in music performance expertise, but these only relate to refining deliberate practice routines in the acquisition of perceptual-motor skills.<sup>24</sup> The only other form of referential processing Paivio recognizes in musical expertise is perfect pitch acquisition, or the ability to recall a note name label given a single pitch (Paivio 2007, 332). Two areas that Paivio correctly

---

<sup>22</sup> He notes that dance and music representations are not “representational” in the same sense as imagery or language. He does not unpack this further, which implies that it is the referential potential between visual based imagery and language that must provide some differentiated form of representational meaning. It may be that sensory representations when activated representationally are meaningful because they prime the verbal system. It may also be the nature of visual stimuli; they can stand on their own regardless of context. For example, an apple remains an apple regardless of context (i.e., if it is with other fruit, or located on top of a train). However, “objects” in music and dance are more relative; they are constructed in a network of internal relations rather than being more fixed (e.g., a PAC in one key can act as a tonicized half-cadence in other, therefore context matters). In this sense, music only has vague representational meaning (patterns in imagery or affective responses) because they only ‘refer to themselves,’ and not to larger meaning units (language) through referential meaning and are therefore more ‘self contained.’

<sup>23</sup> This recalls the distinction discussed in the first chapter by Karpinski and other scholars in differences in musical imagery ability. Karpinski notes a difference between thinking *about* music and thinking *in* music (Karpinski 2000, 3), where thinking about music is primarily verbal and thinking in music is primarily imagistic or sensory (the opposite of what Paivio suggests here). However, another distinction is made between “meaningful” imagery (thinking “in”) music, and “general sound replay” (thinking “about”) by scholars (see Karpinski 2000; Gordon 2012 on audiation), which may be what Paivio was vaguely getting at here (imagery replay versus “deeper understanding” of some kind).

<sup>24</sup> For example, incorporating verbal feedback from instructors, including those related to specific performances or skills, as well as more general strategies (e.g., how long to practice, what to focus on, etc., Paivio 2007, 331).

identifies as important for music performance expertise are the multimodal nature of musical representations, including the issue of sequential constraints on musical processing (and their similarity to verbal representations as stated earlier), as well as the importance of affective responses and imagery in expertise acquisition (Paivio 2007, 332-333). However, given that the forms of musical expertise discussed by Paivio do not include concurrent and consistent cooperative activity between verbalization and sensory systems, he does not propose how these two areas might be handled by DCT. Therefore, these are the two primary issues that I will unpack to properly adapt DCT to music theoretic expertise. I will do so by specifying an approach to categorization, including multimodal and time-course aspects of musical categorization, followed by a detailed account of three types of introspection—imagery, feeling-of-knowing and interoception.

### Adapting and Updating DCT: Defining an Approach to Concepts and Categorization

Paivio himself never provided a detailed account of concepts and categorization; the theory has, nevertheless, been used as a means for understanding the organization of concepts in semantic memory as grounded in the brain's modal systems (McRae and Jones 2013, 207). In this way of thinking, concepts—the mental representation of categories (groups) of real-world instances (Goldstone, Kersten and Carvalho 2012, 608)—are networks of highly probable, interconnected, multimodal mental representations distributed across both sensory and verbal systems, bound through associative (within-system) and referential (cross-system) connections. Categorization, the process by which mental representations (concepts) determine whether some entity is a member of a category (Rips, Medin and Smith 2012, 177), is defined as the pattern of



activation of representations in an individual's 'concept-network' within and across verbal and sensory systems cut across each modality. This view is consistent with modern neurological theories of concept representation in which a concept is a group of distributed patterns of activity across a population of neurons (Barrett 2017a).

In DCT, these patterns of activation are contextual, meaning that they will vary depending on the context of categorization (surrounding, background context, goals, stimuli) in a systematic manner. For example, DCT predicts systematic differences based on the manner of activation of a pool of representations (e.g., representational through the verbal system), and on selective reliance on either one system (e.g., verbal) or co-operation between systems (e.g., verbal to nonverbal and back and forth). Categorization goals and context will affect which representations are recruited during categorization. DCT predicts differences will vary systematically from instance to instance based on whether category representations are directly or indirectly activated (e.g., perception vs. imagery), and whether one or both systems are used, and in which order. Concept representation in DCT is therefore dynamic; each instance of categorization (use of representations) both shapes and redefines the content and structure of memory (i.e., it is partially recursive). This occurs because the contents and connections between representations in DCT are probabilistic and are defined by prior experience, recent use, and exposure, and vary within individuals over time. This position is accordant with approaches to categorization that posit intrinsic variability in human concepts, within individuals and developmentally over time.<sup>25</sup>

---

<sup>25</sup> Goldstone, Kersten and Carvalho (2013) discuss this in reference to category equivalence classes: how widely variable instances of a category can be treated as equivalent in categorization. They also note that each instance of use of a category is unlike to have the same meaning, also articulated by DCT as the variability in network activation (i.e., changes in probability of representational, associative, or referential activation). Barsalou (1999, 2003a) also posits inherent variability in concepts, to which I will now turn.

Given that Paivio never provided a detailed account of concept representation and categorization processes using DCT, it will be useful to place his theory in dialogue with theories of concept representation. Lawrence Barsalou's Dynamic Interpretation in Perceptual Symbol Systems theory (DIPSS, Barsalou 1999, 2003a) is one such modern, embodied approach to concept representation. While DCT and DIPSS differ in important ways (see Paivio 2007, 118; Barsalou, Ava, Simmons and Wilson 2007, 253-254), when integrated, each helps to supplement the undertheorized portions of the other. DIPSS supplies a precise account of dynamic concept representation and categorization behaviors, particularly when compared to traditional approaches to the topic, whereas DCT provides a more systematic account of the underlying representational structures and processes used in such concept representation and categorization behaviors. I will start by outlining the DIPSS approach to concept representation in light of DCT, continue with a section identifying and resolving conflicts and weaknesses between the two theories and conclude by summarizing and demonstrating the current approach to categorization developed in this framework.

### **Barsalou's DIPSS and DCT**

Much like DCT, DIPSS is a fully embodied theory of knowledge representation. It assumes continuity among perception, cognition, and memory, and rejects the notion of abstract, amodal representation of knowledge. Instead, DIPSS proposes that information is stored in the form of a *perceptual symbol*, which, much like the DCT notion of an *imagen*, maintains perceptual information and the mode of interaction in memory (i.e., visual features through the visual system). DIPSS, unlike DCT, does not make a modular distinction between verbal and nonverbal systems and representations; Barsalou does note that language representations are

fully perceptual and not abstract, formed much in the same way as perceptual symbols.<sup>26</sup> Like DCT, DIPSS also proposes that the perceptual symbol is akin to a pattern of neuronal activation, but generally proposes more depth as to exactly how perceptual symbols and their associations may be distributed across modal systems and up the neuronal hierarchy.<sup>27</sup> A perceptual symbol is therefore a partial record of a brain state during perception; the symbol only comes to represent what was perceived during selective attention, and does not represent the entire brain state during the act of perception. Perceptual symbols have a dynamic aspect: because a perceptual symbol is an associative pattern of neurons, its activational patterns can vary widely depending on context such that later reactivation of that symbol may also only be partial, distorted or modified in a new instance of reactivation.<sup>28</sup> This is similar to the DCT notion of contextual activation of imagens and logogens: as context varies, different logogens and imagens will be activated, and the probabilistic connections between representations will shift over time as a result of this experience.

In DIPSS, a concept representation is akin to what Barsalou calls a *simulator*, a distributed network of associated and interconnected multimodal perceptual symbols stored in long-term memory that come to represent a category (Barsalou 1999, 586; Barsalou 2003a, 1180). The network connecting perceptual symbols is called a ‘frame.’ Frames, like the concept of a schema or script, provide the necessary ‘background information’ for use of concept knowledge (simulators) by associating a wide array of perceptual symbols into a probabilistic

---

<sup>26</sup> DIPSS does not theorize particularly deeply about language representations or language behaviors, it focuses primarily on nonverbal simulation. However, much of the simulation behavior discussed in DIPSS involves language, which makes the integration with DCT advantageous. Language use was later integrated with DIPSS under the Language and Situated Simulation Theory (LASS) (Barsalou *et al.*, 2007); however, given the lack of specificity regarding the structure and representation of language units, DCT still provides a better account of language-imagery interactions than DIPSS or LASS.

<sup>27</sup> Largely explained by invoking association areas called ‘convergence zones’ which are able to functionally reactivate or re-enact sensorimotor and introspective states (Barsalou 2003a, 1180).

<sup>28</sup> Like a connectionist attractor, see Barsalou (1999), 584.

network (Barsalou 1999, 590). A simulator is therefore both the knowledge and the accompanying processes that allow an individual to represent some object or event. In DCT, the DIPSS simulator is akin to the vast interconnections between logogens and imagens, within and across systems. As is the case in DCT, the associations that bind perceptual symbols in DIPSS are developed through experience. However, DIPSS places more emphasis on selective attention as a factor determining the probability of association between perceptual symbols. DIPSS posits that as different members of a category are encountered, attention is drawn to particular components or configurations of features, activating similar patterns of neurons across instances. This results in the formation of simulators for that category over time.

In DIPSS, simulators come in two different types: *property* and *relation* simulators. A property simulator comes to represent a particular feature of a category that is repeatedly processed using selective attention during categorization (e.g., *nose*), whereas a relation simulator comes to represent multiple aspects of a category's member and their configuration (e.g., the spatial relation *above*; that a *nose* occurs above a *mouth*) (Barsalou 2003a, 1180-1181). A virtually unlimited number of simulators can be acquired by the human conceptual system, which provides profound flexibility and dynamic rather than fixed representation. DCT, on the other hand, does not make an explicit distinction between properties and relations, and instead posits dynamic variability in the size of representational units (which in effect captures much of the same variability and unit types).<sup>29</sup> In DIPSS, simulators are used in simulation; the active, online process of re-enacting a perceptual, motor and/or introspective state (Barsalou 2009, 1281). From this view, a concept is a simulator that can construct an infinite number of

---

<sup>29</sup> For example, in DCT, visual logogens can occur in any number of sizes; from individual features of letters to short letter combinations (morphemes), to words, and whole phrases. Similarly, visual imagen units can include individual features (eyes, nose), or holistic wholes (face, body).

simulations tailored to meet the needs of the specific context. From a DIPSS perspective, concepts are viewed as dynamic representational networks, and the act of categorization as a specific instance or implementation of a specific simulation (i.e., use of part of a network). As is the case with DCT, this approach to categorization is dynamic in that it accounts for a wide variety of perceptual and cognitive activities including expertise and growth of knowledge,<sup>30</sup> category inference, prediction, and novel category generation (Barsalou 2003a, 1183).

Categorization or conceptualization—the process of using category information to achieve a goal—is viewed in DIPSS as a form of *situated simulation* or *situated conceptualization* (Barsalou 2003b; 2009). From this perspective, conceptualization is better understood as a skill or ability, rather than a (perhaps epiphenomenal) feature of static, abstracted memory representations. Situated conceptualization, grounded in the many diverse sets of acquired multimodal simulators for objects, actions, and events, are tailored to help an individual in a particular instance. Different simulations are constructed to support different conceptualizations needed, as for example about the concept *bicycle*—riding a bicycle, locking up a bicycle, or repairing a bicycle (Barsalou 2009, 1283). Each of these may simulate slightly different aspects of experience, formulated through accessing different parts of a simulation network (i.e., ‘frame’). Some simulations may be focused more on perceptions of the object in question, such as focusing on a bicycle’s handles; others may simulate more strongly action sequences, for example the movements needed to ride a bicycle; still others may simulate more strongly available introspections or event settings, perhaps the feelings of happiness typically associated with riding a bicycle on a summer vacation (see Barsalou 2009, 1283). Many

---

<sup>30</sup> Barsalou notes that each new simulation, e.g., successful categorization, is stored in memory for later use. If a similar event or object is encountered later, i.e., it matches the existing case in memory, then it is assigned to that category (Barsalou 1999, 587). This process is iterative and recursive.

instances, like the act of imagining riding a bicycle (a situated simulation), will likely involve components of each of these (perception of objects, relevant actions, introspections, and event settings). In this way, no conceptualization is ever formed in isolation, abstracted apart from the objects, events, actions or internal states acquired while interacting with the world. As with DCT, DIPSS claims that such simulations come to represent concepts organized from most to least probabilistic, by “dominance order” (Barsalou 2003a, 1180-1181, Barsalou 2005; see Figure 2.3). For example, when prompted to simulate various properties through verbal instructions (e.g., imagine a *nose*), the simulation generated is typically the most frequently encountered instance in its appropriate context, such as a human nose in the context of a face (Solomon and Barsalou 2001).



Figure 2.3. Dominance Order of Simulators (Barsalou 2005)

## **Reconciling DIPSS and DCT: Language, Introspection, and Abstraction**

While DCT and DIPSS are largely congruent and complementary theories of representation, they do diverge in a few important respects. Paivio and Barsalou have briefly engaged with each other's work and noted some of these conflicts (see Paivio 2007, 118; Barsalou, Ava, Simmons and Wilson 2007, 253-254); however, I interpret these disagreements as partially a function of undertheorizing, misunderstanding, or lack of engagement with the other's research. The points of divergence which I will outline here and unpack more fully below are twofold. Firstly, DCT posits a more vital, central role for a modular language system and provides a more systematic view of memory and cognition as specific interactions within or between representational systems (i.e. selective reliance on a single system, or between systems in the case of logogen-imagen interactions). By way of contrast, DIPSS focuses more on simulators and simulation as stemming from the nonverbal system. Barsalou takes issue with Paivio's emphasis on language, specifically as it relates to the understanding of abstract concepts; here, Barsalou argues for a more central role of introspection over language. I will demonstrate that these differences between DCT and DIPSS are largely complementary; the addition of verbal codes from DCT to simulators and simulation provides a more complete and systematic view of situated conceptualization, and the addition of introspection helps to clarify the role of the nonverbal system in abstract concepts.

Secondly, Paivio criticizes DIPSS for its reliance on representational abstraction, particularly the schematic nature of perceptual symbols and frames, and embraces an exemplar view of memory representation in lieu of one that posits the formation of prototypes. This position is not nearly as incongruent with DIPSS as Paivio seems to believe, and the authors' positions on the nature of the memory trace in long-term memory are more similar than not. The

addition of the DIPSS distinction between simulators (long-term memory organization) and simulation (online use of that information held in working memory) actually provides a powerful means to discuss abstraction, expertise and individual differences in memory representation and use. Below I will outline the differences between DCT and DIPSS in terms of their approaches to abstraction in long term memory, with reference to conventional approaches to memory and concept representation (explicit/implicit, declarative/procedural, semantic/episodic, exemplar/prototype). I will then provide a more detailed account of simulation by integrating it with Baddeley's theory of working memory.

The concrete-abstract continuum: on the role of language and introspection

In DCT, concepts are understood to fall along a concrete-abstract continuum. Concrete concepts are those more grounded in nonverbal (imagen) based representations; their concept networks contain many associated imagens, each containing properties analogous to those which external objects possess, as garnered through the sensory systems, and which are directly referentially connected to language. Examples of such concepts are 'apple' and 'horse.'

Contrastingly, abstract concepts are those primarily grounded in language (logogen) based representations; their concept networks contain many interconnected logogens and are indirectly connected to imagens in the imagery system through associations to more concrete (typically object) logogens. Examples of such concepts are 'truth' or 'religion.' These concepts are abstract because their meaning stems primarily from a large network of verbal associations rather than direct connections to imagens. These concepts can still be grounded in nonverbal representations, but this occurs through an intermediate, verbal representation. For example, *religion* can be 'grounded' through association to *church*, a more concrete concept containing direct referential connections to imagens which can therefore result in conscious imagery of churches (Paivio



2007, 46). This concrete-abstract dichotomy is supported by experimental findings showing that naming and imageability reaction times are faster for concrete pictures and words than for abstract ones (Paivio *et al.* 1988). It is also supported by recent neurological findings showing that concrete concepts engage more imagery networks in the brain, while abstract concepts are more centralized in the verbal system (Wang *et al.* 2010).

Barsalou *et al.* (2007) is particularly critical of Paivio's approach to abstract concepts. While Barsalou acknowledges many similarities between DCT and his proposed LASS (language and simulation in conceptual processing) theory—such as the grounding of representations, both verbal and nonverbal, in sensory modalities, and on the selective use of mixtures of verbal and nonverbal processing to accommodate different tasks—he is critical of DCT's reliance on depth of processing in the verbal system alone to explain abstract concepts. This can be understood as the notion that abstract concepts arise primarily from verbal associational processing in the verbal system, rather than processing in the nonverbal system. Barsalou *et al.*'s (2007) LASS theory places substantially more emphasis on the simulation system (i.e., nonverbal processing), proposing that abstract concepts are more grounded in introspection stemming from simulation rather than verbal association. Introspection here is rather vaguely defined, but includes affect, motivation, intentions, and metacognitions (Barsalou 2009, 1281). Barsalou suggests that introspections can become represented by a perceptual symbol, much like a perception or an action (Barsalou 1999, 600); however, he does not specify which introspective states would be represented in a particular symbol type, and how.

Paivio is generally skeptical of the invocation and use of introspection as a tool for revealing aspects of memory performance (Paivio 2007, 14). While his research has successfully used subjective accounts of 'imageability' for concrete words as a means of assessing the

potential for referential processing and has investigated the effects of explicit instructions to use imagery on task performance (Paivio 2007, 59), Paivio still suggests that consciously available, introspective evaluations, such as the subjective assessment of vividness in imagery, are problematic as constructs for use in experimental research (ibid.). However, DCT does account for some introspection in the form of affect in emotional processing (see Figure 2.1 above). Unlike DIPSS, in DCT emotion is not thought of as a distinct representation (e.g., perceptual symbol/imagen), but is instead a result of activation of networks of other representations. Emotions are therefore a type of internal response developed in reaction to (or through interactions with) objects, events, and people. To reiterate, they are not a type of imagen—but are instead represented as dormant connections between representations for language, objects, and events—or as felt sensations during a particular event (Sadoski and Paivio 2014, 65). In DCT, emotions are treated as a sort of nonverbal modality, the representations of which are acquired and stored only in the nonverbal system. Emotions are not directly represented in the verbal system, much like taste and smell, as logogens are not acquired through these modalities. That is, ‘emotions’ as a sensory modality are not directly used in language production, unlike, for example, the motor modality, which acquires and uses both sensory representations (motor-imagens) on the nonverbal side and verbal representations (motor-logogens) on the verbal side. Emotional states are, however, indirectly connected to emotional language, which allows for communication about emotional states. According to DCT, this connection is mediated by other sensory representations (e.g., imagens) with which affective responses are tightly associated, and through which referential connections to language representations are made. Therefore, in DCT, affective responses are directly associated with nonverbal representations gathered from interactions with objects, events, and people. These imagens of objects, events, and people are

themselves connected referentially to affective language, suggesting that emotional representations have no direct cross-system (referential) connections to language. This position posits that activation of an emotional state must follow activation of a relevant stimulus or image, meaning that an affective response is a form of cognitive evaluation (Paivio 2007, 95). Abstract language is therefore more likely to activate large networks of representations—many associated logogens and their referential connections to imagens—making these words more context dependent. While people may report concurrent spatial or haptic images arising from abstract words, such as the prepositions *under*, *inside* and *with*, these arise from the activation of dynamic situations in which actions involving these prepositions occur—essentially, the activation of associated representations in a network of related representations (Sadoski and Paivio 2014, 63).

Paivio's brief account of emotion and affect from a dual-coding perspective has some radical implications for cognition and processing. In particular, the dual-coding assumption that emotional states are mediated through nonverbal representations suggests that such states are likely to be unavailable in the absence of a prompting stimulus or event and are therefore always "indirectly" activated through association in the nonverbal system. This also posits representational distance between affective states and emotional language because referential cross-system connections are also only mediated through other imagens and are not directly connected to emotional states. This position seems quite counterintuitive, especially as emotional language appears to be very closely connected to affective states. Paivio notes that this conception occurs because emotion words acquire generalized affective (sensory-like) qualities analogous to referential meaning through conditioning so that the activation of an emotional state through language becomes increasingly more probabilistically certain and rapid over time

(Paivio 2007, 95). However, because these connections are still mediated through associated nonverbal representations, DCT hypothesizes that processing times for emotional states would be slower through language than through nonverbal activation. Paivio (1978) confirmed this hypothesis, finding that when judging pairs of objects for pleasantness/unpleasantness, that object pictures had the fastest response times, with concrete nouns second fastest, and abstract nouns the slowest.

The dual-coding view of emotional representation has since been challenged by research in abstract representation and language, supporting Barsalou's claim that introspective states are important for abstract concepts. Kousta *et al.* (2011) contradicts the predicted processing times posited by DCT, finding instead that, when context availability and imageability are controlled for, there exists a residual processing speed advantage for abstract words over concrete. Kousta *et al.* (2011) suggests that emotional valence accounts for this advantage as abstract words were rated as more emotionally valenced than concrete words, and that this difference predicted the difference in processing times. Contrastingly, DCT predicts that because emotional states are only connected directly to other nonverbal representations, which are in turn connected referentially to language (directly to concrete words, and indirectly through association to abstract ones), that concrete words should be more emotionally valenced than abstract, and therefore maintain a processing advantage (Vigliocco *et al.* 2013, 289). In order to explain these findings, Kousta *et al.* (2011) propose that concrete and abstract concepts differ in terms of whether sensory, motor or affective information carry the greatest weight, with sensory-motor information more important for concrete concepts and affective information more important for abstract concepts (p. 14). The authors suggest therefore that concrete and abstract concepts function to bind different types of information together, with concrete concepts binding verbal

and sensorimotor information and abstract concepts binding verbal and affective information (Vigliocco *et al.* 2009).

Paivio (2013) argues that Kousta *et al.*'s (2011) findings do not contradict dual-coding assumptions and predictions, claiming that, in fact, the authors misinterpreted DCT (Paivio 2013, 282). Paivio suggests that, because the tasks used by Kousta *et al.* (2011) were verbal in nature, they predominantly activated verbal pathways and did not involve imagery, thereby negating potential effects of imageability and concreteness (Paivio 2013, 284). Similarly, Paivio argues that the findings from Kousta *et al.* (2011) only apply to the word/non-word lexical task used in each experiment and cannot be generalized to processing beyond that. Paivio ultimately views Kousta *et al.*'s (2011) embodied proposal for abstract words to be compatible with DCT, noting that their experiential information is equivalent with his notions of 'nonverbal,' as both modal and affective experiential representations are encoded in the nonverbal system.

I propose that DCT can capture emotional representation, but that more work is needed to clarify what representations comprise affective states and how they are connected. While Paivio suggests that emotion words can gain generalized affective responses akin to referential meaning, he does not expand on this. Similarly, the DIPSS position regarding introspective states as perceptual symbols provides an avenue for accounting for previous findings, but the lack of specificity regarding the nature of introspective representation does not provide the needed clarity.

To summarize, DIPSS claims that the DCT account of abstract concepts should rely less on verbal associative priming, and instead incorporate introspection as a primary nonverbal associate of abstract language, a finding supported by more recent research into concrete versus abstract language response times. Barsalou claims that introspection should therefore take the

level of a separate nonverbal code—a perceptual symbol or imagen—but provides no means for discriminating between different types of introspective states (e.g., affect, metacognition).

Paivio, on the other hand, attempts to account for introspection through emotional responses, viewed as a by-product of nonverbal processing. From this perspective, introspective states in the form of an affective response are a felt response to the activation of a network of representations in a specific context. Rather than choosing between these two accounts of introspection, I propose that both are viable and can be systematically explained in the current project, with appropriate adaptation to adequately account for introspection.

Later in this chapter I will account for three distinct aspects of introspection—imagery, metacognition, and internal sensation in the form of interoception. For imagery I will discuss subjective sensations of imagery vividness and control over imagery, as one aspect of metacognition. As a second measure of metacognition, I will discuss network priming and the associated ‘feeling of knowing’ that can occur when networks are primed but no specific item selection occurs.<sup>31</sup> Lastly, I will account for internal sensation and affect by integrating modern research into interoception as a primary feature of emotional construction. Interoception will take the place of a separate imagen/perceptual symbol code, allowing for specificity of representation of internally available sensations within the body. From this perspective, concreteness effects for abstract concepts can occur through multiple introspective channels, both as a general sensation arising from information priming within a large network of both verbal and nonverbal

---

<sup>31</sup> As is the case with the so-called ‘tip of the tongue’ phenomenon (Brown and McNeil, 1966; Schwartz and Metcalfe, 2011). Here, a word undergoes retrieval failure meaning that the specific target word cannot be recalled, however due to priming and partial activation of representations, information *about* the target is still available (e.g., auditory features of the word, such as it begins with a “T,” as well as availability of similarly structured target words, such as ‘train’).

representations (context dependence), and from specific attachments of internal bodily sensations ('responses' as interoception) to a particular set of activations within verbal/nonverbal networks.

The structure and content of long-term memory

DCT and DIPSS differ slightly in terms of their approaches to the structure and content of long-term memory, specifically with regards to the modularity of long-term memory (modular vs. unitary) and schematicity or abstraction of the trace itself. These issues are important to address when making claims about memory expertise, as it is necessary to hypothesize with precision about the changes to long-term memory that may occur with increasing expertise. I will begin by discussing approaches to long-term memory structure, followed by differences between DCT and DIPSS in terms of schematicity and abstraction of the long-term memory trace. I will conclude this section by contextualizing a combined DCT-DIPSS approach to the long-term memory trace with reference to existing theories of concept representation (i.e., exemplar and prototype theories).

*Long-term memory structure.* I will first address the issue of the structure of memory. Traditionally, approaches to long-term memory have been from two vantage points: firstly, a unitary position that views memory as a single system, where the focus is more on functional differences in memory use (a ‘processing’ view), and secondly, a multiple systems view of memory which posits the existence of modular, distinct systems, each with their own distinct operations (a ‘systems’ view, see Foster and Jelicic 1999). These distinct systems include procedural memory (for skills, typically automatic and unconscious), perceptual memory (memory for objects, used in identification), short term memory (reflecting availability of recently encountered cognitive inputs), semantic memory (general knowledge about the world) and episodic memory (conscious recollection of past, experienced events) (Tulving 1991, 12). While these systems are understood to partially overlap, the multiple systems view posits that each system is neurally distinct and operates independently during tasks for which it is specialized. Particularly relevant here is the distinction between perceptual, semantic, and episodic memory, as conceptual representation is traditionally believed to reside primarily in semantic memory, while remembered life events, typified by accompanying sensations of conscious awareness, reside primarily in episodic memory (Tulving 1972; 1993). Both systems (semantic and episodic) are assumed to be separate from mode-specific perceptual memory; in other words, they are divorced from perceptual representation.

Both DIPSS and DCT reject the distinction between perceptual memory and other memory systems, as they both argue for congruence between perception and cognition, knowledge and memory. However, they do differ slightly in their approaches to modularity vs. unitary organization. DIPSS posits a unitary view of long-term memory which blurs the traditional distinctions between multiple modular memory types—episodic vs. semantic,



procedural vs. declarative (Barsalou 2003b). Contrastingly, DCT explicitly embraces some modularity—particularly in the distinction between systems that are specialized for verbal and nonverbal processing respectively—where other forms of memory representations (semantic, episodic, procedural) are represented by connections within and across the verbal/nonverbal divide (see Paivio 2007, 33). In this way, an integrated (DCT + DIPSS) view of long-term memory organization is one that is primarily unitary, where modularity is represented by selective access to existing traces: the distinction between verbal and nonverbal systems is functionally modular, however given the same modalities cut across each system, they may be more functionally than structurally distinct.<sup>32</sup> This means that distinctions between traditional memory modules (i.e., episodic, semantic) stem more from selective access to particular traces (i.e., more of a ‘processing’ view) rather than the separation of verbal/nonverbal representations further into long-term memory subtypes.

For example, the episodic/semantic distinction is maintained in DCT, with the caveat that the distinction is not a perfect, fully modular one. Semantic ‘knowledge’ in DCT occurs when original episodes are blurred together over many instances, and/or the episodic knowledge is forgotten over time (Sadoski and Paivio 2014, 68). Similarly, episodic-like memory may still be available from ‘semantic’ traces: semantic access to the concept ‘cup’ may selectively rely on verbal representations (e.g., types of cups, such as teacup, Styrofoam cup, paper cup, etc.), but this verbal access will also prime some aspects of episodic information available in the nonverbal system through referential processing, such as the association of different types of cups in different contexts (e.g., paper cups evoking associations with picnics, teacups with indoor lunch

---

<sup>32</sup> For example, verbal and nonverbal representations in the auditory mode (e.g., heard words and music) demonstrate partial neural overlap in several brain areas, suggesting that such networks may be shared between the two. It is unclear that this overlap entails sharing of neural activation however, so it is currently unknown if the systems are more unitary or more modular in structure at the neural level (see Peretz *et al.*, 2015).

or dinner, see Sadoski and Paivio 2014, 70). Specific episodic memories may also be available within the network of ‘cup knowledge,’ such as the memory of having received a particular cup as a gift. Different types of knowledge, such as procedural knowledge, are reflected in selective access of motor imagens for actions involving cups (e.g., picking up, drinking from, washing, etc.). This blurring of perceptual and conceptual information in a unitary view of memory is central to DIPSS, which imbues selective access to conceptual knowledge with traces of perceptual grounding, something Barsalou refers to as a kind of ‘being there conceptually’ (Barsalou 2002).

To summarize, the approach to long-term memory organization in the current project is one that is more unitary, where modularity is maintained via selective access to parts of a distributed set of memory traces. This approach is similar to some modern theories of integrated long-term memory (semantic/episodic integration, see Humphreys et al. 2020), and theories of semantic memory where traces are distributed across modality and verbal/nonverbal systems (e.g., Forde and Humphreys 1999; Warrington and McCarthy 1987). It is important to note that the modular distinction between long-term memory and working memory is still maintained in the current framework, as it is vital in being able to distinguish changes made to existing memory traces (long-term memory) and changes in the ability to actively *use* parts of long-term memory during online cognition (i.e., selective maintenance in working memory).<sup>33</sup> This will be discussed further in the section below on dynamic interpretation in situated simulation.

---

<sup>33</sup> This reflects the distinction between logogens/imagens and the ‘response’ (selection and output) in DCT, and the more nuanced distinction between simulators (existing networks of knowledge) and simulation (online use of parts of those networks) in DIPSS. It is also important to note is that the distinction here is primarily between long-term memory and working memory, where the traditional distinction of short-term memory as a separate module is merely viewed as a portion of long-term memory that fades over time (see Anderson 1999, 166).

*Long-term memory content.* The second issue to address is the nature of the long-term memory trace itself. Paivio criticizes Barsalou's DIPSS for the invocation of abstraction at the level of representation, particularly the schematic nature of the perceptual symbol and the frames that organize them into simulators, noting that this invocation simply increases the explanatory burden and reduces the explanatory power of the theory (Paivio 2007, 118). In DIPSS, representation is slightly more abstract compared to DCT; properties of external objects are not stored in an analogous format in memory as they are with imagens and logogens, but are instead stored as an abstract schematic reduction in the form of a simulator which, which through the use of a frame, can be used to reinstate (or construct) a simulation of an object or event.<sup>34</sup> Figure 2.4 shows the DIPSS approach to simulators, which are stored as a partial pattern of activation of neurons in conjunctive association areas. During simulation, conjunctive neurons (perceptual symbols organized into simulators) reactivate lower-level neurons in sensory feature maps to recreate the perceptual state of the experienced object. This is a kind of reconstruction of memory that will differ from the original instance of perception (i.e., it will be modified, imperfect, altered, etc., due to reconstruction error, interference, etc.). Part of the difference between DIPSS and DCT stems from Barsalou's attempt to contextualize DIPSS within a neural framework that revolves around action-planning and information integration in conjunctive areas, while DCT is more confined to smaller effects of memory.<sup>35</sup> The two theories are slightly opposed regarding the nature of abstraction at the level of the memory trace, particularly as Paivio claims that what constitutes a frame or simulator from a representational perspective (i.e., what exactly 'makes up' a frame) is somewhat ambiguous.<sup>36</sup>

---

<sup>34</sup> The distinction here is subtle. A slightly impoverished explanation of the differences between DCT and DIPSS representation is that in DCT the visual imagen for an apple is in part directly represented in the visual system (sensory properties encoded) whereas in DIPSS they are not directly represented; what is encoded instead is an

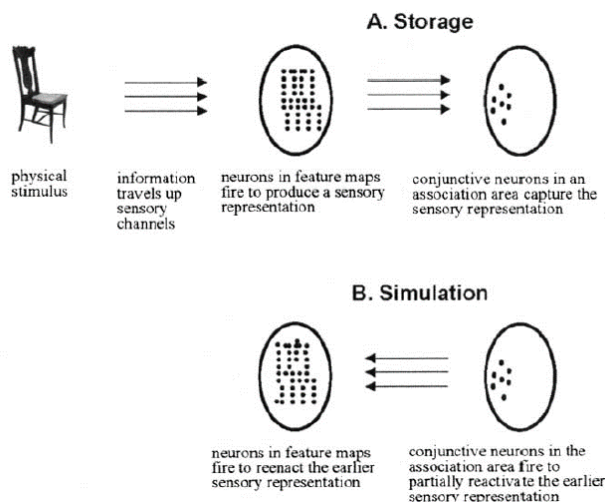


Figure 2.4. Approach to Storage and Retrieval (Simulation) from Barsalou (1999)

In this section, I will argue that these two theories are more concordant regarding abstraction than Paivio notes in his critique of DIPSS. In order to clarify this, I will situate DCT and DIPSS within traditional approaches to concept representation that also differ fundamentally in their approach to abstraction: exemplar and prototype theory. This will help to demonstrate that both are in fact comparable in their approaches to the long-term memory trace, particularly within an expert memory framework that posits expertise stems from *both* changes to long-term memory traces (e.g., revision to existing traces, addition to information through additional traces, distributed across modality and verbal/nonverbal systems) and to *controlled access* to these traces (e.g., skill in information retrieval, long-term working memory or *situated*

---

abstract ‘instruction’ on how to reconstruct an apple in memory for use in either perception or cognition (e.g., in imagery).

<sup>35</sup> In DCT, representational units are, like DIPSS, contextualized within a neural framework. However, Paivio prefers to situate DCT around multimodal coding rather than around single or unitary coding neural frameworks (see Paivio 2007, chapters 7 and 9 on the multimodal brain and a common coding approach, respectively).

<sup>36</sup> The content of the frame according to Barsalou are predicates, attribute-value bindings, constraints, and recursion (Barsalou 1999, 591). Frames are essentially akin to schemas, which traditionally are discussed as abstract entities stored amodally as propositions. The DIPSS position suggests that a frame is a kind of unimodal entity (abstract instructions) on how to unpack a representation in the specified modality. Frames are used for both sensory and language representations in DIPSS, so they are not mode-specific language representations either. It is important to note that the frame concept in DIPSS was not expanded or specified in later versions of the theory, so it appears to have primarily been an attempt to reconcile DIPSS with previous computational approaches to representation.

*conceptualization*). I will start by briefly outlining the differences between exemplar and prototype theories of representation, after which I will situate DIPSS and DCT within these frameworks to clarify the current projects' position on the nature of the long-term memory trace.

Traditionally, there are two contrasting approaches to concept representation: prototype and exemplar. Both are amodal, meaning that they assume conceptual knowledge is stored in an abstracted non-perceptual format. However, they differ with regards to the nature of the memory trace. Prototype theory was developed to account for findings that demonstrated a graded structure of category membership within taxonomies, rather than an 'all or nothing' membership typified by the classical view. For example, Robin is a better fit than Penguin for the category BIRD, which results in faster classification speeds (see Smith and Medin 1981, 69).<sup>37</sup> The traditional prototype view assumes that conceptual knowledge is stored in a summary representation that *only* contains the central tendency information for various features of category members (e.g., see Rosch 1973; Rosch and Mervis 1975; Hampton 1995a, b). Each time a new categorization decision is made, category inclusion is determined by similarity of the instance ('exemplar') to the features stored in the stored in the prototype representation. Every time this occurs, the prototype representation undergoes a form of 'update' or *information revision* to reflect the newly encountered instance (Barsalou 1990, 66). In this way, the prototype approach to concept representation features *no information duplication* as the representation is centralized (i.e., stored in a single 'location') and is decontextualized (i.e., abstracted away from its context) as only the probabilistic information is stored in the prototype (Barsalou 1990, 67). This also reflects a kind of *information loss*, whereby idiosyncratic or non-probabilistic

---

<sup>37</sup> Classical view maintains that categories have necessary and sufficient conditions for defining a concept, whereas the prototype or 'probabilistic' view suggests that concepts are represented only by the properties that are characteristic or probable (allowing for membership to be graded rather than all-or-nothing, see Medin and Smith 1984).

information about an instance is lost when the information is stored in memory.<sup>38</sup> While prototype models are well-suited to explain and model fuzzy category membership, they struggle to account for category learning, as it is unclear how and by what processes information extracted from instances are stored into a prototype (Ross and Makin 1999, 212).

Exemplar models, to which I now turn, have generally been found to better account for learning.<sup>39</sup> The exemplar approach assumes that every instance of a category encountered is stored in memory, and that categorization of new instances is conducted by comparing the similarity of the new instance to all previous instances in memory. Traditional exemplar models feature high a level of information duplication (as each instance is stored in a separate trace), and little to no information loss or revision (each instance is stored permanently, and new traces are formed in lieu of revising existing ones, e.g., Medin and Shaffer 1978; Medin and Florian 1992). While these early exemplar models were able to account for much of the data that prototype models accounted for, they struggled to explain how people were able to operate using abstractions (Ross and Makin 1999, 215). Later exemplar models accounted for prototype effects (e.g., typicality, category size, differential forgetting of prototypes) by assuming that prototypes can be essentially ‘constructed’ at retrieval: that is, by accessing a large pool of exemplars simultaneously, the ‘summed average’ (i.e., the ‘prototype’) of relevant features can be accessed through parallel activation (see the MINERVA model, Hintzman and Ludman 1980; Hintzman 1986). Exemplar models have also typically performed far better than prototype models at category learning, particularly as these models incorporate effects of selective attention and

---

<sup>38</sup> Barsalou (1990) notes that some proposed prototype models can store idiosyncratic information in the centralized summary representation in a similar manner to a connectionist neural network (p. 68).

<sup>39</sup> One such prototype model that has claimed to account for category learning was proposed by Minda and Smith (2001, 2002; also see Smith and Minda 2000), however exemplar models that account for attentional allocation have been shown to provide a better account of the categorization data (Nosofsky 2000; Nosofsky and Johansen 2000; Nosofsky and Zaki 2002; Zaki et al. 2003; Zaki and Nosofsky 2004).

learning rules typically used in connectionist neural networks (see Nosofsky 1986; Nosofsky, Kruschke and McKinley 1992; ALCOVE model in Kruschke 1992; 1993). These models have been particularly useful in accounting for learning data in language acquisition (e.g., phonetics, phonology, semantics, syntax, etc., see Gahl and Yu 2006; Abbot-Smith and Tomasello 2006), providing support for experience-based over innate universal grammar-based approaches (e.g., Chomsky 1957; 1995). While exemplar models have provided good accounts of concept representation and learning, other models have combined aspects of exemplar and prototype theory to account for partial abstraction of memory traces (e.g., conceptual model of Abbot-Smith and Tomasello 2006; varying abstraction models, see Vanpaemel and Storms 2008; Divjak and Arppe 2013). More recent category learning literature has moved away from discussions about the nature of the representations stored in long-term memory and have instead focused more on the learning systems and processes in online category learning.<sup>40</sup>

Dual-coding theory explicitly embraces an exemplar approach to concept representation over a prototype view, maintaining that knowledge is not abstract but is derived from concrete past experiences and retains some perceptual information from such experiences (Sadoski and Paivio 2014, 68). The DCT view is also proximate to exemplar models that posit parallel multiple traces (i.e., exemplar-cloud approach, Hintzman 1986). For example, when presented with a nonverbal stimulus, the activation of imagens occurs through the sensory system through similarity, ‘homing in’ on the group of representations that is most similar to the stimulus presented. DCT also accounts for systematic differences in long-term memory activation through

---

<sup>40</sup> Much of this work has been inspired by the attentional allocation findings in exemplar research (e.g., Nosofsky 1985, Kruschke 1992, see Bartlema, Lee and Vanpaemel 2014). Newer theories of category learning account for such attentional differences by examining different types of learning tasks, such as rule-based (explicit, declarative) versus information-integration (implicit, procedural) tasks, and propose dual-system models to account for systematic differences between category learning types (Ashby and Maddox 2005; Minda and Miles 2010; Chandrasekaran, Koslov and Maddox 2014; Roark and Holt 2015; Ashby and Valentin 2017).

its approach to processing differences; *indirect* activation (e.g., via imagery or referentially through the verbal system) will activate a larger pool (or ‘cloud’) of imagens than *direct* activation (e.g., representational activation through nonverbal system, presentation of a picture or object in lieu of linguistic presentation). This is because referential connections from a logogen will have more connections to a larger pool of imagens than will a picture of an object, which will ‘home in’ on a smaller pool of imagens through direct similarity (Paivio 2007, 49). DCT is also in concurrence with exemplar approaches that posit abstraction at retrieval by allowing for selective access to the verbal and nonverbal systems.<sup>41</sup> The ‘pure’ exemplar view in DCT is made somewhat problematic however given the way in which the theory handles abstraction at the level of the imagen/logogen, particularly with the approach to information revision characteristic of prototype theories (Barsalou 1990, 67). For example, DCT notes that a ‘good exemplar’ can come to ‘stand in’ for a category (Paivio 2007, 119), noting that such a ‘good exemplar’ can functionally serve as the central tendency in a multidimensional distribution, much like a prototype (Sadoski and Paivio 2014, 33). Such exemplars become frequently used, and therefore frequently undergo information revision as the contents and probabilistic connections are modified by experience (e.g., through elaboration, distortion, interpolation, etc., *ibid.*). Such a viewpoint implicitly embraces features of abstraction typified by prototype theories. A frequently accessed exemplar can undergo extensive information revision, essentially transforming and exemplar representation into a category’s central tendency, ultimately making exemplar and prototype theories empirically indistinguishable (Barsalou 1990, 66).

Barsalou’s DIPSS is explicitly offered as an embodied alternative to exemplar, prototype, and connectionist models of concept representation (Barsalou 2003b). The approach adapts the

---

<sup>41</sup> For example, selective access to the verbal system is understood to be particularly important for abstraction at retrieval (Sadoski and Paivio 2014, 69).



most advantageous aspects of each of these theories—abstraction in the form of information revision from prototype theory, retention of idiosyncratic information in distributed representation from exemplar theory, and the dynamic nature of connectionist nets<sup>42</sup>—and integrates it within a theory that posits nonmodular and distributed modal organization. Concepts are represented by large networks of interconnected perceptual symbols, which are organized around actions or *simulations* rather than predetermined taxonomies (Barsalou 2003b, 522). From this perspective, a concept is *situated* because what part of the concept-network that will be activated *depends* on a particular context and/or set of actions: simulation, or selective access of representations, will vary based on context and goals of the person doing the categorizing. This view is more accordant with exemplar theory which posits selective access to exemplar clouds as opposed to prototype theory which posits that the same representations are accessed during *each* act of categorization. However, because the contents of pre-existing traces can be modified or revised, the approach embraces aspects of abstraction inherent in prototype theories. The act of simulation—using a part of a simulator—is essentially recursive as simulation is constructed using part of a concept network, the existing traces in that network can be modified, and aspects of the simulation can be added (in the form of additional traces, or elaboration) to the simulator, which provides more detail to the exemplar cloud. For example, Barsalou (2003a) notes that upon initial interaction with a car, a schematic (i.e., feature or property-limited) simulator for *wheel* may develop (i.e., information is filtered out). After continued interaction with this category, more detail is extracted from subregions of category, facilitating the acquisition of related simulators, such as *tire* and *hubcap* (Barsalou 2003a, 591). Essentially, over the course of

---

<sup>42</sup> Connectionist networks are generally more dynamic and distributed than traditional exemplar theories, which again, posit limited information revision to existing traces. In a connectionist network, a concept is represented by a large space (or network) of interconnected nodes (Barsalou 2003b, 520-521).

concept acquisition, existing traces may be modified (addition of more detail, enhancement of features that are attentionally relevant for categorization) and new traces will be added.

In summary, the DCT and DIPSS approaches to the structure of contents in long-term memory are less different than the authors initially suggest. The current approach, stemming from DIPSS, is most proximate to an exemplar model that embraces aspects of abstraction in the form of information revision to existing traces, and in selective access to parts of a concept network to create abstractions at retrieval (e.g., Hintzman and Ludam 1980; Hintzman 1986). The primary differences between the current approach and other hybrid exemplar-prototype approaches (e.g., Abbot-Smith and Tomasello 2006), is that all long-term memory traces are embodied and mode specific, meaning that information revision occurs in an existing trace *in a specific mode*, and the addition of traces are also modality specific. There exist no amodal representations, a position held by all previous types of concept representation theories discussed. As will be shown, ‘abstraction’ will be primarily a function of selective access to traces in simulation (i.e., abstraction at retrieval like in exemplar theories). As expertise develops, simulation will also become more nuanced and selective because the contents of long-term memory become more elaborated. Because these traces can also be revised, more detail can be added to existing traces, also permitting simulation to be selectively more detailed, or more sparse or abstract. In the next section, I will discuss in more detail the approach to situated simulation, which involves selective access of information from long term memory held in working memory.

#### Situated conceptualization and working memory

The last topic to address is the approach to situated simulation, which maintains the modular distinction between long-term memory and working memory (hereafter LTM and WM).

Both DCT and DIPSS embrace this distinction; however, neither theory fully discusses the nature of the interaction between long-term memory use and its effect on information held in working memory. DCT specifies that representations (imagens, logogens) selected from long-term memory will be used in a system-specific “response” (either verbal or nonverbal, see Figure 2.5), but does not specify beyond this. DCT also posits that the response will vary as a function of imagen/logogen selection from LTM, which is reflected in the systematic activational pathway types (representational, associational, referential) demonstrated through empirical differences in memory effects (e.g., concrete peg hypothesis discussed above). Therefore, there is a need to specify both the LTM and WM structure, and their interactions. In order to better detail the connectivity present in LTM, I will modify the types of connections that can exist between representations, including highly probable and lower probable connections, as well as bi- and uni-directional connections that can differ in probability based on their direction (see Figure 2.6).

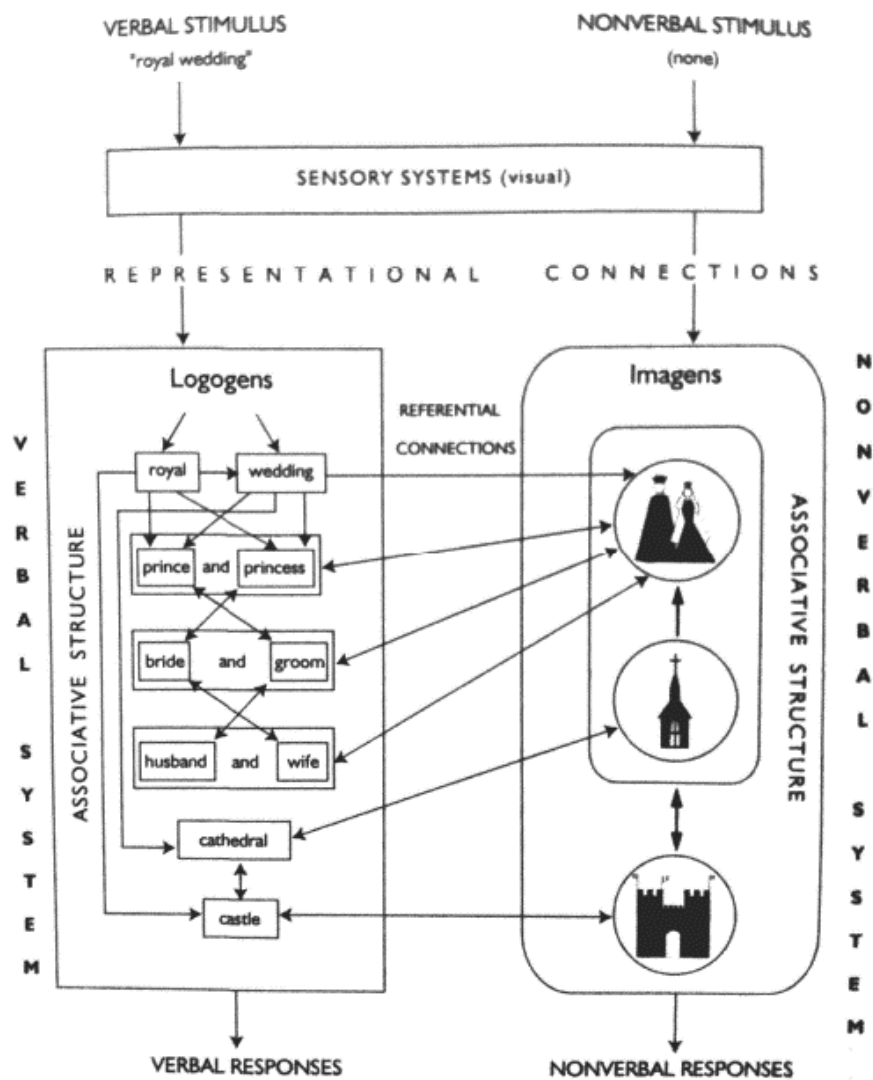


Figure 2.5. Example of Mental Model for Phrase 'Royal Wedding' (Sadoski and Paivio 2013, 58)

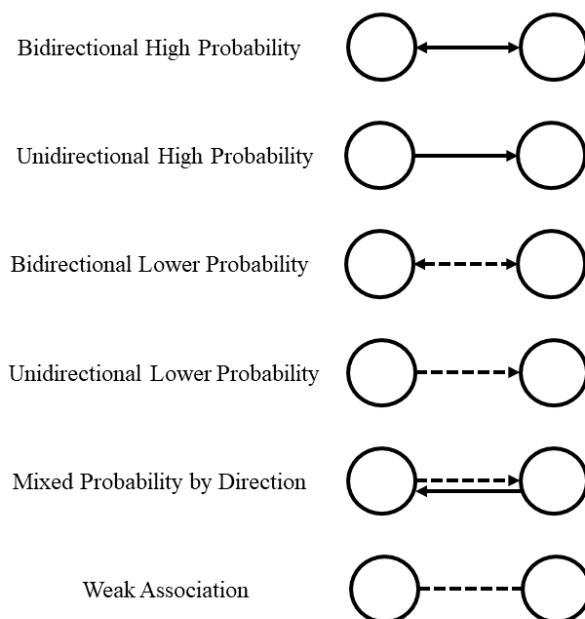


Figure 2.6. New Connector Types for LTM

DIPSS similarly posits selective access to pools of memory traces, noting that simulation will differ based on context-dependent variables, but does not systematically outline how selective access to LTM may differ in WM across categorization instances. In this section, I will briefly outline my approach to situated simulation as selective maintenance of representations from LTM held in WM. I start by reviewing Baddeley’s multicomponent working memory model (Baddeley and Hitch 1974; Baddeley 2003; Baddeley 2007).

The concept of working memory stemmed from research into the insufficiencies of unitary short-term memory (STM) theory to account for impairments to cognitive functioning in online processing (Baddeley 2007, 6). The term *working memory* refers to a limited-capacity temporary storage system that is used during human cognition. The original model, proposed by Baddeley and Hitch (1974), is a multicomponent model that contains three separate systems: an attentional control system called the central executive responsible for guiding information

selection, and two mode specific subsystems for temporarily maintaining the selected information, the visuospatial sketchpad (for visual, spatial and some kinesthetic information maintenance) and the phonological loop (for subvocal articulatory maintenance of auditory information, originally developed to account for verbal rehearsal) (see Baddeley 2003). Information maintained in these subsystems decays rapidly and requires active rehearsal in order for the information to be maintained and transferred to long-term memory.<sup>43</sup> These different subsystems allow for an explanation of information interference effects on memory, such as the effect of articulatory suppression (concurrent verbalization to disrupt the phonological loop) on disrupting serial memory (Baddeley 2000, 419). This position is accordant with DCT's verbal/nonverbal split that hypothesizes that the different systems are optimized to handle sequential and synchronous information respectively, and that information maintained in the same mode (e.g., auditory) can lead to interference effects rather than benefits to memory performance (Paivio 2007, 49). However, the original multicomponent model of working memory had some limitations, particularly when accounting for effects of nonverbal auditory information, which appears to not always be held in phonological loop, and in accounting for how mode specific information can be integrated together in working memory and become available to conscious awareness (Baddeley 2000, 421).<sup>44</sup> To account for this, the episodic buffer was introduced to the multicomponent working memory model. The episodic buffer is also a capacity-limited maintenance system controlled by the central executive, which holds modally integrated information (e.g., complex images) stored in the form of conscious awareness (ibid.,

---

<sup>43</sup> The model holds a similar modular view of LTM that DCT does, separating out language and vision. However, the WM model also maintains a separate long-term store for episodic and semantic memory. For the current project, I will maintain that episodic, semantic and short-term memory are merely different traces stored within a unitary LTM that spans across mode (audition, vision, etc.).

<sup>44</sup> The finding regarding the divergence of information in the phonological loop is also supported by research into auditory and musical imagery, which shows that some musical imagery, such as imagery for timbre, does not rely on the phonological loop (see Smith, Reisberg and Wilson 106-107).

see Figure 2.7). Information can therefore be distributed across working-memory systems to optimize processing and reduce interference, much in the same way that selective maintenance across verbal and nonverbal systems in DCT is understood to benefit memory.

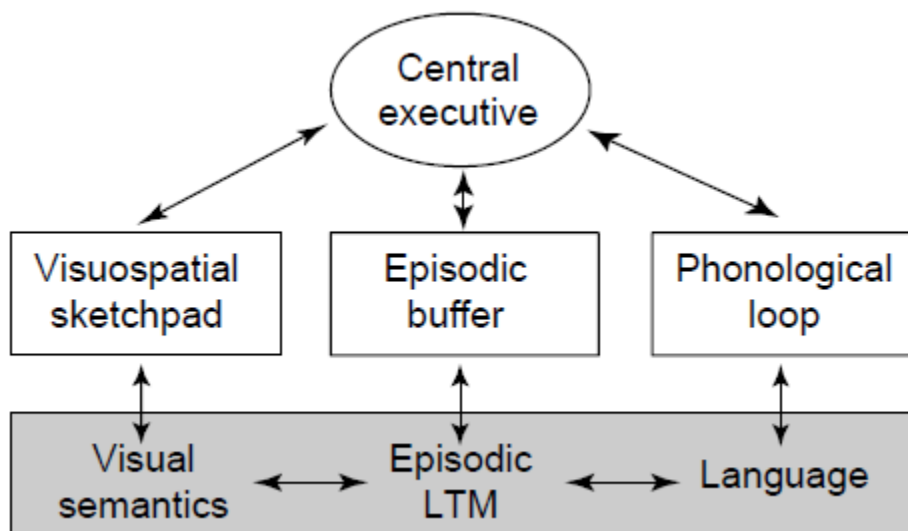


Figure 2.7. Baddeley's Working Memory Model (2000)

In the current project, LTM and WM will be shown using a modification of the traditional DCT outline, with the stimulus or input shown at the top, the verbal and nonverbal split in LTM in the middle and WM taking the place of the 'response' section at the bottom (see Figure 2.8). Rather than showing the three-part split of WM as typified by the Baddeley model, WM here will be shown as a unitary maintenance space where, as in DCT, the modality of information is specified. Information maintained in the auditory mode will be distributed, understood to be partially maintained in the phonological loop and partially in the episodic buffer. Because this project is mainly concerned with what information is currently being selected, used, and manipulated in WM, the WM space will only show information currently available to conscious awareness. The interaction between WM and LTM via the central executive will be shown via a color-coded scheme of traces in long-term memory. A green trace

has been selected for maintenance and is currently being used in WM (see Figure 2.9). Blue indicates an LTM trace which has been *primed* but has yet to be selected by the central executive. Such traces can impact implicit or unconscious information processing and decision making but are not actively maintained or manipulated in WM. Inactive traces in LTM will be shown in black, while traces that are being actively *suppressed* by the central executive will be shown in red. In terms of information processing, primed traces would demonstrate the easiest and most rapid access (e.g., fastest reaction time) for the central executive, followed by inactive traces in black, with suppressed traces in red being the most difficult and slowest for the central executive to reactivate and retrieve. These distinctions will allow for a nuanced and systematic understanding of memory expertise acquisition in simulation as skilled control over the central executive indicated by ease of access to information held in LTM as well as fluent retrieval and maintenance of that information in WM.

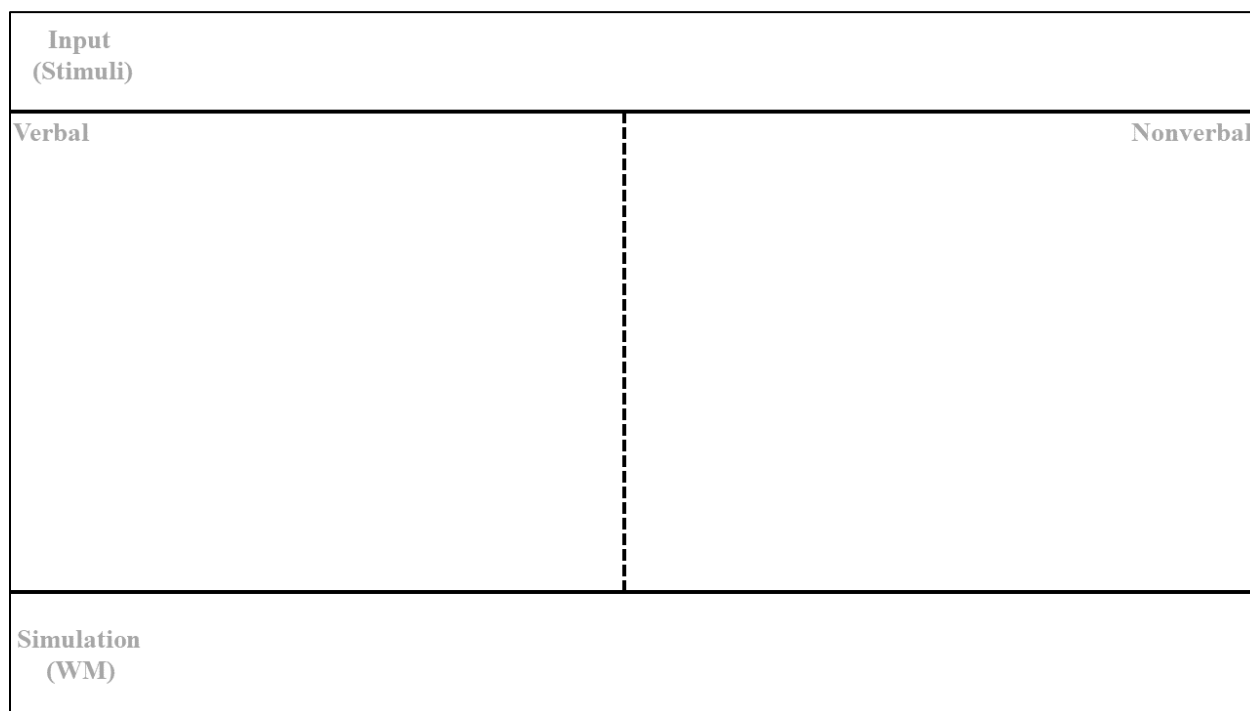


Figure 2.8. New DCT Layout Showing Input (top), LTM (middle) and WM (bottom)



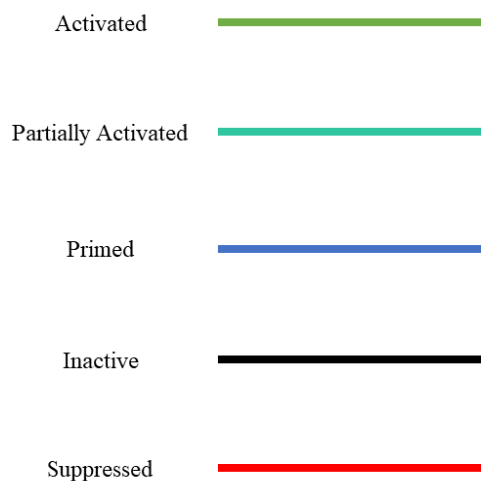


Figure 2.9. LTM Trace Colors Showing Activation Types

### **Re-Defining Categorization: Situated Conceptualization as Spreading Activation in DCT**

I will now provide a summary of the current approach to categorization skill (i.e., situated conceptualization) as patterns of spreading activation (representational, associational, referential) in DCT. I use ‘imagens’ and ‘logogens’ as terms for representations, and ‘simulator’ and ‘simulation’ to refer to concept-networks in LTM and use of those networks in WM, respectively. The current approach relies on *both* language and nonverbal systems, rather than DIPSS’ focus on the nonverbal system. This approach yields two beneficial outcomes. Firstly, that once simulators for words become linked to simulators for (nonverbal) concepts, they can control nonverbal simulations (p. 522). With more experience and expertise, these cross-system referential connections become more automatic. Secondly, language aids in the development and acquisition of categories by expanding the simulator network and explicitly connecting

simulators to other simulators, allowing for a taxonomic-like organization of concepts and enhanced connectivity between nonverbal representational.<sup>45</sup>

Situated conceptualization—simulation involved in categorization behaviors—is viewed as systematic patterns of spreading activation in DCT. Given the more unitary approach to LTM organization in the current project, the related behaviors—recognition, identification, and categorization—are assumed to be handled by the same memory systems, and merely reflect differences in activation of representations and their use, which can also include a time-course component that will be vital in the discussion of musical categorization. Previous research into categorization and recognition has found some dissociations between the two activities, particularly in brain damaged patients who demonstrate selective impairment on recognition, but no impairment to categorization (e.g., Knowlton and Squire 1993). This suggests that recognition and categorization may be handled by different memory systems operating in parallel. Unitary exemplar views however (Nosofsky 1988; Nosofsky and Zari 1998) have explained this dissociation by proposing different patterns of activation within a single set of representations, rather than separate representations for categories and objects. This position is also supported by findings suggesting a common target representation for recognition and categorization (Maxfield and Zelinsky 2012). Research also suggests a time-course component to differences between recognition and categorization. Visual object detection, for example, occurs earlier in visual processing than does visual basic-level categorization (see Mack and Palmeri 2010), with detection being easier for participants than categorization (Bowers and Jones 2008).

---

<sup>45</sup> This position is also held for the learning of abstract concepts, such as relational concepts. Language operates as a kind of *cognitive tool kit* that supports representation and reasoning that would be unavailable in a purely nonverbal system (Gentner 2016).

In this section, I will demonstrate that differences between recognition, identification and categorization are reflected in spreading activation within and across verbal/nonverbal systems (see Figure 2.10). Recognition, or object detection, entails the initial activation or priming of a set of representations belonging to a presented stimulus through direct representational activation (see Figure 2.11). Here the representations can either be primed (unselected by the central executive, below conscious awareness), or partially activated, indexed by a sensation of familiarity.

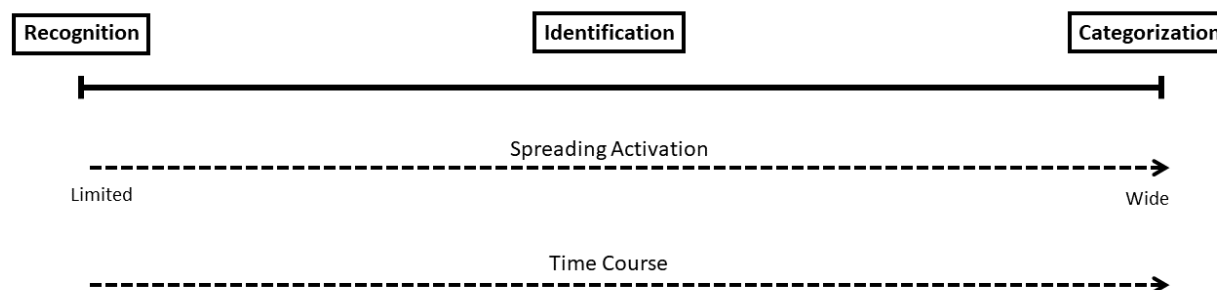


Figure 2.10. Recognition, Identification, and Categorization as Spreading Activation in DCT

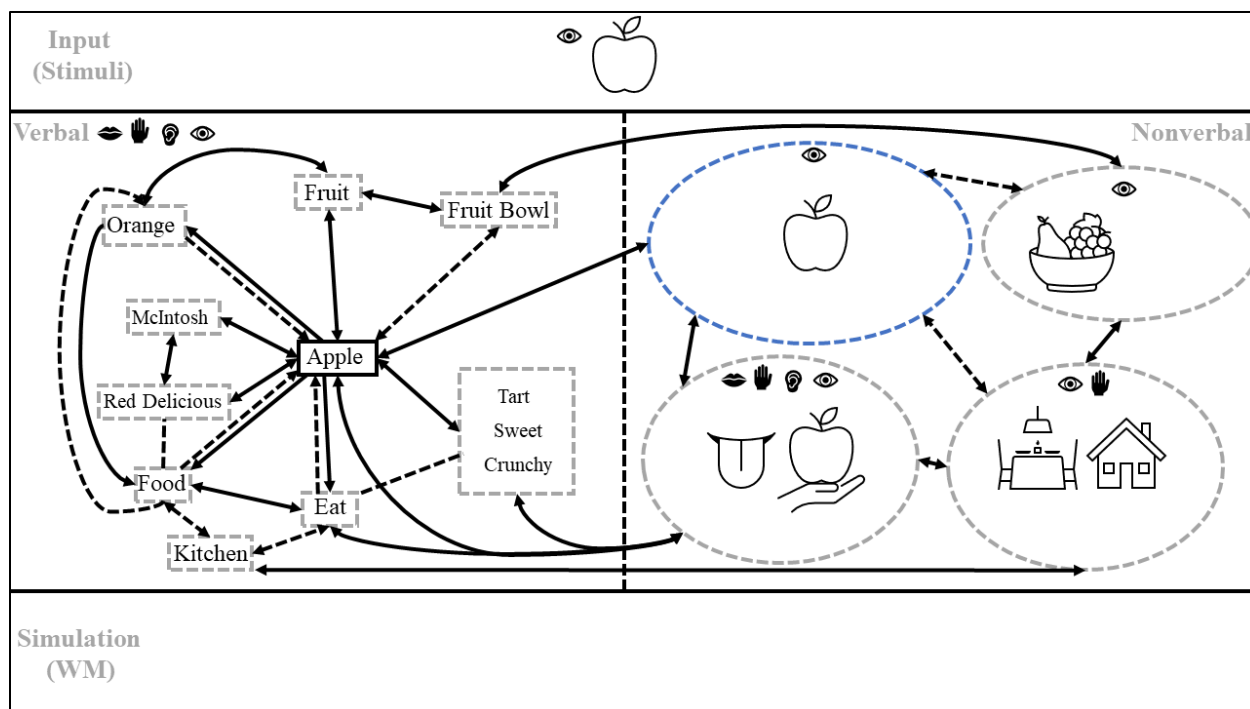


Figure 2.11. Initial Recognition of a Visual Stimulus through Representational Activation.

Identification involves activated traces and spreading activation, both associational within system to a larger pool of representations, and referential activation to the verbal system, indicating availability of a verbal label (Figure 2.12). Lastly, categorization involves more spreading activation outwards within and across systems, reflecting the availability of more information than is explicitly present in the stimuli in the form of primed or selected information within a simulator (or ‘concept network’, see Figure 2.13). Though the core identification representations may be the only traces actively selected, the priming of the other traces in the simulator network represents the availability of information related to the target stimuli—taxonomic words and associated nonverbal representations providing ‘context’ for the activated category. Such primed traces can be easily selected by the central executive to support a categorization behavior, such as naming the presented stimulus as a ‘fruit’ (superordinate level)

(see Figure 2.14a), or in imagining related contexts, such as sounds, tastes and actions involved in eating (Figure 2.14b).

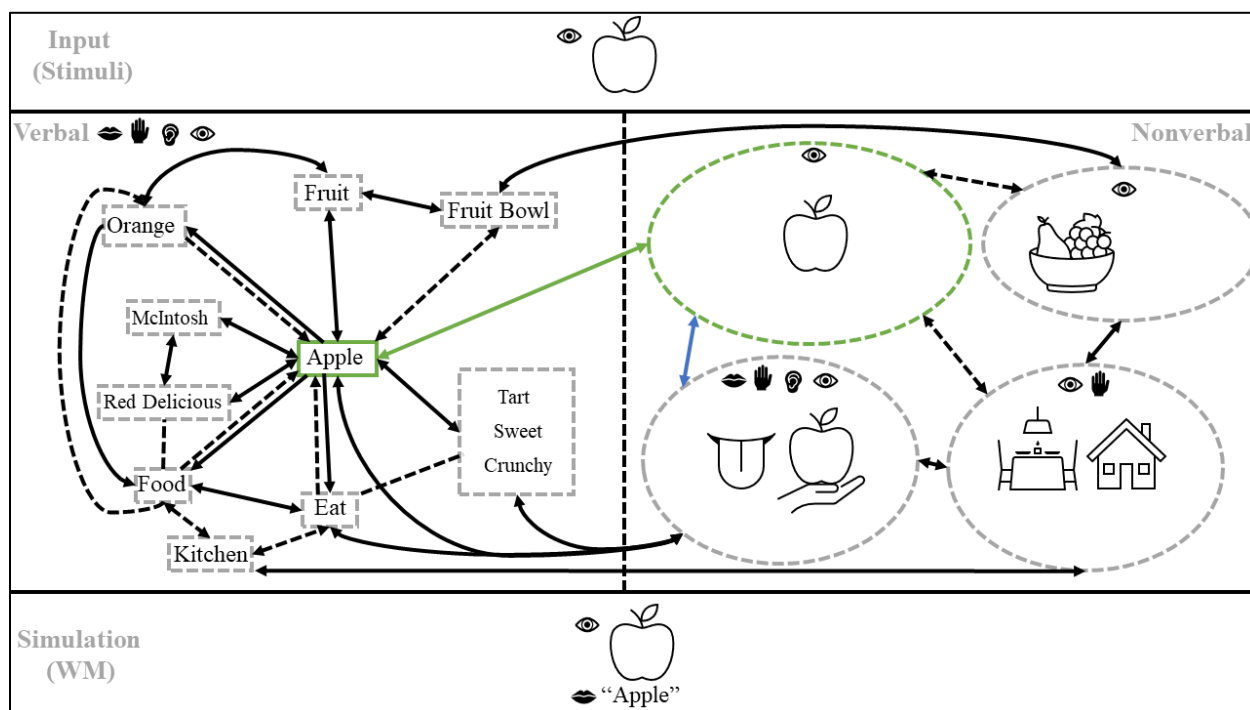


Figure 2.12. Identification of Apple through Referential Activation

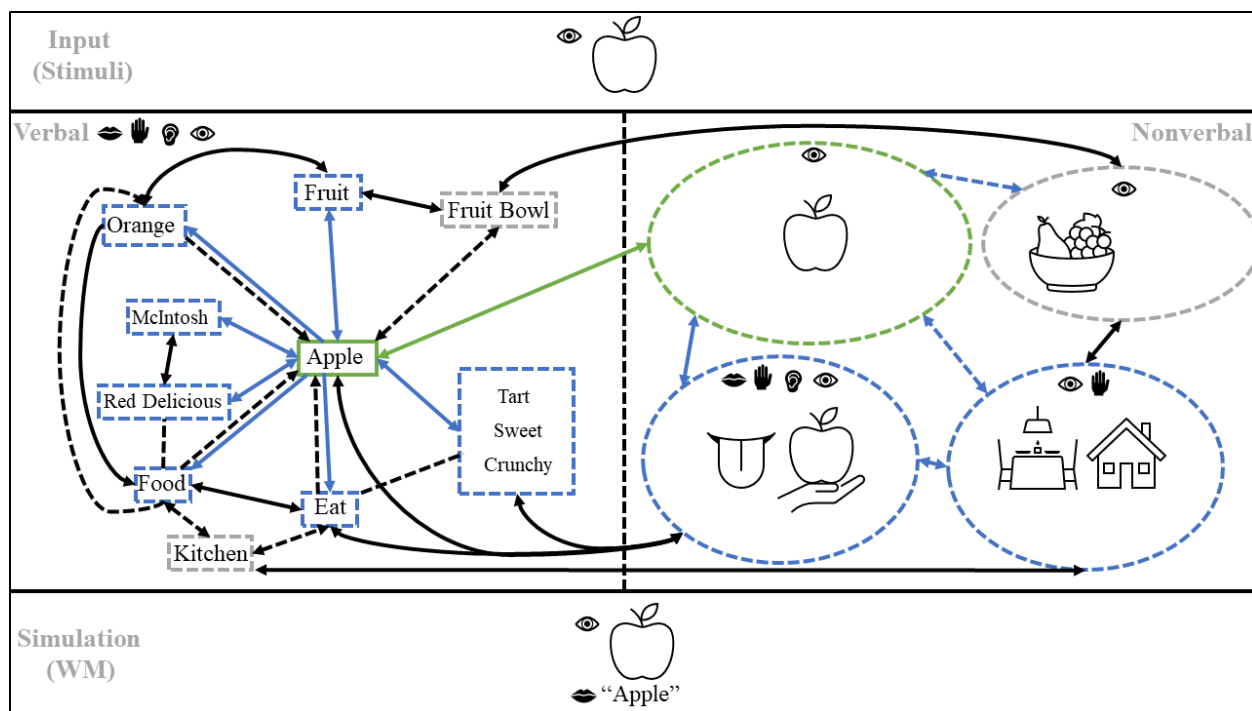
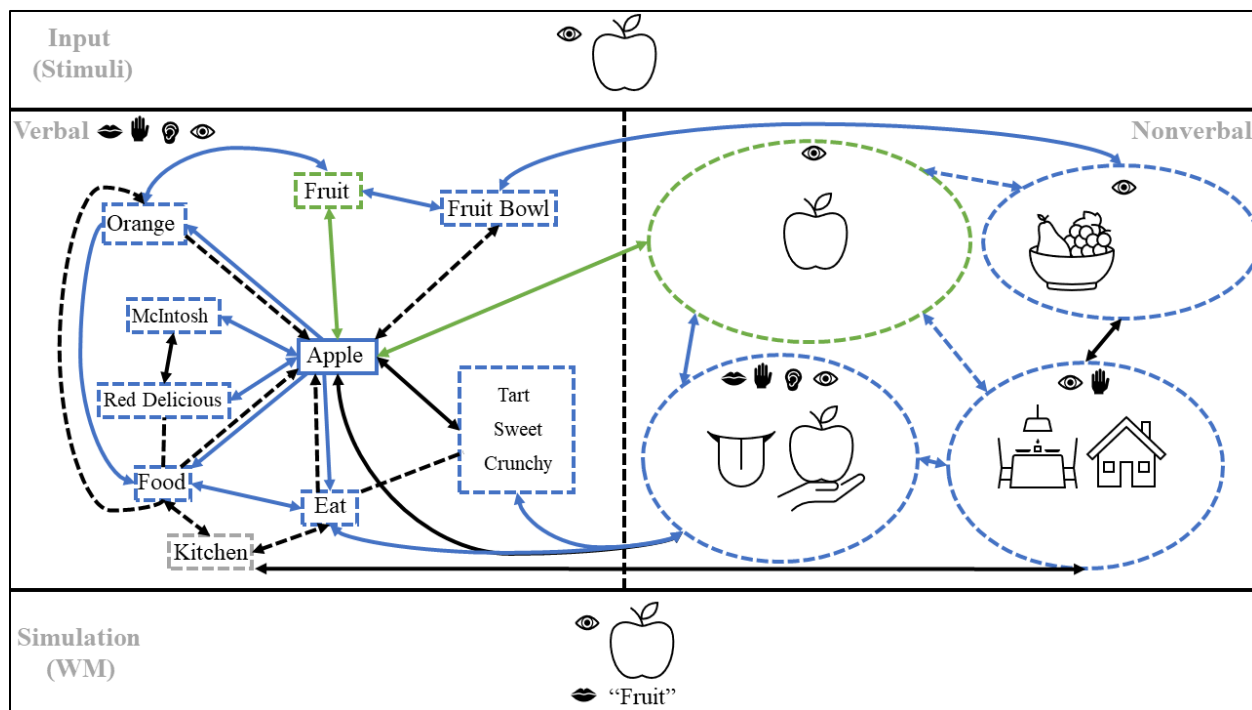


Figure 2.13. Availability of Categorization through Referential and Associational Activation

(a). Categorization Decision through Associational Activation of “Fruit” Logogen



## (b). Imagined Apple Taste through Associational Processing in the Nonverbal System

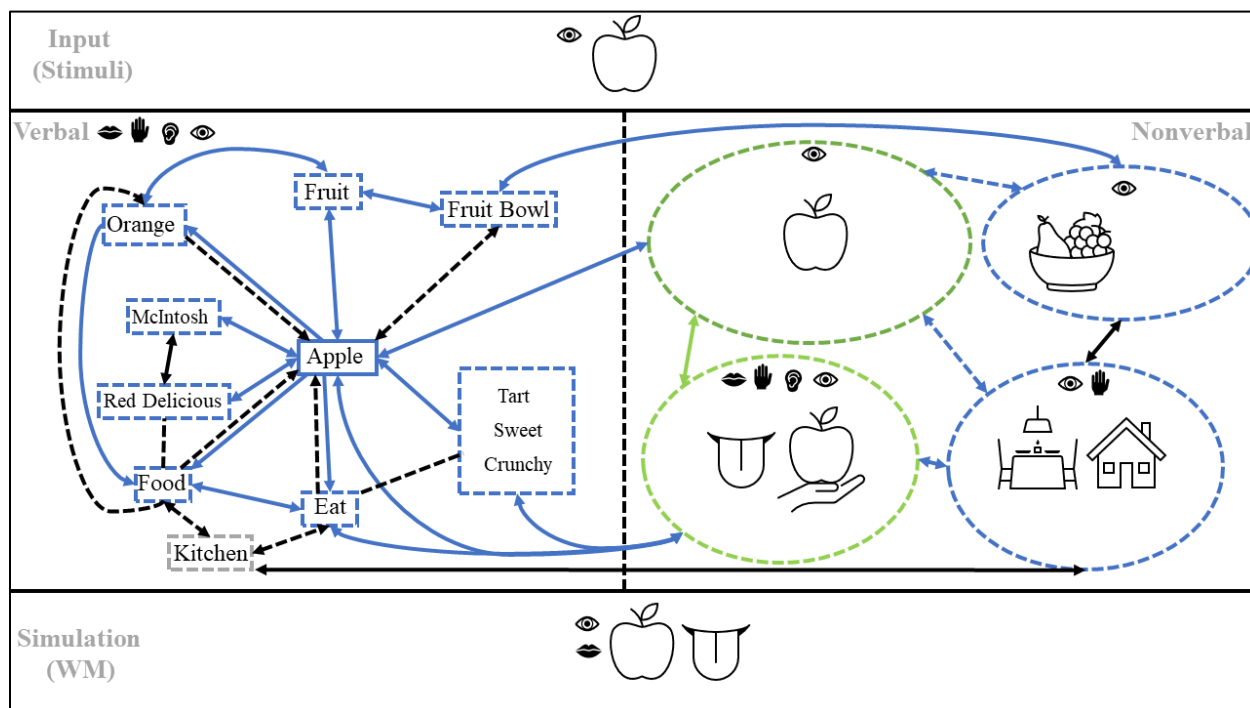
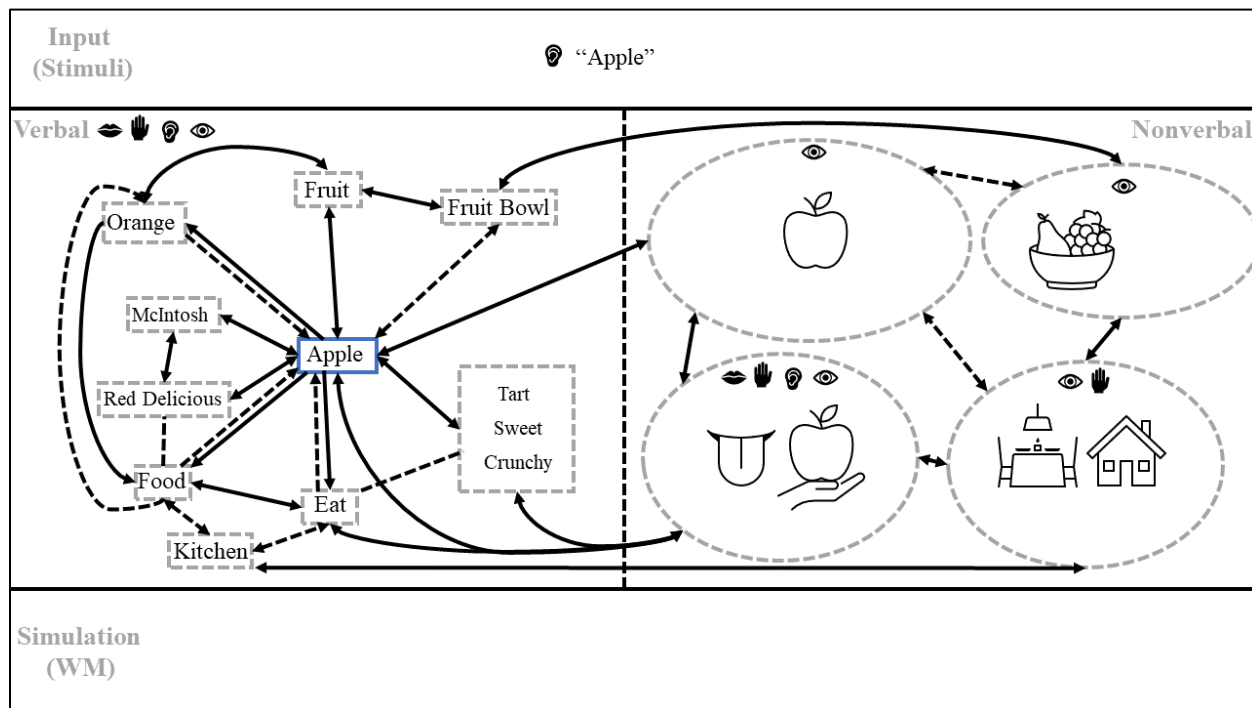


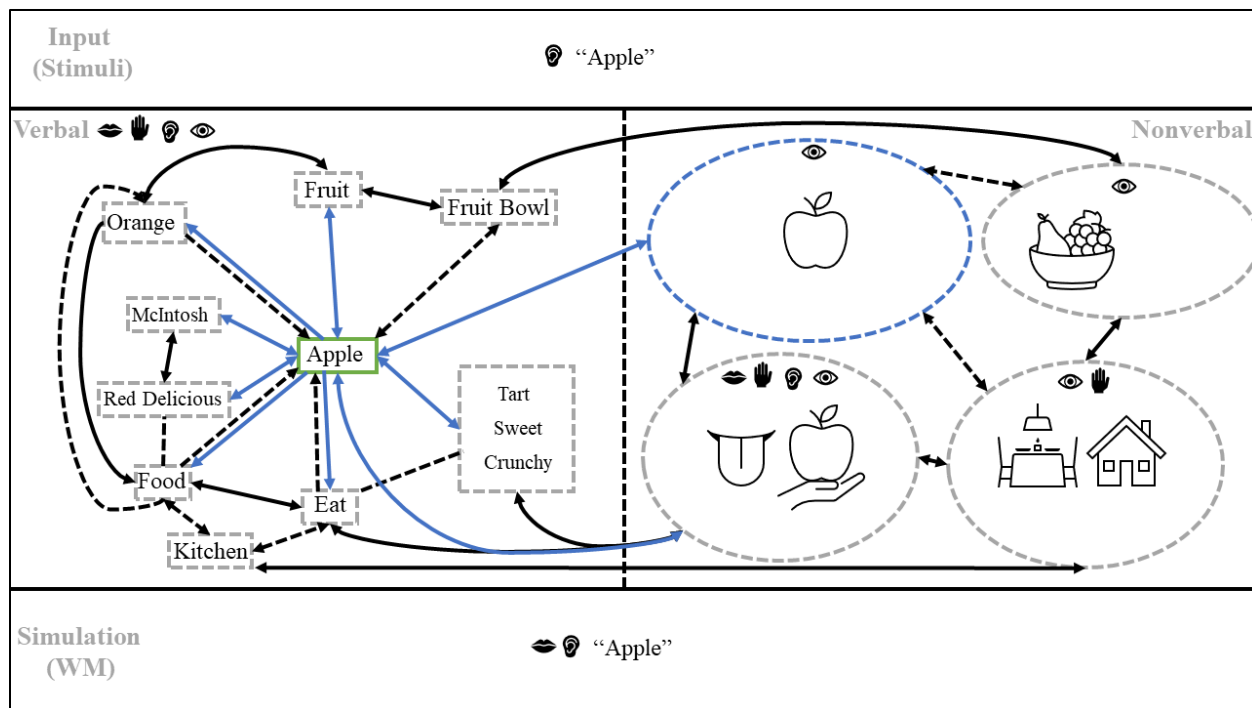
Figure 2.14. Two Types of Simulation. Categorization Decision for "Fruit" (a) and Imagined Taste (b)

A similar pattern of activation would also be involved for verbal stimuli, with initial activation of the word through the visual system (Figure 2.15a), spreading activation to associations and cross-system referential connections (i.e., word identification as understanding of its referent, Figure 2.15b), and lastly, category availability through further spreading activation within and across systems (a wider 'semantic' meaning, Figure 2.15c).

(a). Verbal Recognition: Initial Representational Activation of “Apple” Logogen



(b). Identification: Logogen Activation and Spreading Activation





## (c). Categorization Availability

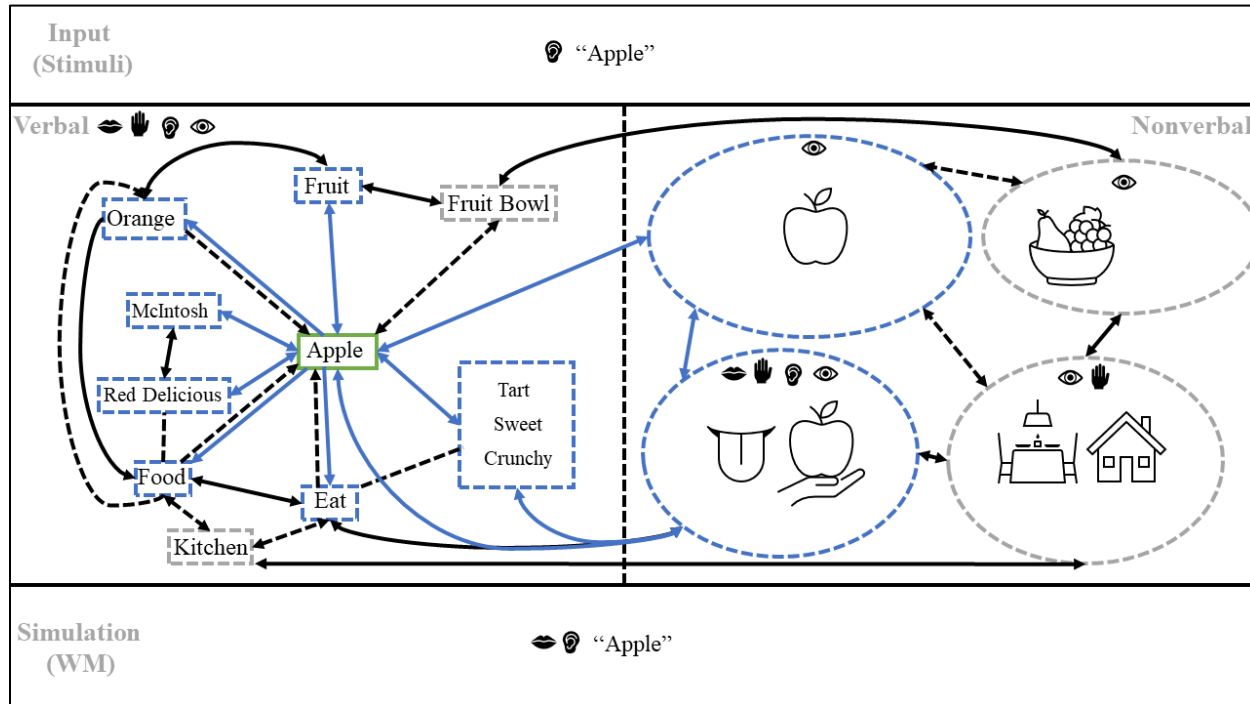


Figure 2.15. Recognition (a), Identification (b) and Categorization Availability (b) from Verbal Representational Activation

Distinguishing between recognition, identification and categorization helps to specify a time-course component of memory use that will be vital in discussing musical categorization. While in practice, these types of processing may be available nearly at the same time, particularly for overlearned items (i.e., recognition, identification and categorization have similar reaction times), it is an important distinction to retain when discussing learning. Showing the activation patterns in LTM and their use in WM will also allow for discussions of developing memory expertise. Recall that DCT posits memory facilitation due to selective reliance on imagery and verbal systems, and that each is specialized for different types of processing (synchronous vs. sequential), such that the imagery system is best for tasks involving discrete memory (e.g., tasks that are not constrained sequentially, such as free recall), whereas the verbal

system is best for tasks involving sequential memory (e.g., serial learning) (Paivio 2007, 80). However, if representations are stored and/or accessed in the same modality (e.g., verbal string and auditory sequence stored in auditory mode), memory can actually be impoverished rather than facilitated as information in the same modality can lead to interference in working memory (Paivio 2007, 49). Therefore, stimuli and resulting representations that occur in the same modality (e.g., auditory), and have a similar organizational structure (e.g., sequential) will compete in memory for resources in encoding, retrieval, and use. This is particularly problematic for music theoretic expertise, especially as verbalizations and auditory representations overlap in their primary modalities (auditory and motor) and their organizational structure (sequential and hierarchic); when such representations are used in imagery, they are maintained in the phonological loop which leads to a high probability for interference (Salamé and Baddeley 1989). This suggests that in the acquisition of music theoretic expertise, certain concepts may benefit from the additive benefits of dual coding, including verbalizations, while others will require more deliberate practice to overcome potential interference effects. This will be discussed in detail in chapter 4.

### Accounting for Introspection: Imagery, Feeling of Knowing, Emotional Construction and Interoception

This project positions introspective judgments as a central feature of music theoretic expertise in situated conceptualization. Introspection, in the form of affect and metacognition, is vital for abstract concept representation and simulation. However, introspection in both DCT and DIPSS is not particularly well specified, covering as it does a wide array of phenomena including emotion and affect, motivation, goals, and unspecified metacognitions. In this section, I will

outline the aspects of introspection relevant to the current project. These may change with, or have impacts on, expertise acquisition, and include subjective assessment of vividness and control in active imagery use, here defined as the perceived vividness (lifelikeness) of simulation in WM, as well as the perceived ease of control over simulation in WM (a form of metacognition). Another feature of introspection is also a separate aspect of metacognition, the sensation of memory function during priming without item selection—or ‘feeling of knowing’—essentially, the availability of sensations of familiarity when memory networks are primed, but little to no active nonverbal simulation occurs in WM. Finally, I will discuss interoception as a distinct form of imagen representation in the nonverbal system. This will replace Paivio’s vague use of ‘emotion’. Such representations are central to emotional construction—the interpretation of internal states (cognitive, affective), one versus another.

#### Imagery and Introspection: Subjective Assessment of Imagery

Imagery is often defined as the persistence of an introspectively available sensory experience, including one constructed from components drawn from long-term memory, in the absence of direct sensory stimulation (see Hubbard 2010, 302). The subjective nature and inaccessibility of imagery to external observation has proved a major challenge to empirical researchers. Traditional empirical methods stemming from the study of visual imagery have been adapted for auditory imagery and typically use a subjective measure to predict outcomes or performance on an objective measure. The subjective measurement most often takes the form of a questionnaire in which participants are prompted to generate some form of imagery from a verbal description and then rate properties of their introspectively perceived imagery on a Likert scale (e.g., ranging from 1–7). The predictive validity and reliability of such measures has

historically been quite inconsistent within the visual domain, leading many scholars to conclude that they have poor predictive validity in general (McAvinue and Robertson 2007, 196). More recent questionnaires have demonstrated more robust relationships between subjective and objective measures. These questionnaires divide subjective imagery scales into subtypes, such as spatial imagery and object imagery, which are queried using specific, objective tasks involving those imagery processes, such as mental rotation for spatial imagery and degraded image tasks for objects (McAvinue and Robertson 2007, 204–205). The more accurately a subjective measure captures the properties of imagery relevant to completing a specific objective task, the better that measure is for predicting objective performance.

In the auditory realm, a recent subjective measure, the Bucknell Auditory Imagery Scale (or BAIS, Halpern 2015), has shown promising predictive power for behavioral task performance in the domains of auditory and musical imagery. This measure is designed to capture two distinct types of processing in auditory imagery: generation (the ability to bring an image to mind) and transformation (the ability to make changes to a generated image). These processes are measured using relevant qualitative features of imagery, vividness (for generation) and control (for transformation). The BAIS measures auditory imagery broadly, and includes environmental, speech and musical items.

The vividness rating scale measures image generation ability by prompting individuals to subjectively judge how life-like their imagery is. In the questionnaire, a verbal cue is presented to subjects requiring them to construct an auditory scene, after which they are required to rate the vividness of their auditory imagery on a scale from 1 (no image present at all) to 7 (as vivid as real sound). The vividness subscale on the BAIS has been highly correlated with several behavioral and neurological measures: high vividness scores have been shown to predict accurate

singing of individual pitches (Pfordresher and Halpern 2013), longer melodic sequences (Greenspon, Pfordresher and Halpern 2017), as well as memory for previously heard melodies (Herholtz, Halpern and Zatorre 2012). Vividness has also predicted the explicit use of imagery strategies for scale degree imagery in the Pitch Arrow Imagery Task (Gelding, Thompson and Johnson 2015), as well as the number of involuntary musical imagery or earworms experienced by individuals (Floridou et al. 2015). Recently, the vividness subscale was shown to predict the ability to perceive expressive timing patterns (Colley, Keller and Halpern 2018). Performance on the vividness subscale has also been correlated with increased brain activation during the encoding of imagined melodies (Herholtz, Halpern and Zatorre 2012), and increased activity during melody reversal (Zatorre, Halpern and Bouffard 2009), as well as increased grey matter volume of the brain in the supplementary motor area (SMA, implicated in subvocalization) (Lima et al. 2015).

Auditory imagery control entails the degree to which one has control over their ability to manipulate and make changes to their musical imagery. In the control subscale, subjects are instructed to construct an auditory scene, after which they are prompted to make some sort of change to the generated image. They are then required to rate ease of change of their auditory imagery on a scale from 1 (no image present at all) to 7 (extremely easy to make the change). Compared to the vividness subscale, the control subscale has shown slightly less predictive power, which is likely due to imagery control's greater complexity and reliance on a wider range of related cognitive mechanisms, such as working memory (WM). This subscale has nevertheless been able to predict better performance on the Pitch Imagery Arrow Task (Gelding, Thompson and Johnson 2015), has been correlated with lower error rates in the imitation of pitch sequences

(Greenspon, Pfordresher and Halpern 2017), and was predictive of the ability for musically untrained subjects to synchronize with expressive music (Colley, Keller and Halpern 2018).

In the current project, imagery will refer to sensory information of any modality, verbal and nonverbal, prompted by indirect activation that is held in WM during simulation. The current project will only use the constructs of *vividness* and *control* as two subjective aspects of imagery. Imagery can therefore have low or high vividness based on the level of detail held and perceived in WM (see Figure 2.16). The ability to selectively simulate imagery of various vividness depths (e.g., lifelike, or reduced featured/schematic), is vital for memory expertise, as the ability to selectively toggle image detail means that imagery can be used efficiently, avoiding overburdening WM demands in different contexts. The ability of the central executive to modify aspects of imagery vividness or imagery quality, as well as the ability to switch between different types of images, is reflected in the subjective experience of imagery control—a form of metacognition reflecting ease of transformation of images held in WM during simulation. Subjective imagery control is therefore an indicator of LTWM skill, or the ability to select, retrieve and use different simulators in WM to adapt to various contexts, conditions, and goals (see Figure 2.17).

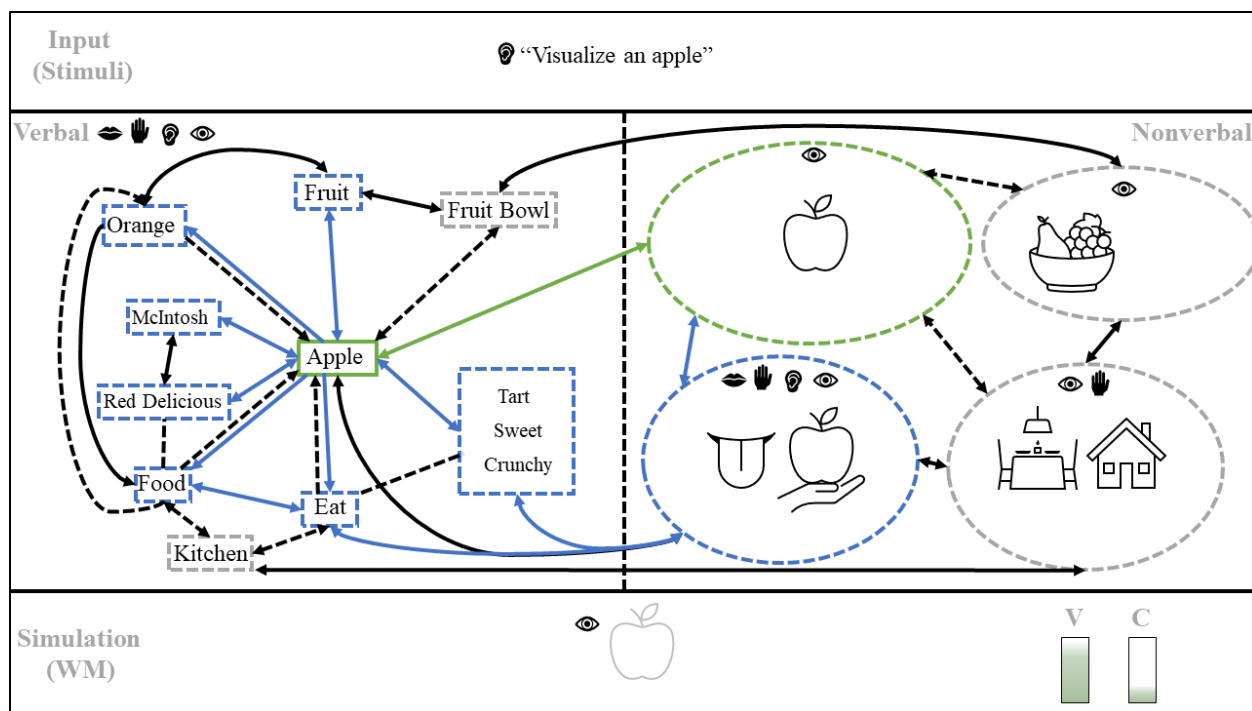


Figure 2.16. Vividness Scale Added to WM

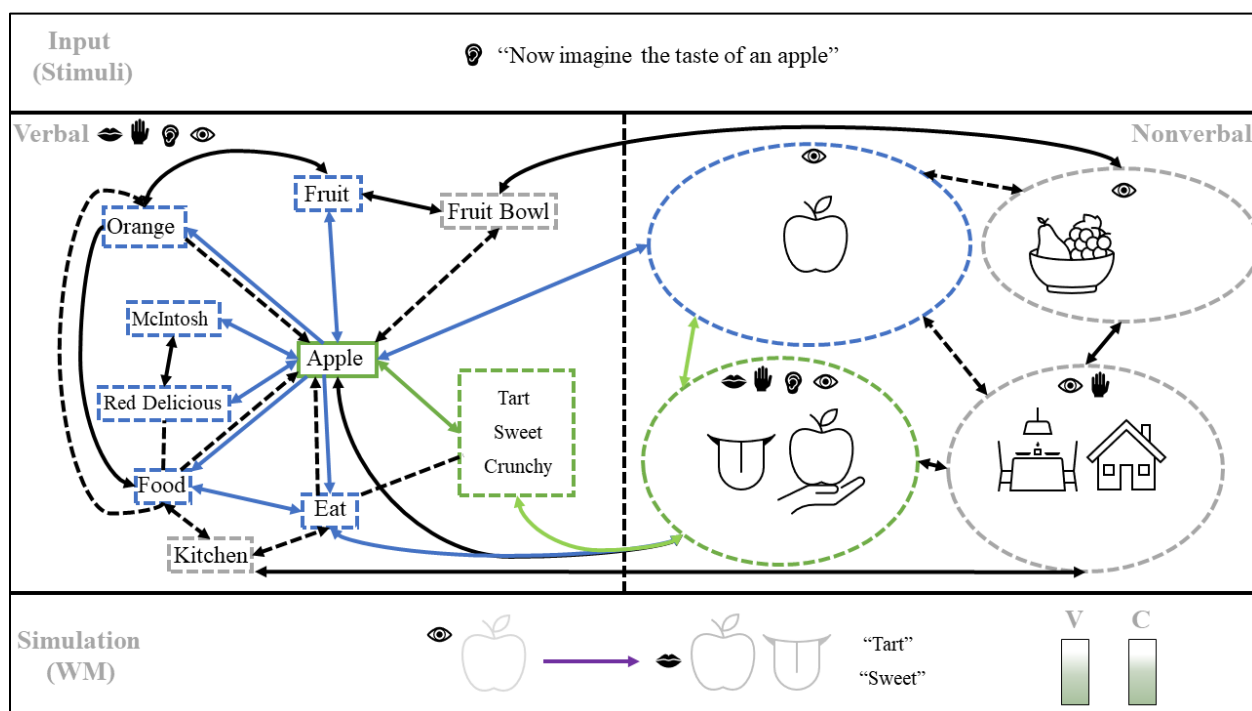


Figure 2.17. Control Scale Added to WM Showing Ease of Change of a Simulation

## On Sensations of Memory: Priming, Item Pre-Selection, and “Feeling of Knowing”

The second category of introspection regards metamemory—or the internal sensation of memory function and processes, including knowledge of and control over such processes (Dunlosky and Bjork 2008). While there are many subjective types of metamemory judgements, this project will focus on feeling-of-knowing (hereafter FoK), which primarily refers to the subjective awareness of retrieval processes during memory use (see Figure 2.18). Traditionally, FoK judgments refer to the sensation of memory when a current target is stored in memory, but is not recallable (Neilson and Narens 1990, 130). Such a sensation is typified by the tip-of-the-tongue state, where a cue prompts retrieval of a known item, but that item cannot be retrieved (Brown and McNeill 1966; Schwartz and Metcalfe 2011). Aspects about associated knowledge can be accessed however, as well as potential or partial features of the target (see Figure 2.19). Here, the prompt “What does Snow White eat?” prompts the retrieval of the word ‘apple;’ however, due to retrieval failure, the word cannot be recalled. The person trying to recall may be able to identify surrounding associated knowledge held within the simulator—the target is a fruit, not an orange, is round, or even that word starts with an ‘A,’ but cannot retrieve the word. However, because they are *certain* that the target does in fact exist in memory, the FoK judgement, shown on the bottom right corner (next to vividness and control), is high.



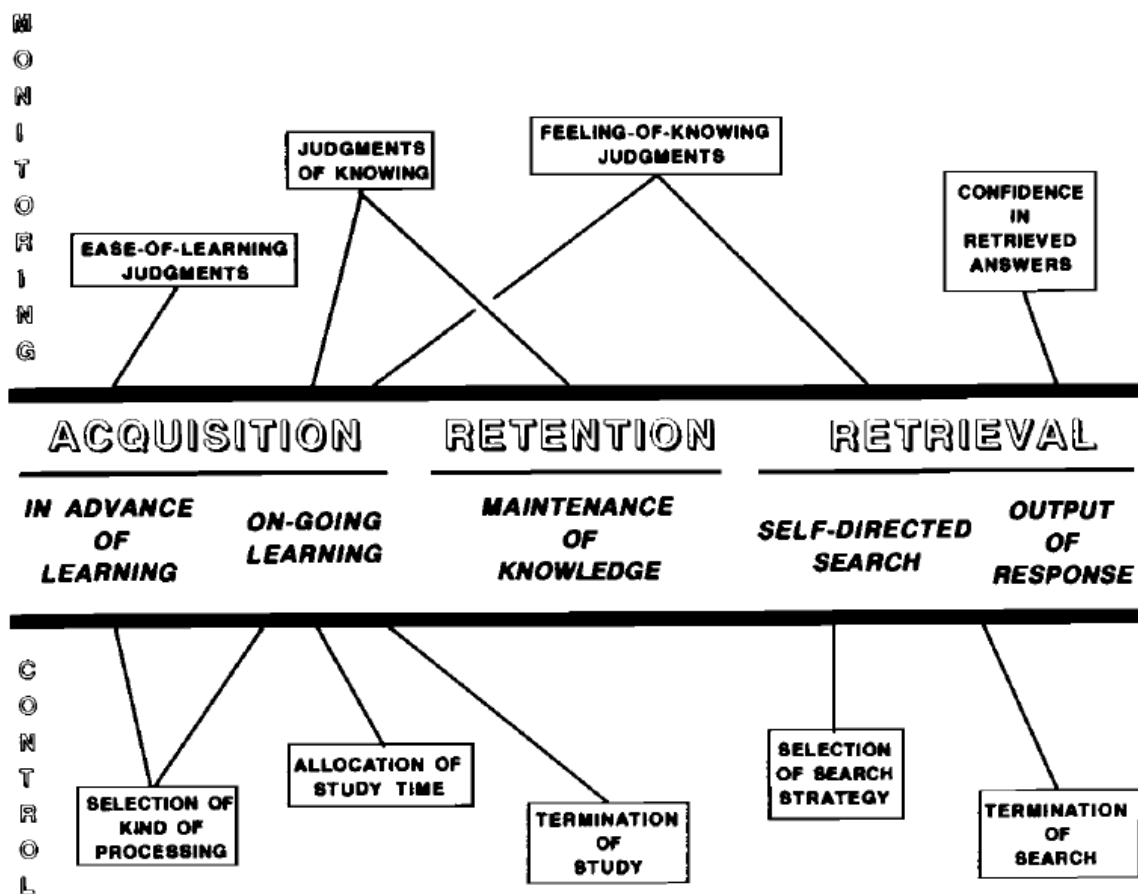


Figure 2.18. Relationship Between Various Metacognition Judgements and Learning Processes (Neilson and Narens 1990, 129)

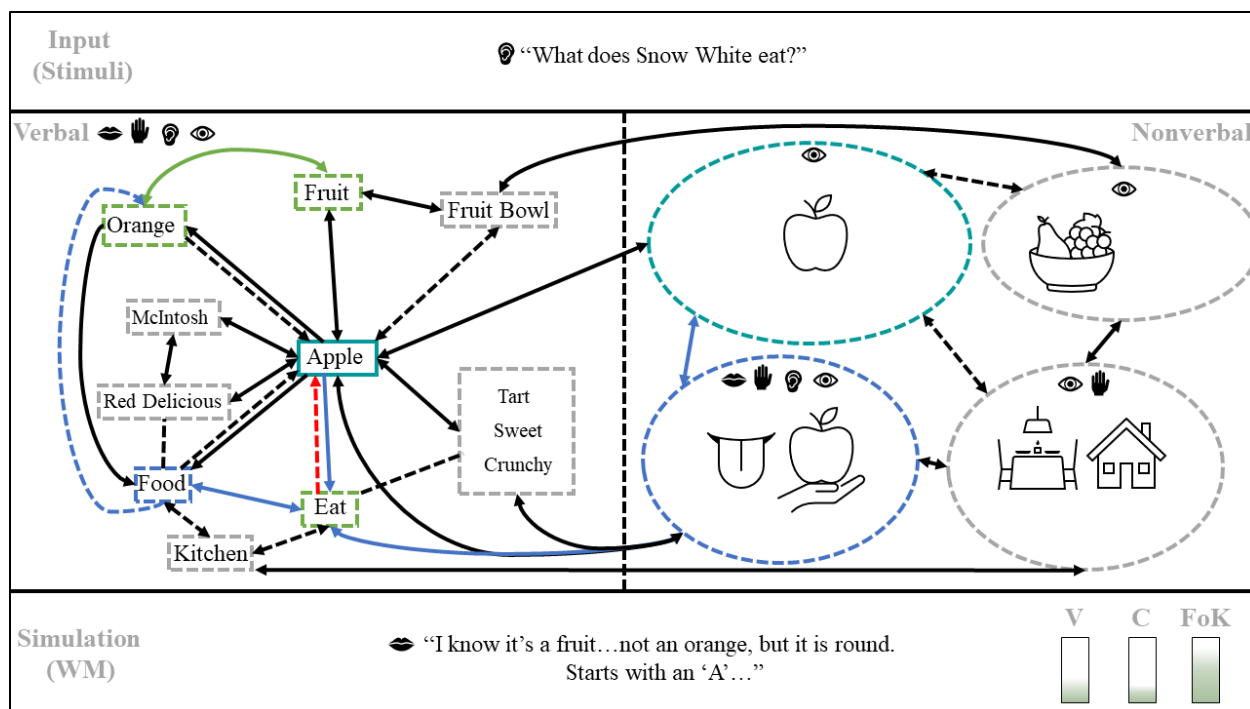


Figure 2.19. Feeling-of-Knowing (FoK) During a Tip-of-the-Tongue State

Typically, the better encoded a target is—even to the point of overlearning—the better and more accurate FoK judgements are (Neilson and Narens 1990, 145-146). ‘Standard FoK’ refers to processes during target retrieval. The familiarity of the cue for a target is also relevant for FoK judgements. This is sometimes referred to as ‘preliminary FoK’, and refers to the viability of the provided cue early in the retrieval process (Van Overschelde 2008, 59; Schwartz and Metcalfe, 1992). Generally, the longer someone is willing to search in memory for an item, the higher the FoK judgement (Van Overschelde 2008, 60).

This project will expand on the notion of preliminary FoK to include sensations of retrieval that occur during priming of representations in a simulator, reflecting the availability of unselected but primed items in that simulator. Here, FoK will refer to the state of feeling that occurs when a concept network is primed and ready for use. In general, the more items in a simulator, and the stronger the encoded representations and highly probabilistic interconnections

for those items, the higher potential FoK that exists for that simulator. If a fruit has only ever been learned through academic study—say textbooks with pictures—this would yield a limited number of representations in its simulator (see Figure 2.20). When a representation in that simulator is activated, and before other representations are retrieved, the corresponding FoK would be rather low because there are not many retrievable items. If a simulator is much more developed for that same category, activation of the simulator would prime many associated representations, resulting in a higher FoK rating (see Figure 2.21).

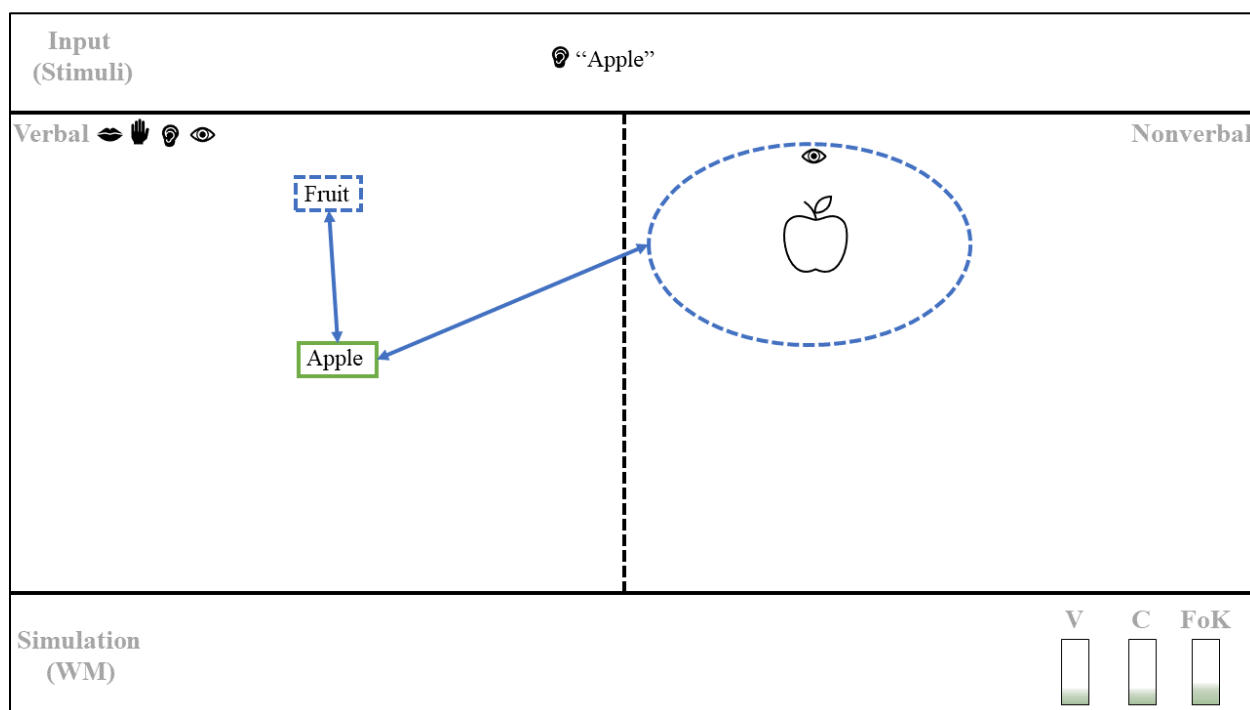


Figure 2.20. Sparse Simulator with Low FoK

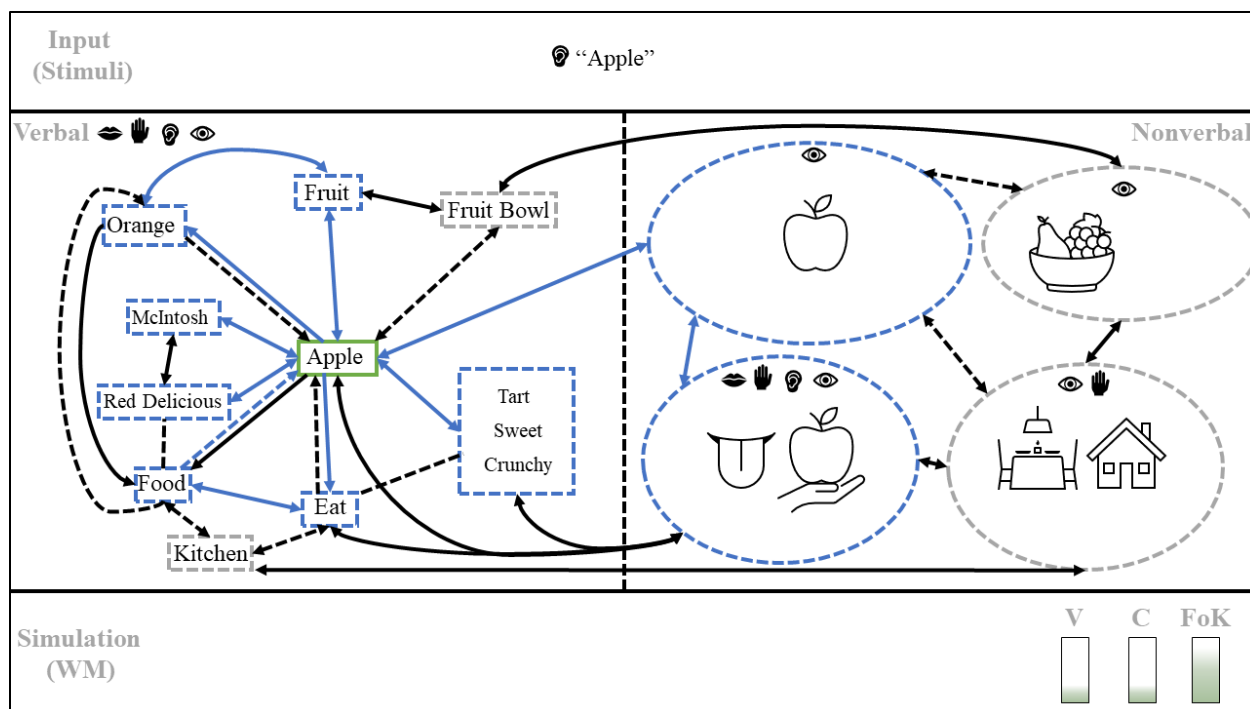


Figure 2.21. Elaborated Simulator with High FoK

## Emotional Construction and Interoception

Lastly, we turn to interoception in emotional construction. In this section I will integrate Barrett's conceptual act theory (CAT) of constructed emotion (2017b) with DCT principles to provide a more complete account of introspective states related to emotion and affect. I will use these findings to expand upon Paivio's nonverbal representations to include and account for interoceptive states which are crucial representations used for emotional categorization in cognition (Barrett 2017a). Interoception, generally understood as sensations arising from inside the body (from muscles, organs, etc.), has been expanded in recent years to include 'common sensation' in the body.<sup>46</sup> While interoception as a subjective measure is typically discussed as

<sup>46</sup> Craig (2015) expands interoception to include *thermoreception* (temperature) and *nociception* (inputs activated by damaging or threatening stimuli) (p. 3). These are included in so called body awareness or *interoceptive awareness*, which Craig argues arises from a distinct interoceptive cortex and dedicated pathways in the brain (Craig 2015, 4-5).

three separate constructs (awareness, sensibility, and accuracy; see Forkmann et al. 2016, Garfinkel et al. 2015), here I will only refer to interoception generally rather than using each construct separately. In this section, I will demonstrate that interoceptive representations are vital for distinguishing between different affect-based or qualitative concepts; these internal representations help to distinguish between related but distinct simulator networks. In combination with sensations of FoK, interoceptive representations allow individuals to learn to detect differences in simulator network activation by further grounding an FoK in particular embodied states. This claim is supported by research that demonstrates consistent relationships between interoceptive states and metamemory, FoK in particular (Garfinkel et al. 2013; Chua and Bliss-Moreau 2016; Fiacconi, Kouptsova and Köhler 2017).

Much like the cognitive appraisal approach outlined by Paivio, Barrett's theory of constructed emotion involves evaluation and categorization as a central feature of emotion perception. Her approach is similar to Barsalou's DIPSS in that it views emotions as abstract categories that are used to construct or simulate a concept. An object concept may be used to simulate objects in perception by directing attention towards specific properties in categorization; in an analogous manner, emotional concepts do the same with sensory information arising from both outside and inside the body (i.e., interoceptive states) (Barrett 2017b, 29). Barrett theorizes that emotions are wholly context dependent, dynamic abstract categories more akin to events than objects, that only exist in in social contexts and in minds (Barrett 2014, 292);<sup>47</sup> she does not view emotions as having intrinsic or static representational

---

<sup>47</sup> This is particularly pertinent in emotion perception because both externally perceived and internally perceived emotions depend on context. For example, it is possible to interpret someone who is crying as exhibiting either 'sadness' or 'happiness' depending on context (e.g., funeral context for sadness, family reunion for happiness). Similarly, different emotional states may be interpreted from the similar interoceptive states based on context. Barrett (2017b) provides a personal example; she interpreted the interoceptive state of a flushed face and fluttering stomach as attraction while on a date, only to realize later than evening after returning from the date that it was in fact a result of the flu (p. 30).

states, Emotional categorization is situated conceptualization: the knowledge used during categorization is dependent on context is enactive, and is used by an ever-predicting brain to prepare for situated action (Barrett, Wilson-Mendenhall, & Barsalou, 2014). Emphasizing the importance of language in this process, Betz, Hoemann and Barrett (2019) demonstrates that words provide stability and consistency in the perception of emotional states from pictures of faces compared to free labeling of these states. Hoemann, Delvin and Barrett (2020) similarly suggest that infants learn abstract emotion categories through language. Emotion words stabilize the interpretation of large sets of overlapping internal states paired with highly variable contexts.

Similarly, a DCT perspective views emotion concepts as abstract categories dependent on language for evaluation. Rather than being highly abstract networks of associated logogens, emotion or affective concepts are situated in a sort of middle ground of abstraction; they are dependent on language and language networks for stability and interpretation but are also equally grounded in interoceptive representations. Adapting DCT to account for interoceptive states as a concrete and distinct form of sensory representation, I will replace the vague notion of “emotion” as a primary sensorimotor system (see Figure 2.2) with the more precise term interoception (see Table 2.1).<sup>48</sup> Little is known about the organizational structure of such representations. I suggest that they vary widely depending on context,<sup>49</sup> but in general can be easily paired with other sensory or language representations without substantial cost to attentional resources.

---

<sup>48</sup> It is important to reinforce the distinction between interoceptive imagens and other sensory representations as there is often a considerable overlap due to concurrent presentation, creating strong within system sensory associations. This is particularly true for motor imagens and interoceptive imagens. For example, motor imagens are formed through interaction with external objects and therefore represents those features of objects as interpreted through touch (e.g., the *feeling* of playing piano, depressing keys, specific actions, etc.). An interoceptive imagen of the same context would include more generalized internal bodily states and general sensations of movement and gesture while playing (e.g., entrainment, swaying on the piano bench, etc.). This distinction is somewhat subtle, but can be clarified through where attention is directed when these representations are active. With a motor imagen, it's directed toward properties of interaction, in interoceptive imagens, it's directed toward internal sensations.

<sup>49</sup> In terms of their organizational structure, interoceptive imagens are more variable as they are more context dependent and because interoceptive states consistently co-occur with all other sensorimotor systems. They can

Table 2.1. Updated Sensorimotor Systems in DCT

<b>Table 2.2</b>		
<b>Updated Sensorimotor Systems in DCT</b>		
<i>Sensorimotor Systems</i>	<i>Symbolic Systems</i>	
	<b>Verbal</b>	<b>Nonverbal</b>
Visual	Visual Language	Visual Objects
Auditory	Auditory Language	Sounds
Haptic	Braille, Handwriting	“Feel” of Objects
Gustation	-----	Tastes
Olfaction	-----	Smells
Interoception	-----	Internal Bodily Sensations

I view acts of emotion or affective categorization as specific instances of activation patterns among a network of representations. These include within-system and cross-system activation between interoceptive imagens and logogens in a given context. Interaction with an apple may result in hunger pangs (sensations of pit in the stomach, grumbling, etc.). Figure 2.22 represents the activation of representations through association within the sensory system. Viewing an apple (activating apple imagen), the interoceptive imagen for ‘hunger pang’ is also activated; the evaluation or interpretation of this internal state is likely to be ‘hunger.’ DCT views this as referential, cross-system activation of the word ‘hunger’ and activates HUNGER as a concept through associational priming in the verbal and sensory systems. In a different context, the network of activation may change, resulting in a different evaluation or interpretation. The same interoceptive imagen (pit in the stomach) may be activated by viewing an unfinished dissertation chapter; evaluation in this instance may be fear or anxiety (see Figure 2.23).

---

therefore be more or less synchronously or sequentially organized depending on their context. For example, hunger pangs have some synchronous organization (i.e., sensations of pit in the stomach, and bubbling sensation), and some sequential organization (e.g., gurgling sensation and change over time, which may depend on how hungry one is!). Other interoceptive states, like heart rate sensation, will have more sequential organization than synchronous.

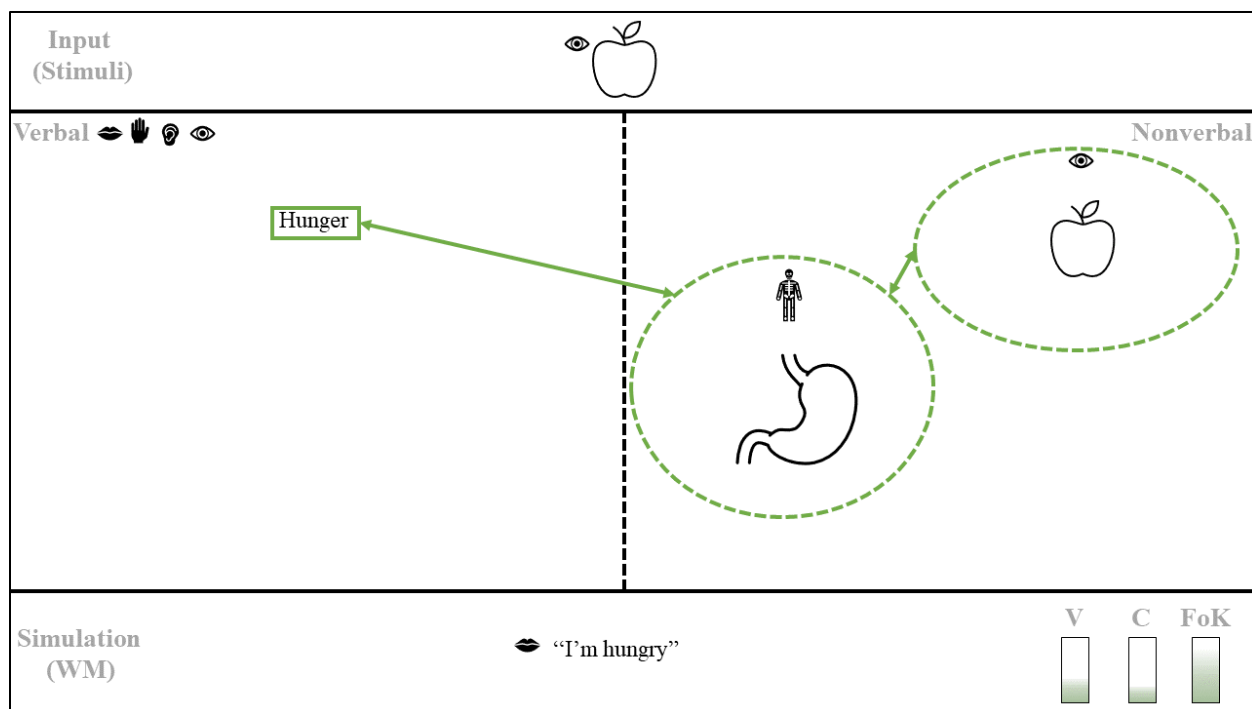


Figure 2.22. Simulation of Emotional Construction "Hungry"

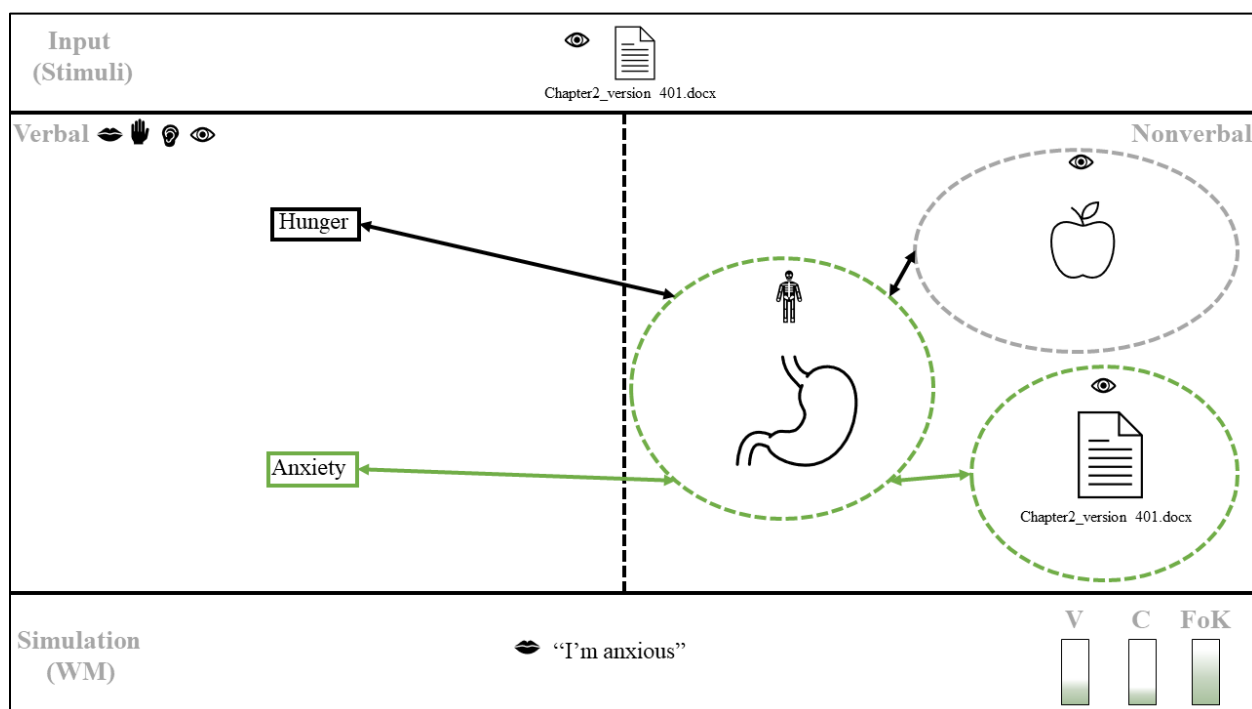


Figure 2.23. Simulation of Emotional Construction "Anxious"



Contrary to Paivio's position that affective language is *indirectly* connected to emotional representations as these are mediated through sensory representations, I suggest that interoceptive states *are* the sensory representations directly connected to language and therefore act as the mediating states between 'objects, events, and people and affective language.' In this sense, interoceptive imagens have in fact, many direct referential connections to emotional and affective language (logogens). The probability of which referential connections get activated in a given instance, as Barrett suggests, is wholly dependent on context (i.e., which other representations in the network are primed). Therefore, the evaluative states afforded by the interoceptive imagen 'pit in stomach' are multiple and varied, and can be influenced by context, such as 'mistaking' anxiety related sensations for hunger, leading to stress eating (see Figure 2.24). Similarly, their activational structure is much more varied as the system can be primed from many different representations.<sup>50</sup> Importantly, research demonstrates that through training and expertise, humans can improve interoceptive awareness, leading to more accurate evaluation of interoceptive states.<sup>51</sup> This process involves learning to distinguish between subtleties of the interoceptive representations in question, and in creating more specific simulator networks for each concept. For example, one may learn to distinguish pit-in-the-stomach feelings for hunger and anxiety, with the help of language (Figure 2.25). This may bring awareness to differences between these two states, such as the hunger-pang being more indicative of bubbling and gurgling sensations, and sounds emanating from the stomach, whereas anxiety pangs may be

---

<sup>50</sup> Basically, 'hunger' or 'anxiety' in this case can be instantiated through a wide variety of pathways; either through activation of a wide range of logogens, priming both internal states and other sensory representations in sensory systems or *visé vera*, or through many different contexts. Again, because similar interoceptive states are common to many different affective states (evaluations), they are 'many ways in' to activating these networks.

<sup>51</sup> See Craig 2015, 7. Humans can learn to improve interoceptive awareness through various activities to hone interpretation of bodily states, like sports or yoga, and also through mindfulness activities (e.g., learning that the difference between 'anxiety' and 'hunger' lies in the co-occurrence of different states; pit in stomach with gurgles for hunger and pit in stomach with increased heart rate for anxiety).

more pit-like, accompanied by nausea and drops in body temperature. Language and their referential connections to more specified interoceptive representations allow for each of these concept-networks to be more precise, subsequently affording more accuracy in emotional construction. Similarly, the FoK values for each of these networks is higher, and each is imbued with particular internal sensations now differentiable interoceptively and verbally.

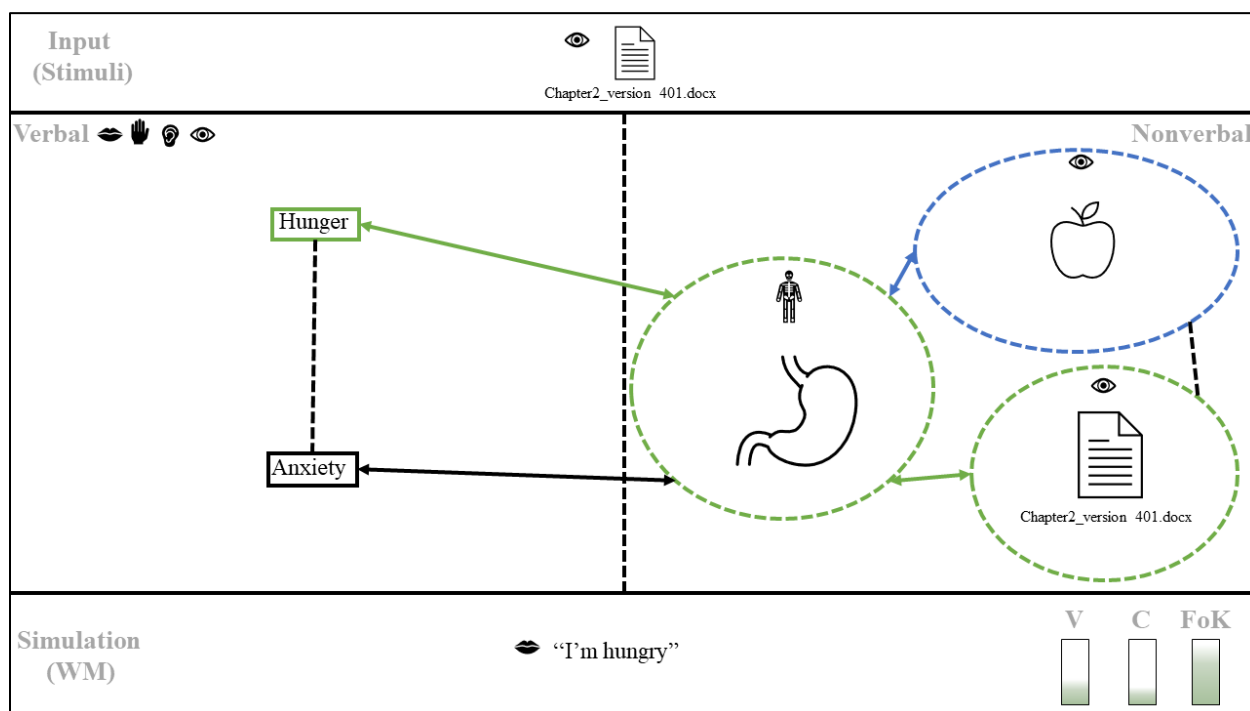


Figure 2.24. Misinterpretation of Anxiety for Hunger

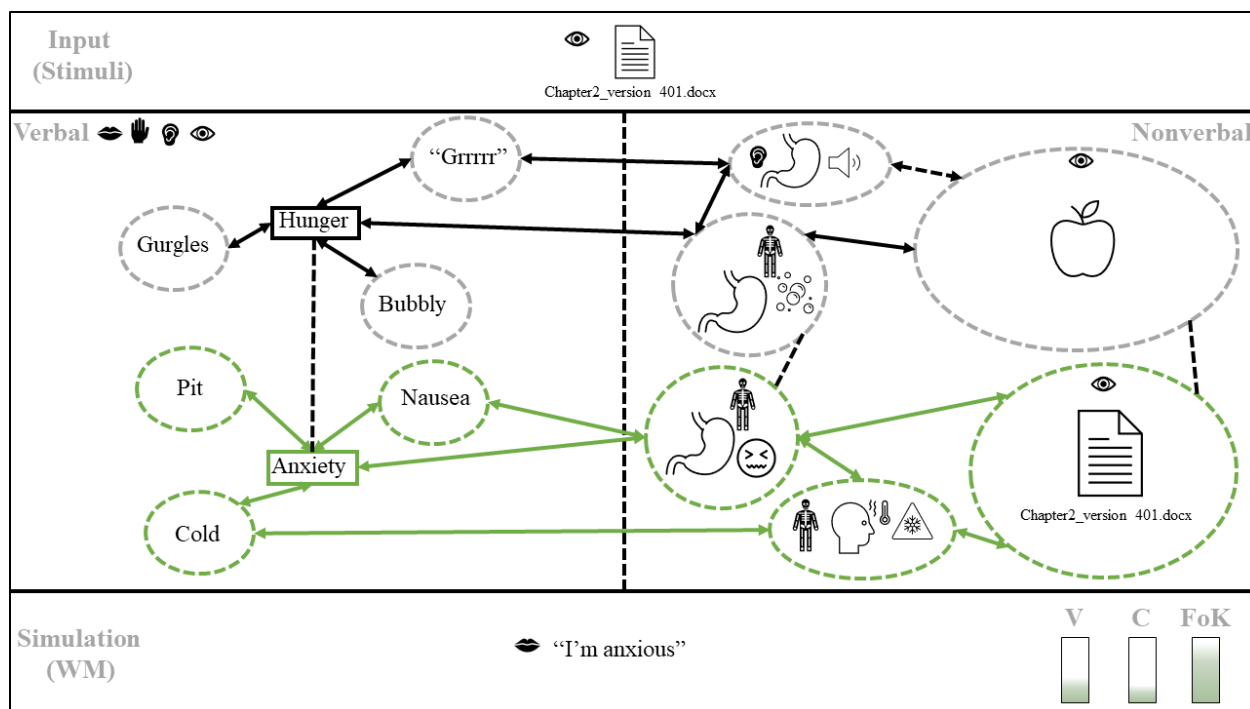


Figure 2.25. Elaborated Simulators for Hunger and Anxiety

Due to the multitude of connections, both within the sensory system and across referentially in the verbal system, interoceptive imagens can give rise to a wide variety of different constructions given different contexts and goals. Over time, emotional and affective language gain more probabilistic connections to specific sets of interoceptive imagens, providing stability in their representation (what Paivio notes as “generalized sensory-like” qualities). This multiplicity inherent in interoceptive imagens affords the ability to ‘ground’ abstract concepts (from a DCT perspective, make them *more* sensory). This position concurs with more recent findings that emotions and affect are considered introspective states used in grounding abstract concepts (Altarriba et al. 1999; Altarriba and Bauer 2004; Barsalou and Weimar-Hastings 2005; Connell, Lynott, and Banks 2018). This approach is also supported by research findings showing content differences for concrete and abstract concepts. Wiemer-Hastings and Xu (2005) for example, found that concrete concepts contain content properties, whereas abstract concepts

contain more features relating to subjective experience rather than item properties. This distinction will be vital for understanding the representational content of music theoretic concepts, their connections within and across systems, as well as their use and affordances in music theoretic expertise.

### DCT Updated: A Conclusion and Summary

In this chapter, I have summarized, updated, and adapted DCT to be able to account for musical categorization behaviors, including introspective judgements vital for discussing qualitative individual differences and the development of music theoretic intuition. DCT has been merged with Barsalou's DIPSS, which maintains a distinction between simulators—conceptual networks in LTM—and simulation—their usage in WM for recognition, identification, and categorization. In providing better specification regarding the connection types within LTM (high and low probability, uni- and bi-directional), I will be able to discuss how LTM structure changes over the course of expertise acquisition. Similarly, in adding in LTM activation types for use in WM (black for inactive, green for active, blue for primed), I will be able to discuss differences in retrieval processes during active memory use. Because I also specify the modality of the input, LTM traces, and WM simulation, I am also able to discuss potential interference effects in long-term working memory. Lastly, by incorporating accounts of introspective judgements into WM simulation—including imagery vividness and control, FoK, and interoceptive grounding in emotional construction—I will be able to discuss the qualitative differences during memory skill acquisition and use in music theoretic expertise. In the next chapter, I will apply the current framework to provide an account of embodied Galant schemata

representation that can distinguish between individual and group-level differences in LTM structure and WM instantiation (i.e., simulation).

## Chapter 3

### **An Embodied Approach to Galant Schema Representation**

The goal of this chapter is to introduce an embodied approach to Galant schema representation using the adapted version of DCT developed in the previous chapter. In the first section, I highlight the shortcomings of Galant schema theory in offering insight into potential effects of expertise between novices and experts, which I argue occur in part due to differences in LTM representation, organization, and access. I show that these shortcomings stem in part from the adaptation of schema theory from the cognitive sciences, which positions abstract, amodal propositions as the primary representational format. As a result, the schema approach as it is typically understood is currently unable to account for differences in representation that may exist between populations as a function of learning and experience, something for which the current framework is able to compensate.

In the second section, I characterize differences between novice and expert Galant schema representation as distinguished by loose versus structured representation (adapted from Barsalou 2003a), which is demonstrated by example in this chapter's third (novices) and fourth (experts) sections. In the third section, I describe the novice representation as one that does not directly represent Galant schemata features, but instead captures a holistic gestalt of the memory state of the perceiver during listening. Because the interaction of encultured listeners is primarily auditory, their representation of Galant schemata is more modally centralized (audition and interoception in the nonverbal system), and is more loosely (or less probabilistically) organized. As a result, I demonstrate that novices' abilities to actively simulate Galant schemata categories are highly restricted. In the final section, I describe expert Galant schemata representation (in the

domain of music theory) as one where schemata features and relations are directly represented in property and relation simulators distributed across modality (vision, audition, motor) and system (verbal, nonverbal). Such simulators are explicitly acquired by theorists using music theoretic concepts for properties and relations (e.g., scale degrees, counterpoint, forms, etc.), which function to encode the attentional focus of an interaction with a schema into separate, but associated, episodic traces in LTM. Such simulators are highly structured—more probabilistically interconnected with one another—and form large simulator pools which come to represent Galant schemata categories. Because this manner of representing categories is more distributed across mode and system (verbal, nonverbal), I demonstrate that simulation ability is more flexible and better able to be adapted to shifting task demands in online categorization.

### When ‘Schemata’ are not Enough: Issues in Abstract Unitary Representation

Schema theory has been indispensable for theorizing and conceptualizing knowledge representation, particularly in the domain of language acquisition. However, it has been critiqued by competing approaches (DCT included) for its ambiguity and lack of operationalizability. Paivio and colleagues have discussed two primary benefits that schema theory has had in research on reading: first, an emphasis on the constructive nature of comprehension, and second, the crucial role of a reader's prior knowledge in the construction of meaning comprehension (Sadoski, Paivio and Goetz 1991, 465). However, they argue that the term ‘schema’ is itself ill-defined and is often used to describe a vast array of cognitive phenomena for which explanatory mechanisms do not exist. The pervasive use of the term, with little to no consensus on what representations make up various types of schemata creates “...an illusion of consensus and has left the impression that we have a more profound understanding of cognition...than we do”

(Sadoski, Paivio and Goetz 1991, 465). The many definitions of schema and analogous abstract entities—such as templates, scripts, and plans—provide no extra explanatory power and therefore, merely increase the explanatory burden, complicating any potential predictive power afforded by the theory (Paivio 2007, 57).

In addition, the nature of mental representation of schemata is quite unclear. Some researchers explicitly invoke abstract proposition as the fundamental unit of representation (see Anderson and Pearson 1984; Kintsch 1998). Others, however, do not specify the nature of the representational units themselves, and instead posit theoretically infinite, recursive embedding of schema units into one another (Rumelhart and Ortony 1977; Rumelhart 1980). Such approaches often explicitly refer to schema representations as abstract prototypes (see Rumelhart 1980, 34), which in turn invokes amodal propositional representation by proxy to prototype theories of mental representation (as discussed in the previous chapter). Theories that utilize the schema concept but in the context of an exemplar approach to representation, as is the case with modern constructional approaches to language acquisition (see Abbot-Smith and Tomasello 2006), still suffer from the drawbacks of amodal abstractions because exemplar theory has, to date, not adequately accounted for multimodal representation.<sup>52</sup> The schema concept has been a crucial heuristic for conceptualizing aspects of cognition—particularly as contextualized, process-oriented patterns of thought, or groups of associated knowledge structures—but is limited in what it offers with regards to the format and nature of such units.

---

<sup>52</sup> It is important to note that the schema concept from its inception was not conceptualized as amodal. Early scholars often refer to schemata as nonverbal, sensorimotor entities (see for example, Piaget 1952). F.C. Bartlett, considered to be one of the founders of schema theory in modern psychology, discussed the vital nature of nonverbal imagery in cognition, something that behaviorists at the time did not view favorably (see Bartlett 1995, 215). The nature of the format of schema representation was not addressed until after the cognitive revolution, through which preference for propositional representation over pictorial representation continued to be the preferred format.



Within the domain of music, the schema concept has been instrumental in forging an embodied and situated cognitive approach to musical understanding. In contrast with other cognitively inspired theories of music, like those developed from universal grammar approaches (e.g., Lerdahl and Jackendoff 1983), Galant schema theory provides a more grounded understanding of mental representation as formed from human interaction with the environment, situated in specific places, times and cultures (see Byros 2009). While Galant schema theory has proven to be a useful heuristic for conceptualizing situated knowledge structures and accompanying processes—habit responses, normative procedures, sets of features, musical behaviors<sup>53</sup>—it is as yet limited in its systematic ability to make claims about the content, structure, and function of representation. In particular, Galant schema theory in its current format has yet to account for systematic differences in representation resulting from expertise that, I argue, should lead to observable effects in perception and cognition.

This issue occurs primarily due to the limitations of adopting schema theory from the cognitive sciences, which, as just discussed, posits the acquisition of amodal abstractions, problematically collapsing representation across modality and system (verbal/nonverbal). The effects of such a perspective are evident in Galant schema theory in several ways. Notably, the collapse across modality is demonstrated through the invocation of visual, score-based corpus analysis techniques to gather Galant schemata category features. Such features, garnered through visual analysis, are considered equivalent to those acquired through listening. Visual corpus analysis is considered to be a ‘metaphor for experience,’<sup>54</sup> such that the categories and features extracted from it represent those that are acquired by encultured, ‘ideal’ listeners (see Bourne 2015, 34). This problem is not unique to Galant schema theory but applies to other theories that

---

<sup>53</sup> See Meyer (1956, 62; 1994, 220) and Gjerdingen (1988, 6; 2007, 16), respectively.

<sup>54</sup> See Byros (2012, 278), Gjerdingen (1996, 280-281), Sears (2016, 114).

use reductional score techniques (particularly those inspired by Schenkerian analysis, see Lerdahl and Jackendoff 1983; Rabinovich 2019).

A similar invocation has historically been made in the domain of language learning, even in the constructional approach to language acquisition. Even though the constructional approach is embodied (i.e., that a language unit has an association with a learned non-verbal behavior or context), the language units themselves are still construed as amodal abstractions (see Tomasello 2005; Abbot-Smith and Tomasello 2006). Within a DCT approach, language representations are formulated from *multiple* modality specific units: any language unit is constructed out of tightly interconnected auditory, visual, and motor logogens, not a single abstracted representation. Productivity and abstraction in language is viewed from a DCT perspective as fluency in intermodal transfer between representations of different modes and systems, essentially allowing for logogens and imagens of different modes to be instantaneously substituted for one another (as discussed in the previous chapter). Any claim, therefore, that score analysis yields representational units and features that are commensurate with those possessed by encultured listeners is inaccurate. I argue that there do exist representational differences between encultured listeners and experts in schemata representation—primarily in the modal and system (verbal/nonverbal) level distribution of such representations—and that my current framework as proposed here is able to illustrate these. While there may certainly be overlap between the units and features garnered from score analysis with those acquired through auditory statistical learning alone (i.e., as is the case with encultured listeners), there will also be differences that have real implications for cognition. In the same way that someone who can speak fluently but cannot read or write may share some representational units with someone who is fully literate (e.g., auditory logogen /*kat*/), they do not possess the representations for word parts in the visual

and motor modalities (i.e., visual and motor logogens *C*, *A*, and *T*), nor the ability to rapidly exchange such visual, motor, and auditory logogens. I will show that the latter type of representational format, which I call structured distributed representation, provides distinct benefits to simulation ability. Therefore, in contrast to pure schema theory, the framework presented here is able to capture and discuss those representational differences that afford particular cognitive abilities to experts that encultured listeners do not possess.

### Novice versus Expert Representation: Loose versus Structured Representation

The primary argument of this section is that Galant schema representations will differ between encultured novices and experts as a function of the types of interactions both groups have with a schema. Therefore, both the representation of a schema held in LTM, and the types of cognitive affordances using these representations, i.e., ability for simulation in WM, will differ between groups. From the memory expertise perspective invoked here, it is understood that memory skill will vary both between groups and individuals, stemming from systematic differences in LTM organization and access.

To highlight these potential differences between novices and experts, I will draw on Barsalou's invocation of loose versus structured representation (Barsalou 2003a). From the DIPSS perspective, 'natural' category learning results in the acquisition of loosely organized property and relation simulators, guided by the directing of attentional resources to particular features of a category and relationships between them. Here I propose that attention selects for different features, resulting in loosely interconnected (lower probability) perceptual symbols (imagens) within a category simulator that can be accessed during simulation. As I will show in the case of novice Galant schema representation, such loosely structured simulators are limited in

their modality and system-type encoding, resulting in a memory network that is *less* distributed over representational codes—i.e., simulators are primarily constructed from associated auditory and interoceptive imagens in the nonverbal system, with limited referential connections to logogens in the verbal system. Such loosely organized simulators result in a limited ability with, and control over, simulation; representational codes are less accessible via executive control and are primarily available through direct representational activation.

As expertise is developed, the number of property and relation simulators grows, meaning that representation of categories becomes more distributed over different code types. The activational pathways between representational codes also becomes more probabilistic—i.e., more structured—meaning that one representation can functionally act as a retrieval cue for another. This increases the probability for co-activation of different representations, which impacts an expert's ability to actively use these representations during simulation. From this perspective, not only does the representational base expand, but the ability of experts to fluently access different representations, including multiple representations concurrently, resulting in an ability to interpret multiple regions or features of a category at the same time (Barsalou 2003a, 1182, 1184). With increased information accessibility also comes higher levels of cognitive control over information access, which results in a more dynamic system that supports skilled interaction with a category across different contexts (Barsalou 2003b, 546). In the case of music theoretic expertise, I will show that music theoretic concepts are explicitly used to produce highly distributed and tightly interconnected simulators to represent Galant schema categories: music theoretic concepts are used to guide particular interactions with schemata categories, directing attention toward different aspects of the interaction (i.e., features, or internal states). Music theoretic concepts therefore act as an explicit means to acquire property and relation

simulators, distributed across systems (verbal, nonverbal) and modality (auditory, visual, motor, interoceptive).

### **What is a Galant Schema? The “Conceptual Peg” Account**

Before diving into an account of schema representation, it will be necessary to provide a definition of Galant schemata from the DCT framework implemented here. In the music theoretic literature, Galant schemata are discussed as multidimensional categories, defined primarily by typically occurring scale degrees in the upper and lower voices which are accompanied by particular harmonies. In ‘prototypical’ schemata, scale degrees and accompanying harmonies co-occur at the same temporal locations, which are also typified by specific metric placements and grouping structures (see Figure 3.1). These co-occurrences result in the conception of each joint temporal location functioning to mark a schema stage, the length and presentation of which varies between schemata of the same category. Category membership therefore occurs along a gradient, as these features can differ noticeably. Despite this, certain versions of a schema are considered more ‘prototypical’ (i.e., more frequent, or more central in a distribution) than others in particular contexts. For example, the Galant Romanesca is considered more prototypical for the Galant period (1720-1790) than is the sequential Romanesca, which was the more standard presentation in the 17<sup>th</sup> century (Gjerdingen 2007, 27). Category membership is understood to fluctuate over time by repertoire, meaning that category representation would likely have also shifted over time in encultured populations given differential exposures (see Gjerdingen 1988; Byros 2012a,b).

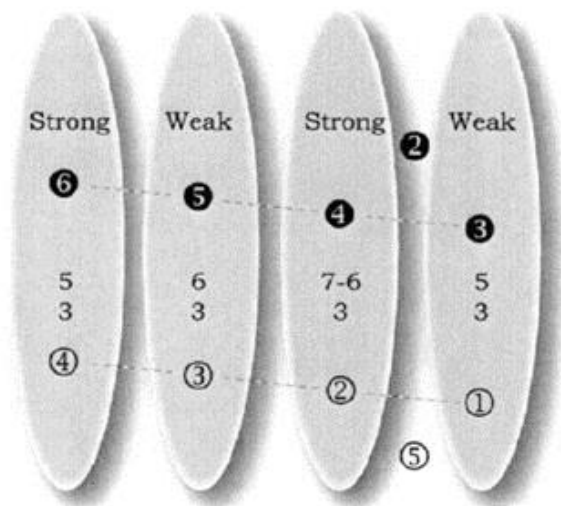


Figure 3.1. Schema Stages for Prinner Prototype (graphic from Gjerdingen 2007, 455)

From the DCT perspective invoked here, Galant schemata—at least extracted as ‘natural categories’—are imagen representations stored in the auditory mode. Rather than information being extracted from such interactions and placed into an abstract representation, what is captured instead is an exemplar representation for each piece that is encountered. Because scale degrees, harmony and metric placement tend to co-occur, each point that marks the beginning of a schema ‘stage’ represents a high imagery point—a form of conceptual peg that may come to define the beginning of an episodic chunk in memory, and one onto which subsequent sequentially organized information is attached. Such schema stages offer points for concrete interactions with the category: in each modality, they facilitate directed attention, ensuring that these points are stored in LTM. In concurrence with music theoretic concepts, each point offers multiple types of interactions that will become encoded and associated to represent the category.

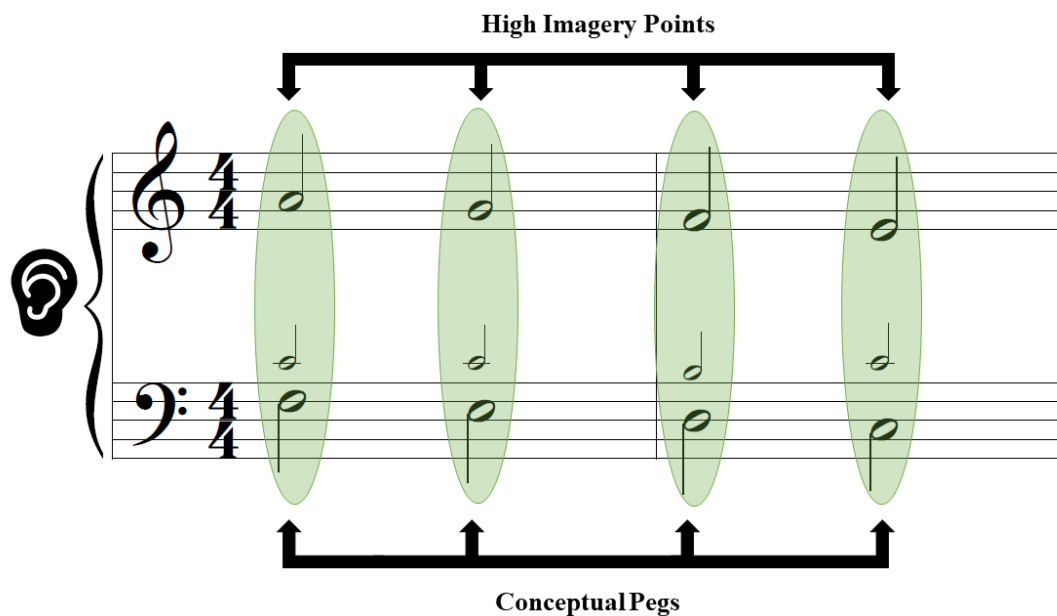


Figure 3.2. Prinner Schema Stages as Conceptual Pegs in DCT

### Loose Holistic Representation: The Novice

Galant schemata are generally understood to be natural categories: meaningful patterns understood by encultured listeners that are extracted through statistical learning—in other words, categories that do not need explicit learning (see Gjerdingen 2007, 16; Byros 2009). This work suggests that the primary features of a schema—scale degrees, harmony, metric placement—are extracted from exposure alone, and are comparable to those features explicitly learned by experts, whether composers or music theorists. Contrary to this, I suggest that the representations acquired and used by experts differ significantly from those of encultured listeners. For encultured listeners, simulators in LTM are more holistic and loosely organized—they are holistic because acquired simulators do not explicitly represent different properties of a Galant schemata (e.g., harmony, scale degrees), but instead capture a holistic gestalt of the experienced perceptual/cognitive state of the listener. The simulators themselves are also loosely organized, meaning that the distribution of representations across modality is limited (i.e., primarily

auditory and interoceptive), and the probability of connections between these representations is lower, stemming from a more limited set of interactions that listeners have with schemata compared to experts who interact with these patterns in more consistent and varied capacities (e.g., visually, haptically, verbally, etc.). As a result, the types of cognitive processes—or simulations—available with loosely connected simulators is highly restricted. Below, I will outline loosely structured representation of schemata in encultured listeners by providing an account of statistical learning from the current DCT perspective. Following this, I will illustrate the limitations in simulation ability given such LTM structure.

### **On Representation: An Account of Statistical Learning**

Given that an encultured listener would acquire Galant schemata ‘naturally’ through ‘passive’ exposure alone, the primary means through which this would occur is by statistical learning. Statistical learning is defined as the automatic extraction of statistical regularities from exposure to various stimuli. The primary area in which statistical learning has been examined is language learning. There are several components to statistical learning that are understood to extract different types of information. Firstly, conditional learning provides information about transitional probabilities—sound combinations within versus between syllables. Secondly, distributional learning provides information related to word boundaries (i.e., chunking) as well as information related to syntax, including between-word (word-pairs) and other non-adjacent probabilities (e.g., sentence structure) (Thiessen, Kronstein, and Hufnagle 2013). For exemplar accounts of statistical learning, conditional and distributional processes are viewed as resulting from common processes of memory: conditional learning is a function of exemplar storage, and distributional is a function of memory integration—the comparison and revision processes that likely occur when a novel exemplar is encountered (Thiessen and Pavlik 2013; Thiessen and



Erickson 2013; Thiessen, Kronstein, and Hufnagle 2013; Erickson and Thiessen 2015). Therefore, statistical learning can be viewed as an inherent feature of memory (storage, integration, and forgetting, see Thiessen 2017), rather than a feature of a separate system or cognitive mechanism. This account of statistical learning agrees with my framework here that views many different processes, perceptual and cognitive, as functions within a unified, embodied system.

Also important to statistical learning are its inherent limitations and constraints—particularly as related to the modality of the stimuli presented, and in the amount of learning that can occur from exposure alone. Some research suggests that rather than being a domain-general mechanism that works similarly across modality (auditory, vision, etc.), statistical learning operates slightly differently in different modes, resulting in the storage of mode-specific information to each respective system (Frost *et al.*, 2015). Along with inherent storage differences of the human brain, statistical learning also differs based the input stimuli structure—for example, whether the stimuli are sequentially (e.g., auditory) or spatially / synchronously (e.g., visual) organized (Krogh, Vlach and Johnson 2013). This is congruent with the approach taken here that emphasizes the modal-distribution of representation rather than centralized, amodal representation. Such processes are also affected by the state of learning (particularly the level of cognitive resources available to the learner) and by previous exposure which can affect feature extraction and attention (*ibid.*). While particularly valuable for acquiring implicit knowledge about statistical regularities, statistical learning is inherently limited in its ability to provide explicit access to previously learned exemplars.

Within the framework presented here, encultured listeners acquire Galant schemata through auditory statistical learning (i.e., listening), which results in the acquisition of loose

holistic simulators. Such simulators do not explicitly represent different properties of a Galant schemata (e.g., harmony, scale degrees), but instead capture a holistic gestalt of the experienced perceptual/cognitive state of the listener. What are encoded in the case of encultured listeners are complete exemplars of individual pieces in the auditory system, along with other distributed, albeit modally limited, representations encoding the context in which the music was encountered, such as motor imagens (e.g., from dancing or movement), vocal imagens (e.g., active singing along, subvocal imitation), internal bodily states (e.g., interoceptive imagens associated with affect), and/or visual information (e.g., visual cues in a live music context). I will begin by detailing an account of statistical learning within the auditory imagens for Galant schemata, followed by a demonstration of the formation of distinct interconnected representations in Galant schema simulators in encultured listeners.

Auditory representations acquired by an inexperienced listener may look something like the following (see Figure 3.3). The chunk size of auditory imagens in LTM may be quite large, and some information in the trace may be degraded or even lost. As one becomes an encultured listener, more exemplars are acquired, and statistical learning mechanisms store and integrate regularities across these exemplars into LTM. In the case of Galant schemata, as more exemplars are integrated into memory, features common to a given category may become enhanced in memory through information revision. When learning a new exemplar, the structure of the LTM trace will be affected by previous learning. For the encultured listener, representational chunk size may be more similar to experts, and regions common across categories (e.g., outer voices) may be enhanced in the auditory imagen trace (see Figure 3.4). Given that attention may be directed more towards melodic attending, particularly at the opening of the exemplar in Figure 3.4, that information may also reflect enhanced encoding relative to other information, such as

lower frequency range (bass part) of the auditory trace. From a statistical learning perspective, conditional probabilities (information regarding features that occur together through time) are represented by the auditory information within each chunk, while distributional probabilities (information regard non-adjacencies) are mostly available through the occurrence of chunks within the exemplar. The availability of distributional information across many exemplars reflects the understanding of Galant schemata ‘constructions,’ that is, the “meaning” of a schema (chunk) as frequently co-occurring with another.<sup>55</sup>

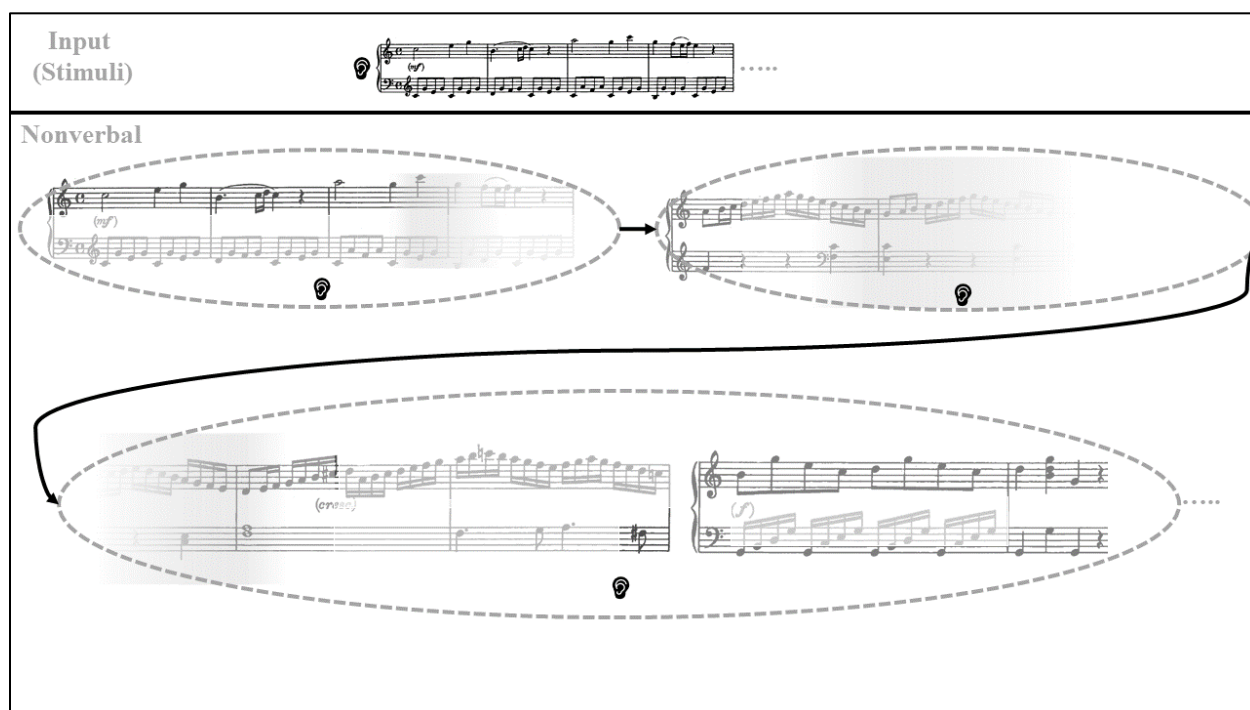


Figure 3.3. Auditory Imagen Organization of K. 545 for a Novice Listener

<sup>55</sup> In the current example (Figure 3.4), the construction available is one of Prinner schema-as-riposte.



Figure 3.4. Auditory Imagen Information of K. 545 for an Encultured Listener

Simulators may store more than auditory information, however. The simulator will also include some modally distributed traces capturing contextual information, such as concurrent visual, motor and interoceptive states during the interaction (see Figure 3.5). Here, interoception in the form of entrainment is shown by projective arrows, which changes over the course of the excerpt (two equal projections at the opening, to alignment with the bass rhythm in the second phrase). Other modal information is also captured, including an instance of subvocalization in the first two measures, which further enhances the melodic portion of the auditory trace. During the second phrase, there is an additional motor trace reflecting the tapping of a foot concurrently with listening. It is important to note the unidirectionality and low probability of these additional traces: it is very unlikely that entrainment or foot tapping (alone) would be a sufficient cue for the auditory trace in the simulator. In this example, the only way to activate the auditory portion of the simulator is through direct representational activation, not through association in the

nonverbal system within the simulator. Figure 3.6 shows the hypothetical LTM structure for the same exemplar given that a live performance was viewed. This would result in additional visual traces in the simulator, reflecting perhaps the fact that the listener had focused primarily on the pianists' hands and body position during listening. This results in the association of smooth hand gestures with the first phrase, and more rapid movement with the second.<sup>56</sup> If such visual representations were acquired from a particular performance (in episodic memory), they may demonstrate some bidirectional connection, such that seeing a recording of the visual performance without audio may activate the auditory imagen through association (and vice versa).

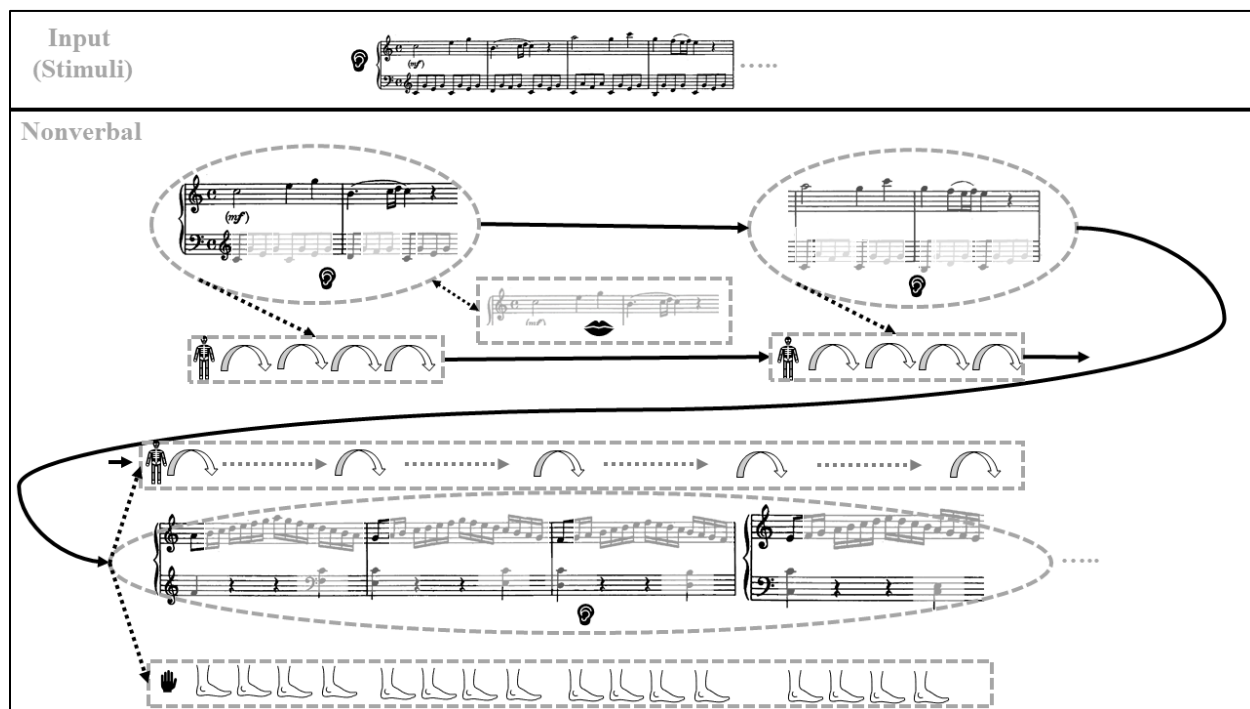


Figure 3.5. Modally Distributed Simulator for K. 545

<sup>56</sup> Such additional traces may facilitate metaphoric interpretations of the excerpt, such that the listener might agree with the description of the first phrase as more floating, and the second as more driving.

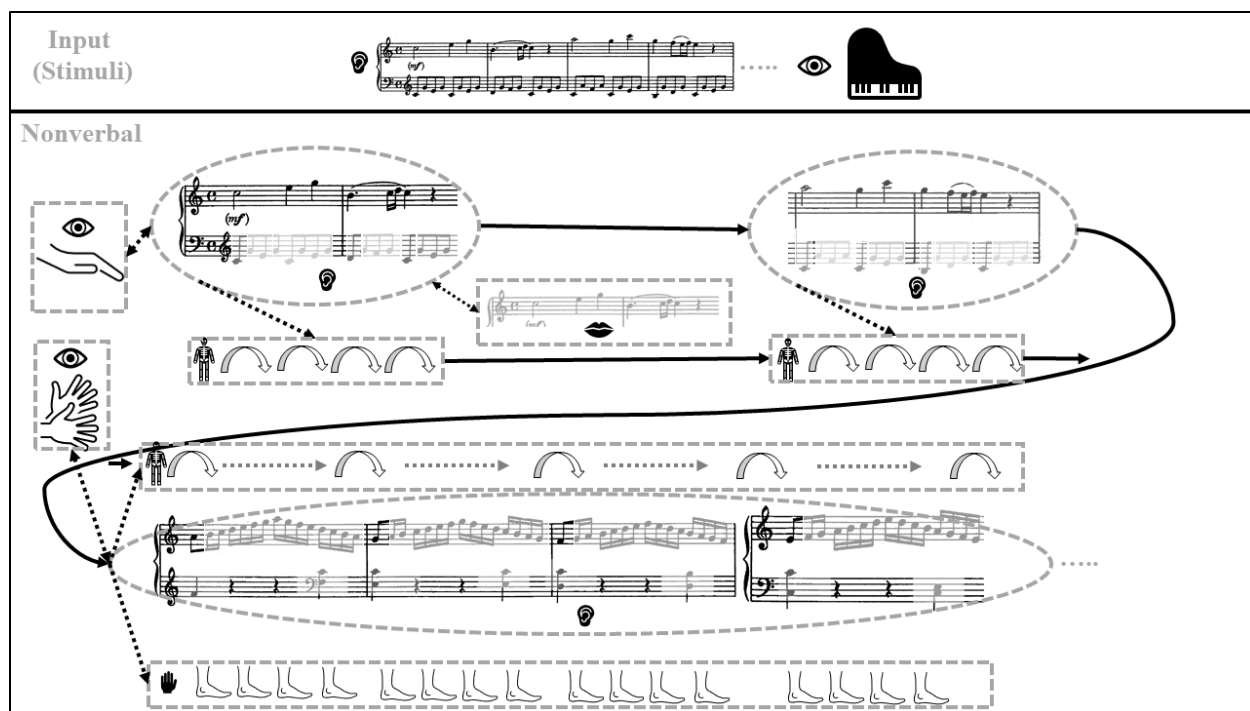


Figure 3.6. Modally Distributed Simulator for K. 545 (Live performance)

An encultured listener who does not have explicit knowledge of Galant schemata categories would also have a very limited set of referential connections to logogens within a simulator. For example, they may only have logogens for the composer and piece name (see Figure 3.7). The referential connections across verbal and nonverbal systems are likely bidirectional but may differ in strength of probability. For example, the logogen for the composer and piece name may have a higher probability connection to the auditory logogen chunk at the opening of the piece compared to the chunk that occurs in the second phrase. In this way, if given the verbal prompt for the piece, one would be more likely to imagine the opening than bar five. However, given an auditory prompt of the second phrase at bar five, a familiar listener would likely be able to identify the piece through referential processing.

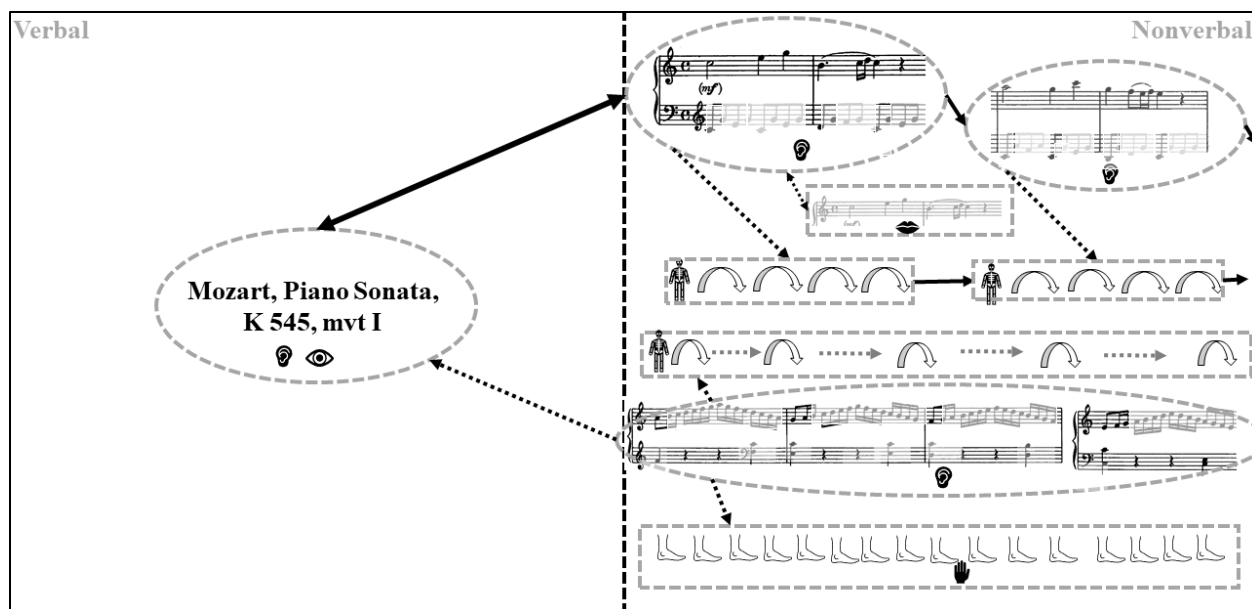


Figure 3.7. Simulator for K. 545 Including Referential Connections to Verbal System

Lastly, other conceptual information important to Galant schema are represented in the LTM information of the encultured listener, including formal function and spatial positioning (e.g., sonata form location). From a music theoretic perspective, theorists have proposed that other conceptual information, such as regarding sonata form, is accounted for by schema embedding. That is, while schemata for Galant patterns emerge from learning, so do separate schemata for forms, which are embedded into one another (see Byros 2015, *Haupttruhpunkte des Geistes*). As the encultured listener is understood here, their knowledge regarding function and form is represented not in different abstractions for these different concepts, but instead comes from similarities in the content and organization of simulators across exemplars for individual pieces. Here, the musical surface is preserved in a listener's representation, so textural aspects of formal function, such as grouping, fragmentation, harmonic rhythm, etc., are maintained in the exemplar trace. Formal knowledge is represented by the general size and ordering of auditory imagen chunks in memory (i.e., spatial location over time). Representation

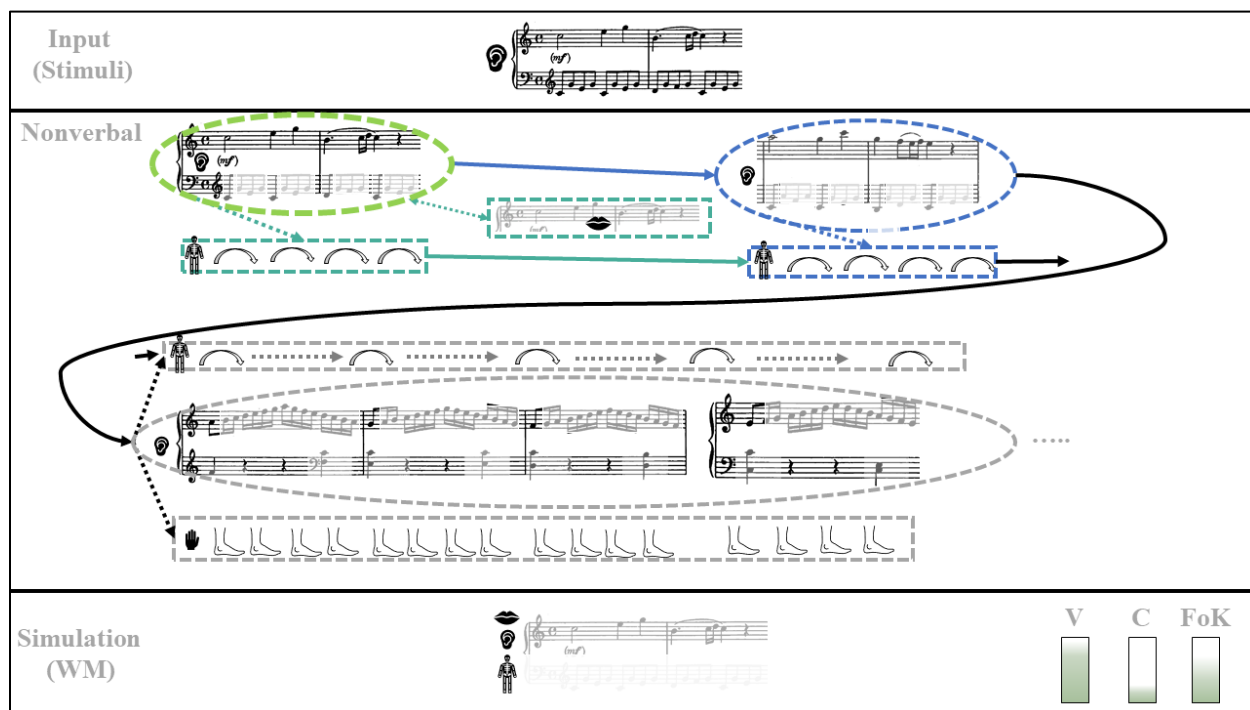
in encultured listeners is therefore relatively centralized: information regarding Galant schema and their usage is primarily confined to exemplars stored as auditory imagens. Given such centralized, holistic representation, access to traces is inherently limited—particularly with regards to ‘features’ of Galant schemata such as harmony and scale degrees. Due to the limited ways in which encultured listeners can access and use these loosely organized simulators; they are similarly limited in simulation ability—the capacity to actively conceptualize and use information stored in LTM. In the following section, I will detail such limitations in simulation for the encultured listener.

### **On Simulation: Activational Pathways and Limitations**

Given the loosely interconnected and modally limited representational base possessed by the encultured listener, simulation—the ability to actively use information stored in LTM in WM—is inherently limited. This is due primarily to limitations in information access (activation of imagens and retrieval mechanisms), as well as limitations in WM maintenance. Firstly, information in LTM is primarily available through representational activation of auditory traces in the nonverbal system. Associational and referential activation within a simulator is restricted due to the limited number of traces within a simulator, and their lower probabilities of connection (see Figure 3.8a, b). While indirect activation is still certainly possible (e.g., imagery without and external prompt), this activation will primarily occur through activation of auditory imagens in the simulator.



(a). Representational activation of auditory imagen and associational processing plus priming



(b). Representational activation of motor imagen with no associational processing

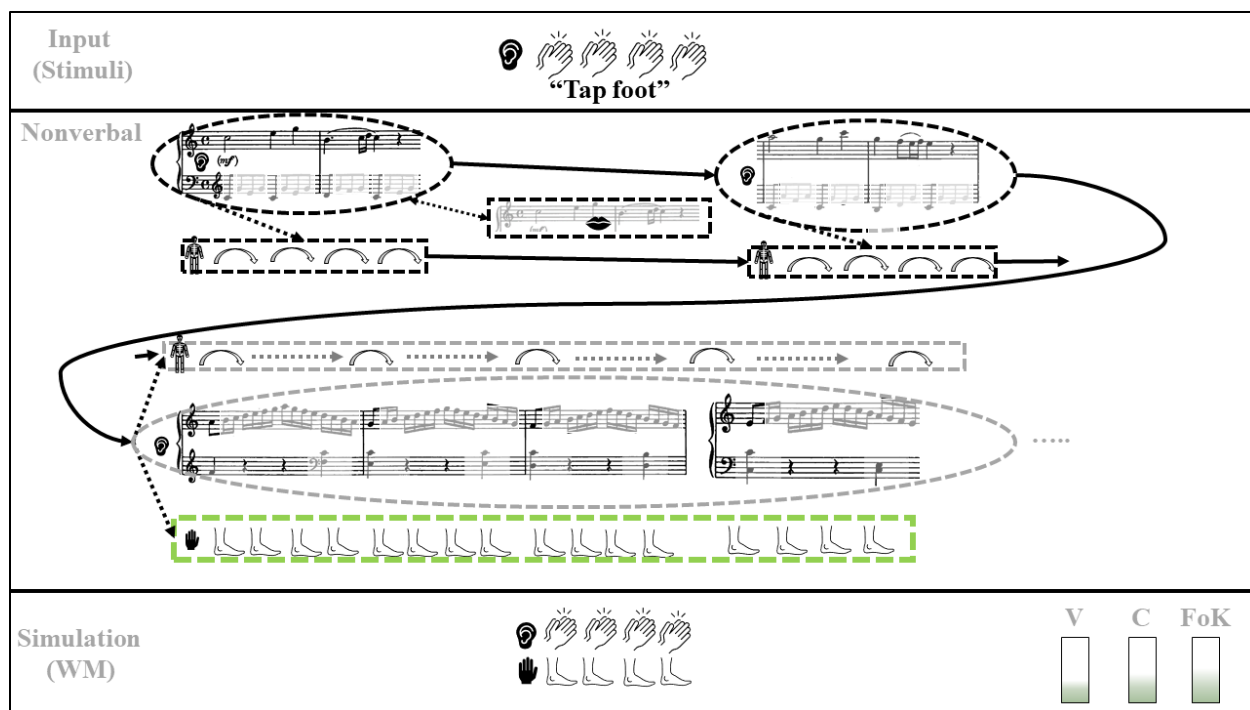
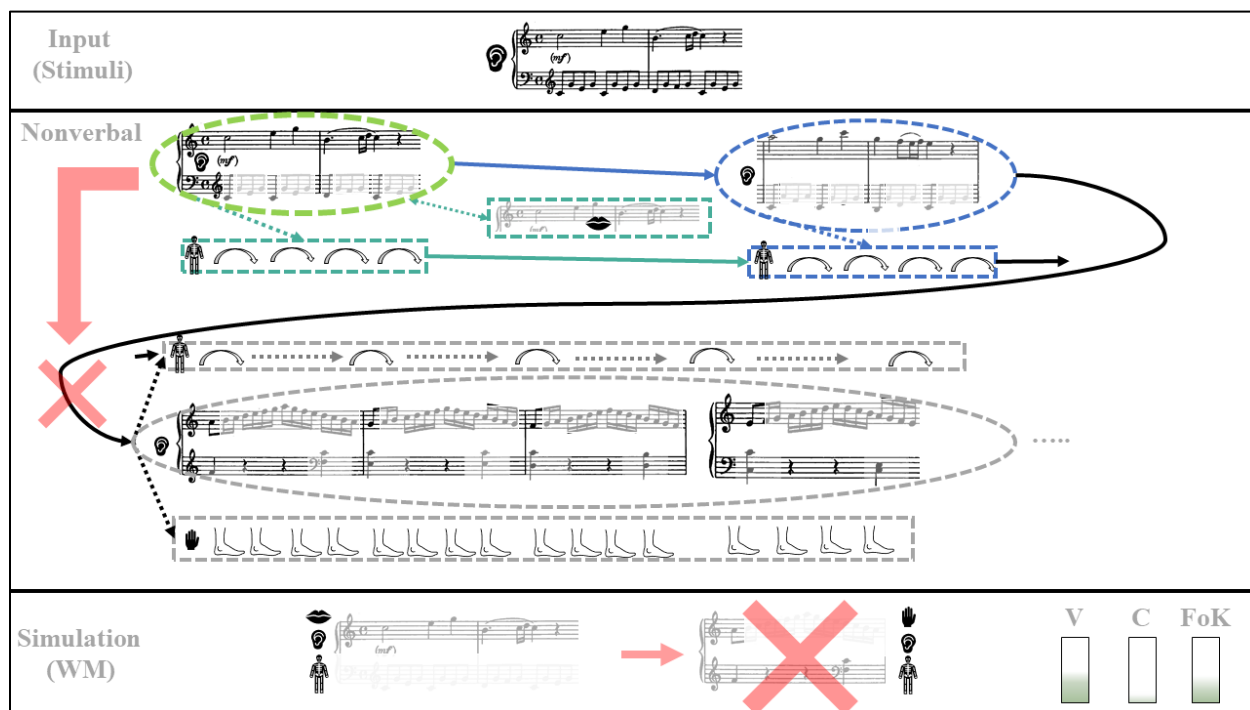


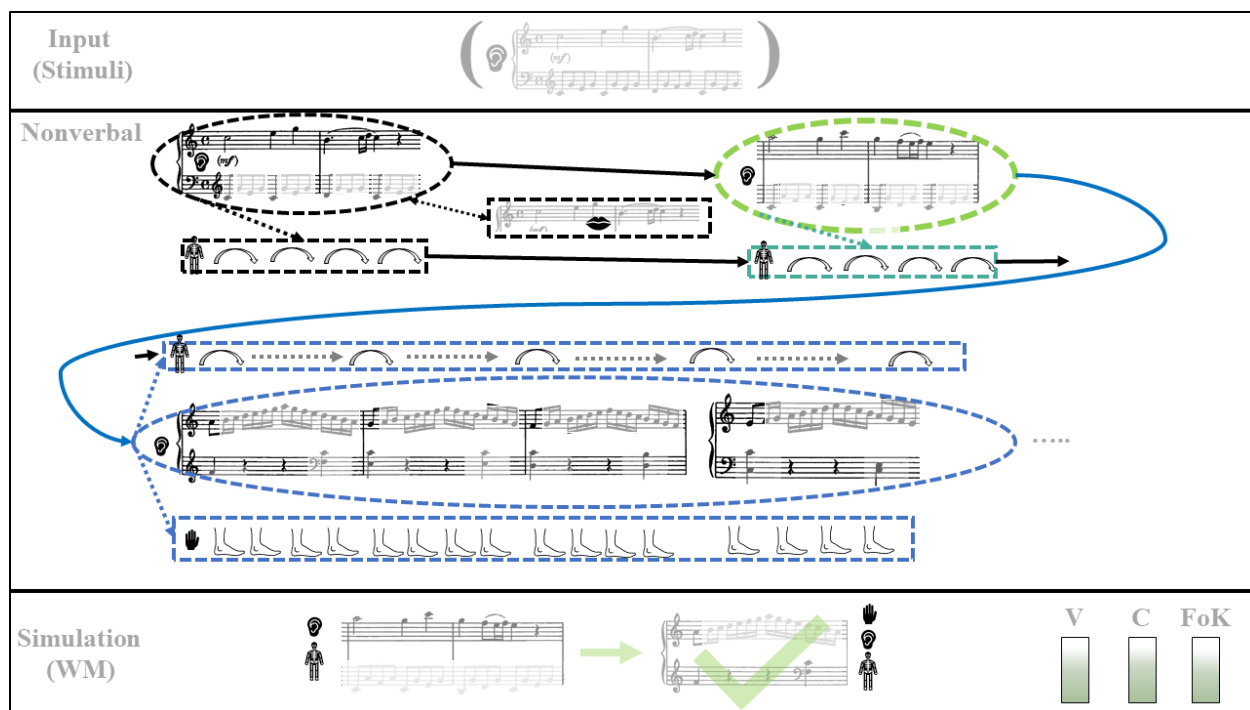
Figure 3.8. Direct Representational Activation of Auditory Trace (a) and Motor Trace (b)

Secondly, retrieval mechanisms within a simulator are limited: the limited distribution of traces and the unidirectionality of most of the traces in the simulator means that retrieval would primarily come through representational activation of auditory traces. Once activated, the encultured listener may have a limited capacity to access different parts of the simulator. The strict sequential organization of auditory traces limits the ability of an encultured listener to access information. For example, despite the added detail in the trace, an encultured listener may not be able to skip across chunks: while the opening of the piece may be available to imagery, they would be limited in their ability to ‘skip over’ a chunk to access another one without first passing through the intermediary chunk in WM (see Figure 3.9a, b). The chunk would be available out of sequence if activated through direct representational activation (Figure 3.9c). Similarly, access *within* an auditory chunk is limited: in order to access particular event in the middle of a chunk, the chunk would need to be cued from the beginning and scanned across in auditory WM to the point of interest (see Figure 3.9d).

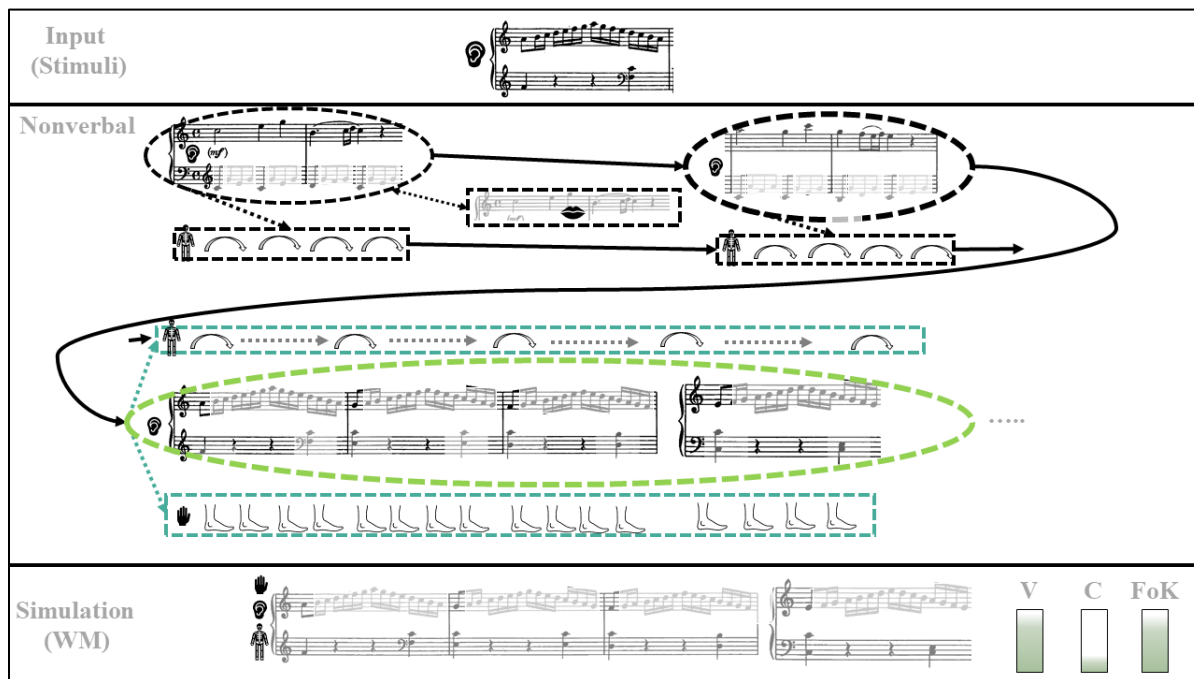
(a). Unable to jump



(b). Chunk access through sequential associational activation between auditory imagens



(c). Chunk accessible through representational activation



(d). Part of chunk available through sequential maintenance in WM

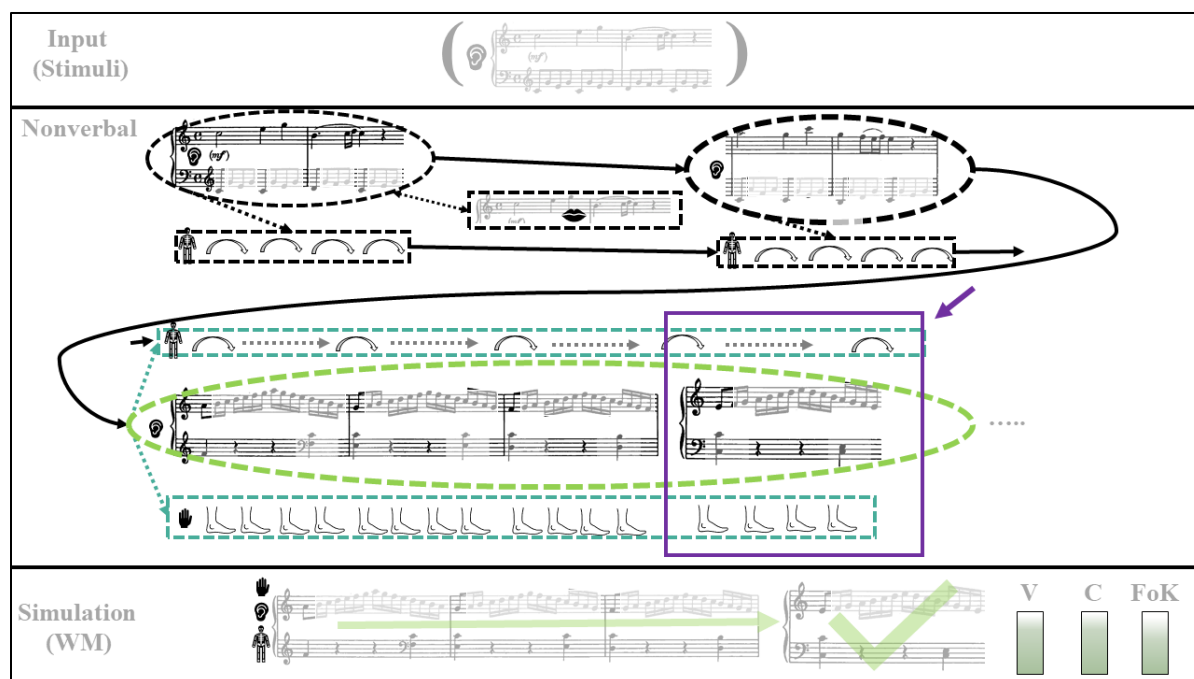


Figure 3.9. Chunk Access During Simulation (Imagery). Unable to jump (a), Sequential Associational Activation (b), Direct Representational Activation (c) and Sequential Maintenance (d)

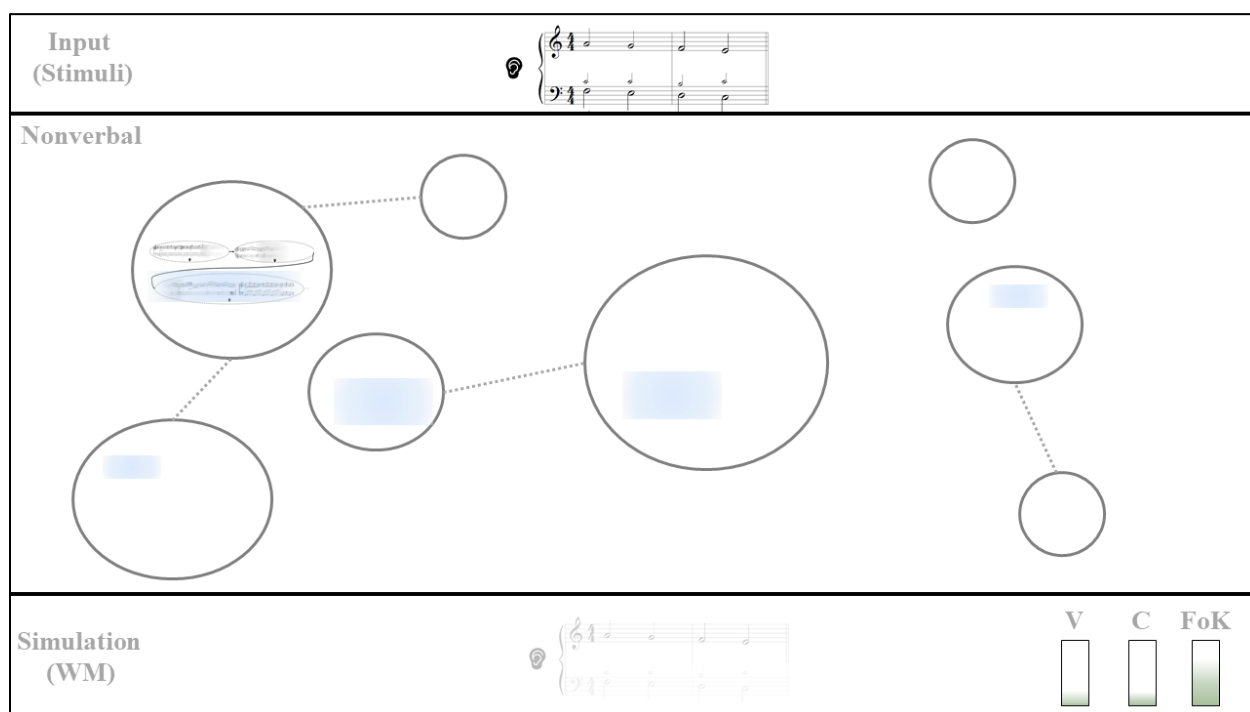
Lastly, simulation in WM is more challenging due to the inability of the encultured listener to exchange a representational unit for one in another mode to avoid overloading WM. Recall that in the DCT account of reading, fluent readers will exchange visual logogens for auditory ones during reading to avoid overloading visual working memory (Sadoski and Paivio 2014, 99). Because the primary representational units of the encultured listener are auditory imagens, they cannot perform such an exchange, and therefore risk overwhelming auditory working memory during simulation.

For the encultured listener, simulation is primarily based around simulators for exemplars. As discussed above, category-level simulations are limited; additionally, the lack of distributed representation limits the possible spreading activation within and across simulator pools. To reframe this, encultured listeners possess more recognitional abilities than identification or categorization ones. and would be able easily perform category *responses* and *evaluations* in reference to representational activation (presented stimuli) but are more limited in their abilities to simulate category decisions apart from this.

Figure 3.10 shows the differences between category level activation in the nonverbal system given a new exemplar for a category between a novice (3.10a) and encultured listener (3.10b). For the novice listener, who may have substantially fewer exemplar simulator pools stored in memory, a newly presented exemplar activates the similar portions of the existing traces. With fewer simulators in memory, the novice listener may only have a weakly held image held in memory very briefly after exposure, which may have relatively low vividness and a moderate FoK, particularly if the presented exemplar primes traces that exist in memory. Comparatively, the encultured listener will have many more exemplar simulators stored in LTM, which will result in a larger portion of exemplar pool simulators becoming active during

representational activation. Because of the higher number of memory traces, increased connectivity between pools, and increased detail in each of these traces, the effect of sensory information held in WM after exposure to the stimuli will result in a simulation that is higher in vividness and higher in FoK. No control of simulation is required for this type of recognition as the information in WM is not being manipulated or accessed.

(a). Novice Listener Category Activation



## (b). Encultured Listener Category Activation

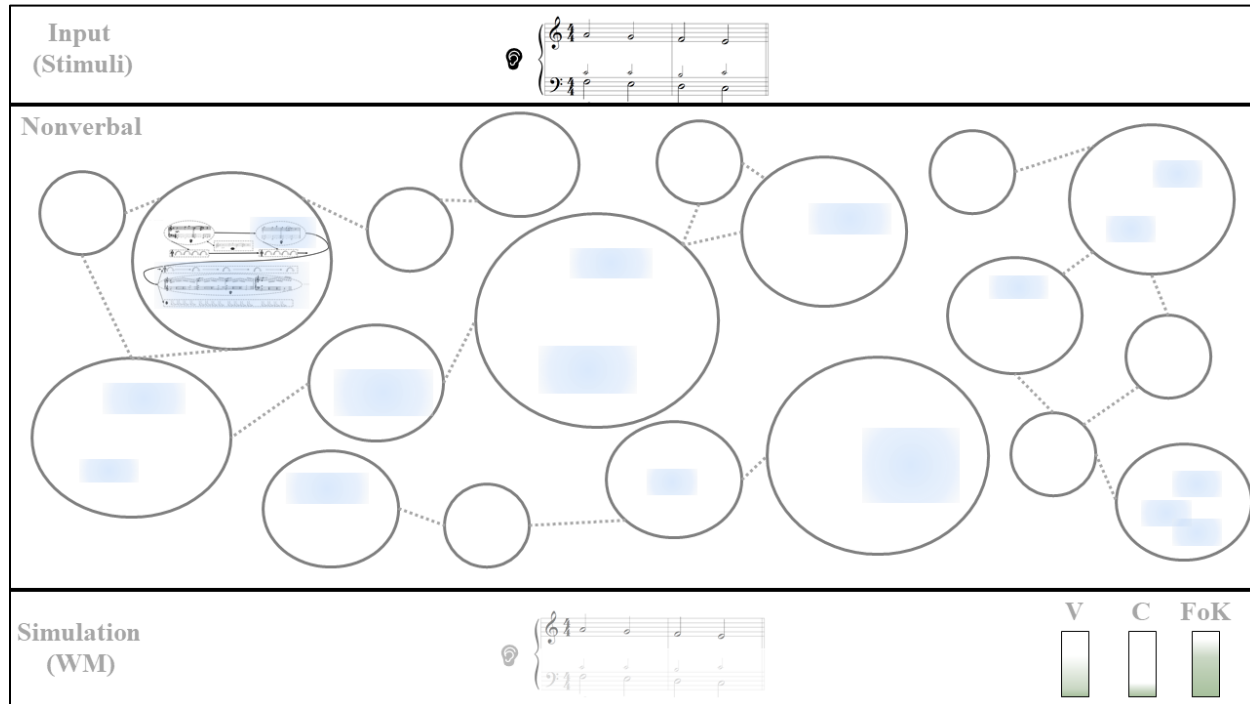
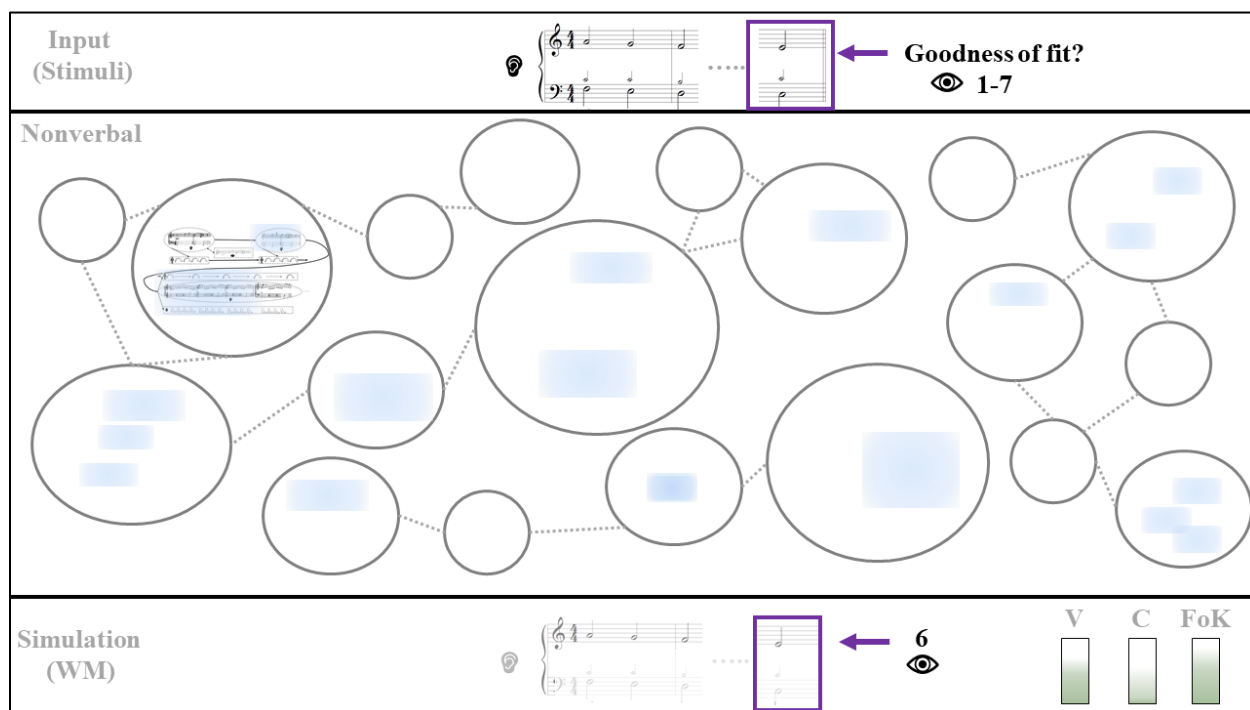


Figure 3.10. Category-Level Activation of Exemplars in Exemplar Pools Through Representational Activation

An encultured listener will be quite able to perform basic evaluations of schemata categories given representational activation of the existing traces in memory. For example, an encultured listener would likely have enough category-level representational information to properly evaluate and provide an appropriately higher goodness-of-fit rating for the terminal stage of a new exemplar for a Prinner schema (Figure 3.11a), and a lower goodness-of-fit rating for a presented variation (Figure 3.11b). Similarly, if asked to imagine the terminal stage of this familiar schema pattern given representational activation, the encultured listener would likely have no issues in imagining and maintaining a continuation in WM (Figure 3.12a). However, if then prompted to imagine an alternative stage in place of the one just imagined, the encultured listener may have more difficulty (Figure 3.12b). Here, the listener would be required to

selectively access a separate subset of simulator pools. Due to the limited variety of traces and retrieval mechanisms in their simulators in LTM, they may be unable to retrieve and instantiate an alternative option in WM. In sum, the structure of LTM traces affords the encultured listener with the ability to easily simulate once auditory imagens have been activated through representational activation. They are, however, more limited in simulation ability apart from this due to the limited distribution of representation in LTM: have a more difficult time in retrieving and maintaining different simulators in WM during simulation. They are therefore quite fluent in recognition of schemata but are limited in simulation ability at the category level due to limitations in spreading activation characteristic of categorization in LTM.

(a). Goodness of fit for terminal stage in Prinner schema





(b). Goodness of fit for variation of terminal stage ( $\hat{6}$  in the bass voice)

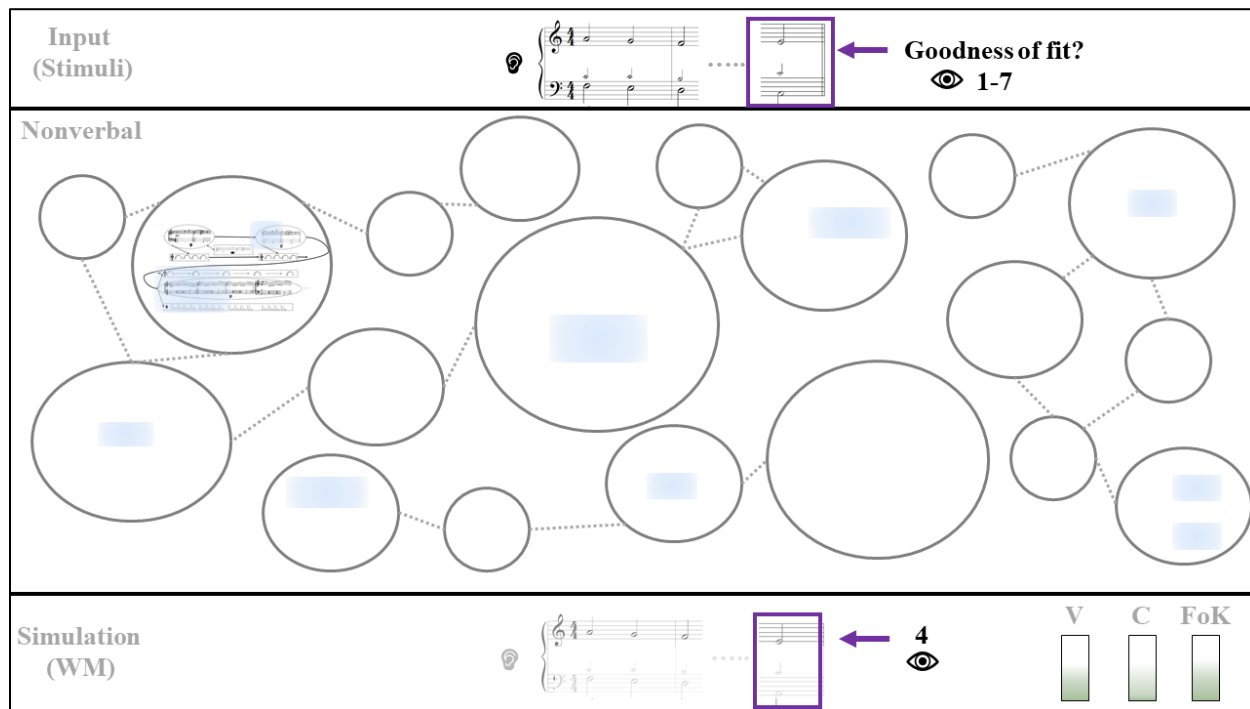
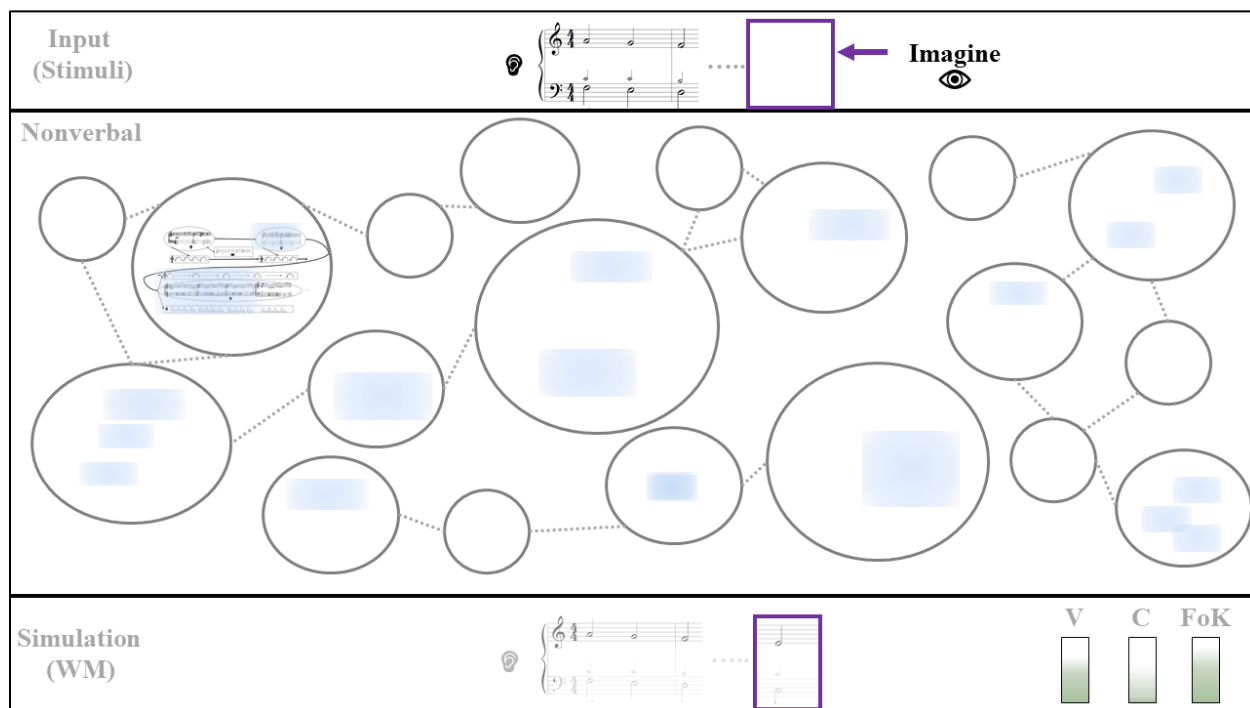


Figure 3.11. Goodness of Fit Ratings for Prinner Schemata and Variations

(a). Imagery completion for Prinner



## (b). Imagery completion select alternate

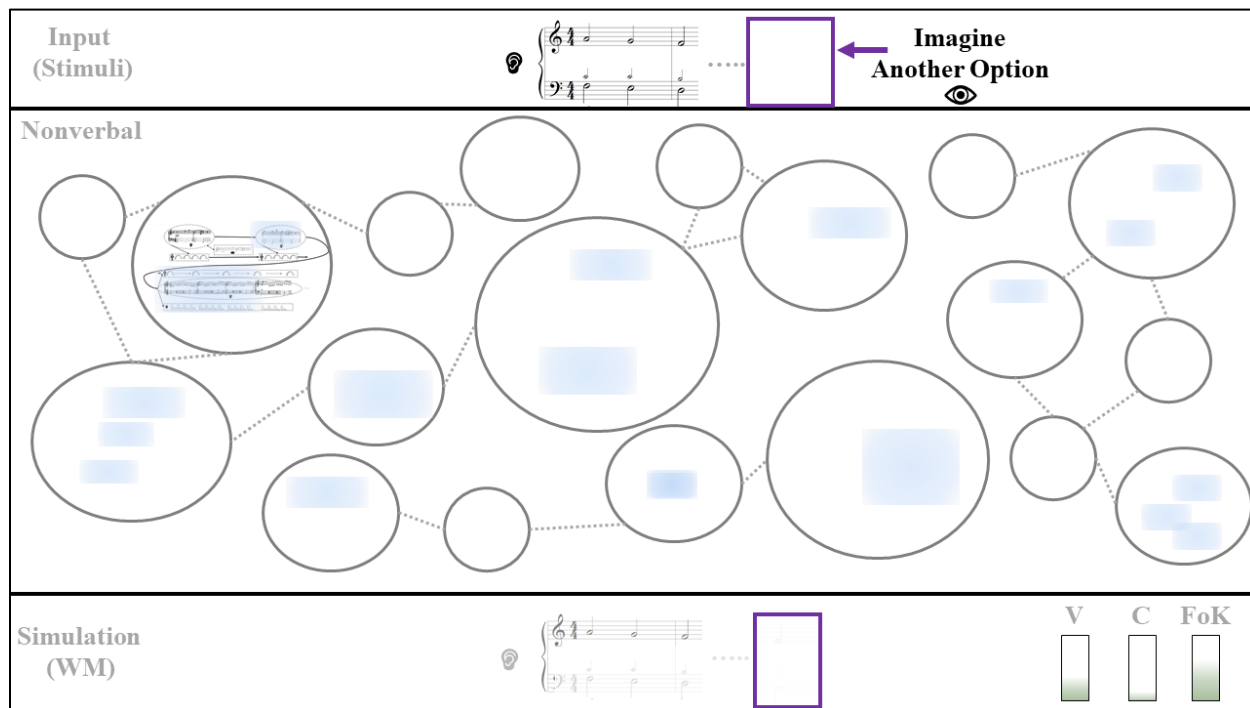


Figure 3.12. Imagery Completion for Original Prinner and Alternate

## Structured Distributed Representation: The Expert

In contrast, expert representation differs because experts explicitly acquire a larger, more modally distributed base of simulators to represent Galant schemata categories. That is, features and relations of these categories are explicitly represented by a distributed set of traces which are more probabilistically interconnected (i.e., structured). Such representational structure provides more fluency in retrieval and maintenance during simulation. Expertise with music theoretic concepts functions to direct attention towards particular features and relations, ensuring that they are encoded and stored in LTM. In this way, music theoretic concepts *are* simulators: they help to focus attention during particular types of interactions with music, encoding the object of attentional focus into separate yet associated traces across verbal and nonverbal systems to form

larger networks of modally distributed representational pools. Given this larger, more modally distributed set of simulators, music theorists can more fluently simulate Galant schemata categories in a wider range of contexts compared to encultured listeners. They are able to flexibly access and maintain different types of information from LTM in WM and use this information for various types of tasks. Explicitly encoding and relating multiple parts of a category allows for flexibility in identification and categorization (see Casale *et al.* 2012).

In this section, I will begin by outlining the representational structure of a Galant schema category (Prinner) for a hypothetical music theoretic expert. To do this, I will first discuss music theoretic concept as simulators and how they enable encoding of properties and relations of Galant Schemata. Secondly, I will provide an example of the verbal-associative network for the Prinner as an example of declarative knowledge which aids in the bootstrapping of category learning. I will then outline how such verbalizations are referentially attached across to the nonverbal system to form simulators that represent Galant schemata. In the second section, I will briefly discuss the benefits this LTM structure has on active simulation in WM.

### **On Representation: Music Theoretic Concepts as Simulators**

Music theoretic concepts afford a variety of different ways of interacting with music, particularly as they use a wider array of modalities than listening alone (visual, auditory, motor) and function across systems (verbal, nonverbal). Music theoretic concepts serve to focus attention to a particular feature and/or set of relations, ensuring that these aspects of the interaction are encoded into a separate trace in LTM. While this may ensure that particular features are encoded in LTM, they also, importantly, divert attention away from others, resulting in the formation of highly structured simulators for certain features and rather loosely structured simulators for others. For example, music theoretic concepts are typically focused on aspects of

pitch structure and spatialization (e.g., form), which de-emphasize other aspects of musical information (e.g., texture, timbre). These features are still encoded within musical interactions but are typically not explicitly acquired in the same way as pitch-related information for example, and are therefore represented by fewer, more modally restricted traces.

Here I will maintain Barsalou's distinction between property and relation simulators (Barsalou 2003a). While all music theoretic concepts, even 'object-based' ones, are relational in nature (e.g., scale degrees reference their relative positionality), concepts that are property simulators focus attention more on a particular 'object' in perception / imagery, while concepts that are relation simulators direct attention toward relations without the need to capture the specific objects in question. The case is the same with everyday concepts. For example, the concept EYE contains a strong association to the relation ABOVE because eyes often occur above noses and mouths in the context of faces; thus, the concept EYE *encodes*, through association, the relation ABOVE through its relation to other objects. Simulating the concept EYE will also likely include this relation, although not as the attentional focus of the simulation. Rather, it would be 'background' for that simulation, or part of its context, Barsalou 2003a, 1181. However, simulating the concept ABOVE can easily occur without the concept EYE, because this relation occurs between many different objects.

This invokes the distinction made previously between concrete and abstract concepts. While concrete concepts are grounded in specific imagen representations of those objects, abstract concepts, such as ABOVE, are more grounded in interoceptive and affective states of the perceiver. Abstract concepts captured by relation simulators are grounded in internally available states where attentional resources are focused *more* onto these states than on particular objects themselves. While such relational concepts are initially acquired through concrete concepts, over

time they eventually gain independence, particularly through the use of language (Gentner and Asmuth 2019), to which I will now turn. The concepts I will focus on are those most commonly used in Galant schema theory, including scale degrees, harmony and counterpoint (see Gjerdingen 2007; 2020). Such concepts were also used in the past by persons engaged in what scholars call the core pillars of conservatory training—*solfeggio*, *partimenti*, and counterpoint (see Sanguinetti 2012, 42; Baragwanath 2020, 288). I will also discuss some concepts anachronistic to the Galant period—harmonic function and formal function—as they are commonly used by modern theorists when interacting with schemata categories (see Caplin 2015).

#### Representing declarative knowledge in the verbal system

The cooperative interdependence of verbal and nonverbal systems lies at the heart of DCT. However, the importance of language for the bootstrapping of learning, and categorization in particular, cannot be overemphasized. More nameable features or cue words for categories improves categorization performance (Fotiadis and Protopapas 2014, Zettersten and Lupyan 2020). Verbal priming improves identification of color (Forder and Lupyan 2019), plausibly through referential priming explained in DCT. Language is vital for categorization development in young children (Ferguson and Waxman 2017), and eventually provides the cognitive scaffolding required for more abstract thought, such as analogical and relational reasoning (Gentner 2016).

Developing music theoretic expertise entails with it the development of a large body of verbal representations, used both actively during theorizing and analysis, and in the communication of findings to others. Galant schema theory, like other theories, also relies extensively on music theoretic concepts, such as scale degrees and chords, to detail features and

relations important to different schema categories. Such concepts function to explicitly encode properties and relations into LTM across verbal and nonverbal systems, and also as a means for fluently retrieving and instantiating different representations from LTM into WM. Verbalization provides an efficient means for maintaining and accessing memory-burdensome representations (e.g., auditory imagens of music), which facilitates control over the depth of simulation, lessening the burden on WM resources. Even those trained in the traditional partimenti systems had extensive verbalization practice in the form of solfège (Gjerdingen 2020, Baragwanath 2020). Below is an example of such a verbal network for the Prinner schema (Figure 3.11) showing the encoding of verbal information available from *Music in the Galant Style* (Gjerdingen 2007) (Figure 3.12). Because of the depth of verbal expertise available to theorists working from the book, the modes of the logogens are assumed to all be present (i.e., visual, verbal, spoken, heard). A more accurate depiction of such a network of verbal associations would show separate representations for each mode (i.e., visual, verbal, spoken, auditory for the logogen “Prinner”). For space reasons, logogens will only be shown once, with each mode assumed for each one.

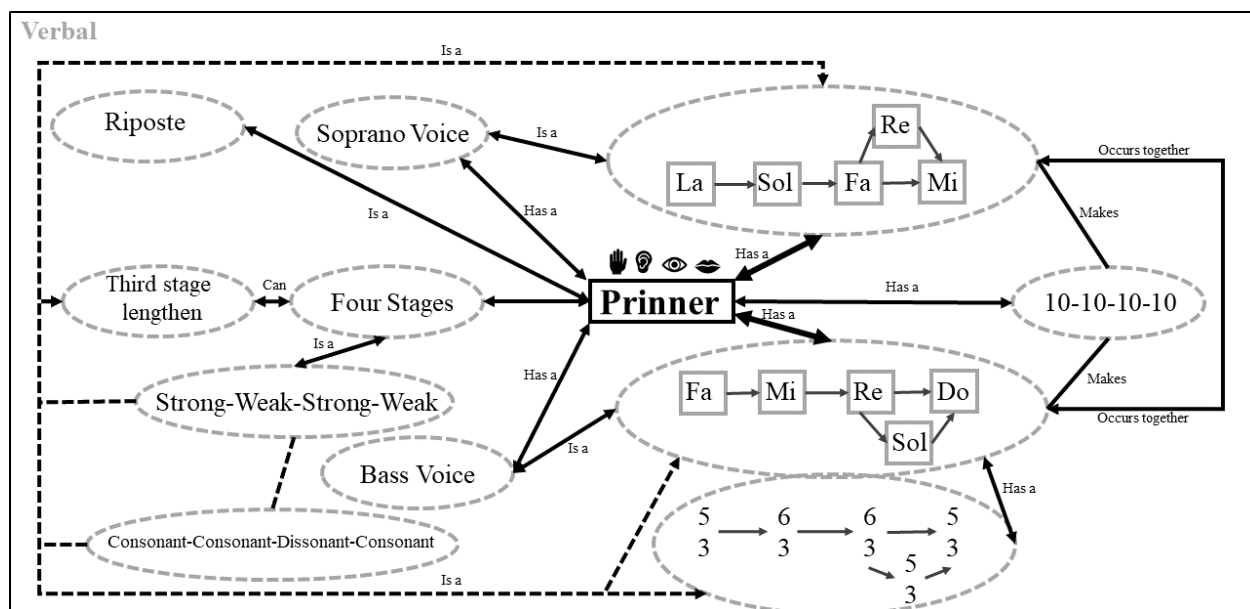
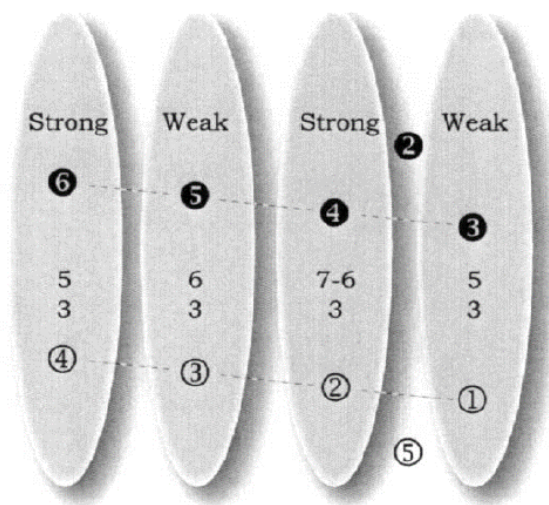


Figure 3.13. Verbal Associations for Prinner Concept



#### Central Features

- Four events presented either with equal spacing, with an extended third stage, or in matching pairs.
- In the melody, an emphasis on the stepwise descent ⑥-⑤-④-③ (to effect a stronger cadence, a high ② is often inserted before the final ③).
- In the bass, an emphasis on the stepwise descent ④-③-②-① (to effect a stronger cadence, a ⑤ is often inserted before the final ①).
- A sequence of chords in  $5/3$ ,  $6/3$ ,  $6/3$ , and  $5/3$  positions. The third stage is often dissonant, while stages one, two, and four are consonant and in the same mode.

Figure 3.14. Prototypical Features of Prinner Schema from Gjerdingen (2007)

## Property simulators

Property simulators focus attention on a particular component of experience, primarily for the attended-to feature of a category. The LTM representation stores the feature in attentional focus *in* the mode of interaction. If the mode of interaction is multimodal (i.e., visual and haptic), separate highly associated traces will be added in each modality to reflect the interaction.

Property simulators can come in many forms, encoding different features of a stimuli (e.g., pitch structure, rhythm). I will detail two types of property simulators used in Galant schema theory which are central to encoding in LTM: scale degrees and harmony. Scale degrees are the primary property around which Galant schema categories are organized, so much so that they have been referred to as ‘scale degree schemata’ (Temperley 2006). The traditional approach to harmony in the Galant schema tradition is figured bass, but I will outline both figured bass and roman numerals here as they are, in practice, used interchangeably (although I will show that they afford different associations in LTM).



*Scale degrees.* Scale degree encoding lies at the heart of Galant schema cognition and is vital to the encoding and relation of many features of these categories. Scale degrees offer a means to anchor attention to a single point of focus in a complex auditory scene, allowing one to encode the perceived ‘gestalt’ of a chord progression into separate parts that move together—essentially transforming a holistic whole into comprehensible and graspable sections. However, it is very important to note that this is not due to auditory encoding alone; much of the cognitive power that scale degrees offer stems from encoding that occurs across visual and auditory modes, further ‘objectifying’ parts as separate entities from their experienced wholes. Scale degrees are therefore fundamental to conceptualization in Galant schemata learning, affording a network of concrete objects and their relations in memory. Auditory, visual, and kinesthetic representations of scale degrees are used in categorization behaviors. Cognition research supports the importance of scale degrees for pitch conceptualization, showing that musicians often use them as conceptual anchor points when identifying pitches (Letailleur, Bisesi and Legrain 2020). Much like a conceptual peg, scale degrees provide a means for grasping and interrogating various stimuli in perception, and function as anchor points for conceptualization in imagery. In Galant schemata, the primary scale degree lines encoded are outer voices, although other voices will be encoded within harmony simulators discussed below. Like other properties and features of categories, scale degree lines are likely organized probabilistically in dominance order such that certain lines (e.g., Fa-Mi-Re-Do) are more characteristic and strongly associated within a simulator for a given category (e.g., Prinner) than others (e.g., Do-Do-Ti-Do).

Here, I will detail the hypothetical encoding of scale degrees during various interactions with a Prinner schema ‘prototype’ presented in score format in isolation, much in the same way that schemata are often initially learned or presented in practice. Here I will show separate

interactions in the auditory and visual modes, which may indeed only occur early in an acquisition period. More experienced practitioners would likely interact in both modes simultaneously, as they view and audiate or view and play at the same time; here, for expositional clarity, I will show auditory and visual interactions separately. First, I will detail the encoding of the bass voice. The hypothetical music theorist has already encoded bass voice solfège (moveable *do*) on the verbal side and uses this verbalization to guide nonverbal interaction and encoding. When interacting with the score alone, visual attention is drawn to the lowest voice while the logogens for the bassline solfège activate. Each solfège syllable would be mapped onto a fixation point, denoting visual attention, which is stored in the resulting associated visual imagen in the nonverbal system (see Figure 3.13). Fixation points are placed in visual scenes where the eye rests momentarily, which permit information uptake and retrieval during visual categorization (Rehder and Hoffman 2005a, b; Blair et al. 2009; Kim and Rehder 2011). These are shown by X's in the figures below. Contrastingly, visual saccades—movement trajectories where the eye does not fixate, and where attention is not focused—are shown by blue lines across the fixation points, demonstrating the trajectory across each line. Information in peripheral vision (e.g., soprano and tenor voice) is still somewhat retained in the trace but shows substantial information loss relative to the bassline fixations. This demonstrates the availability

of the relative position of the bassline as ‘at the bottom’ of the visual scene; detailed information about the other lines are missing in this trace.

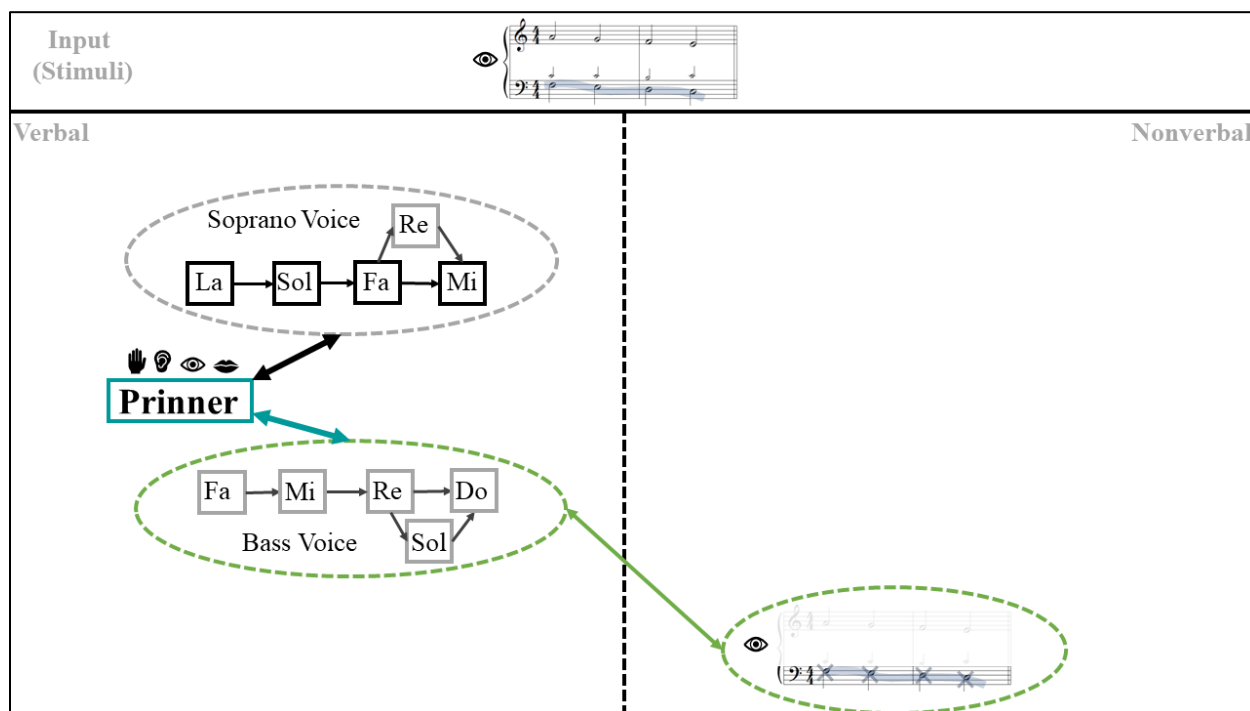


Figure 3.15. Encoding of Visual Imagen for Bassline Simulator

As a second interaction, the hypothetical theorist will encode an auditory trace from the visual cue, which strongly associates the auditory trace with the visual one, shown by the arrows linking the two together. Here, the auditory trace is created by the theorist singing the bassline using solfège (see Figure 3.16)<sup>57</sup>, which also ensures that the auditory imagen has a strong referential connection to the logogen for solfège syllables. The trace could also have been created through other interaction types involving audition, including playing the line on an

<sup>57</sup> The bass line imagen is represented here in the auditory mode as a the ‘absolute’ pitches (i.e., the representation stored in LTM is ‘perfect,’ such that the exact frequency information is retained). This is largely due to difficulties in visual depiction of auditory representation. As most of the population has relative pitch, the encoding of scale degrees lines into LTM is more relational (i.e., the distances between pitches are retained). Little is known regarding the exact mechanisms that differ between perfect and relative pitch—whether, for relative pitch possessors, absolute information is degraded in the trace—or if the differences stem from distinct recall mechanisms. While it is evident that some absolute information is retained in the general population (see Janata *et al.* 2002), relative pitch possessors’ representations would be different from the absolute representations depicted here.

instrument, or listening to someone else play or sing it. If these interactions were added in addition to a sung version (as they often are in practice), the simulator would contain more interconnected auditory image traces. It would also be possible to encode the bassline *in context* by listening to the excerpt with all voices while devoting auditory attention to the auditory stream for the bassline (see Figure 3.17). By creating multiple *separate* traces for the bassline alone, the music theorist ensures that this information is well encoded into memory, which can then be activated and utilized during stream segregation to enhance auditory attention. In sum, the more traces stored within a simulator, the more stable *and* flexible that simulator is for that feature because there will be a higher likelihood of retrieval of a relevant trace. Therefore, it would be most beneficial to store separate but associated traces for a sung, played and heard basslines for the Prinner category, assuring that this feature is accessible in multiple modes. This also allows for the modality of a simulation to be alternated to avoid overloading WM during simulation.

The same process is then repeated for the soprano voice, encoding separate traces for a visual scan and sung interaction, which are associated with the traces for the bassline on the nonverbal side (Figure 3.18). As several new traces are added for the soprano line, and auditory attention is guided towards the soprano during listening to the whole excerpt, the trace that contains the auditory representation for the heard example undergoes revision—the upper frequency portion of the trace gains more detail through directed attention during listening as well as through association with the other soprano traces in the simulator, which may come online during listening to help guide attention in stream segregation. All traces on the nonverbal side are associated with one another, particularly if such interactions are alternated (e.g., viewing and solmizing the bass, soprano, bass, etc.).

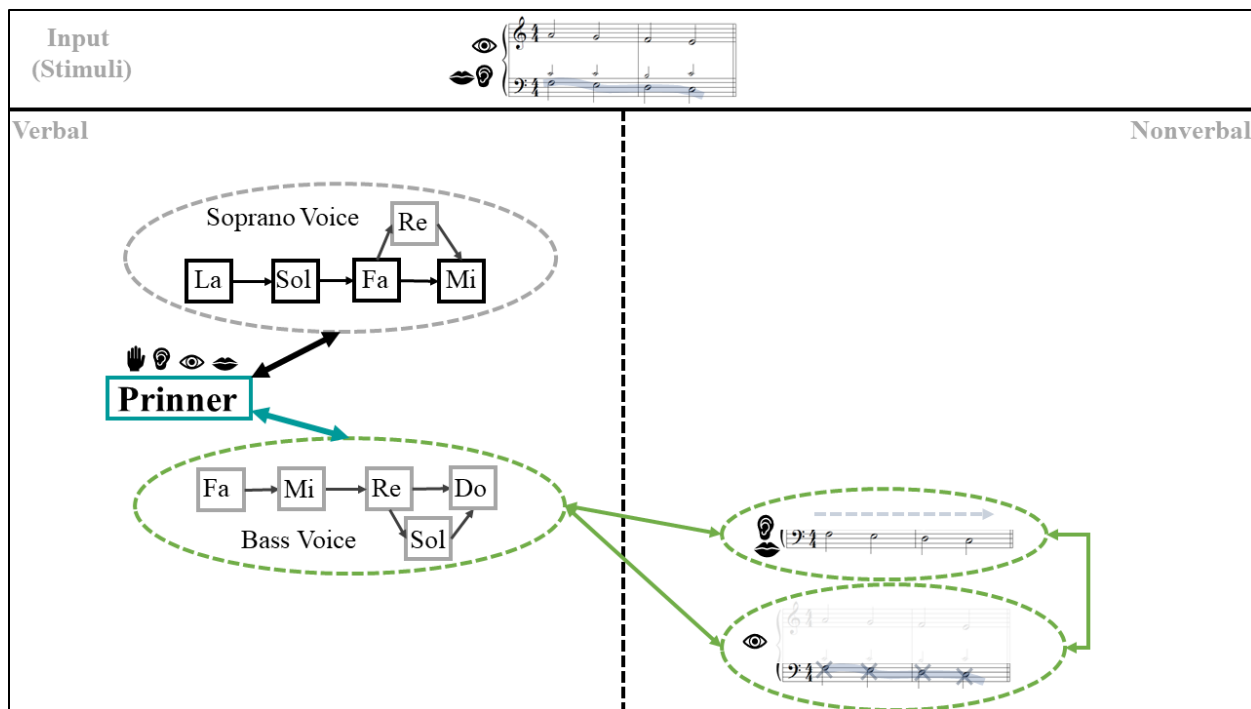


Figure 3.16. Encoding of Associated Auditory Imagen for Baseline

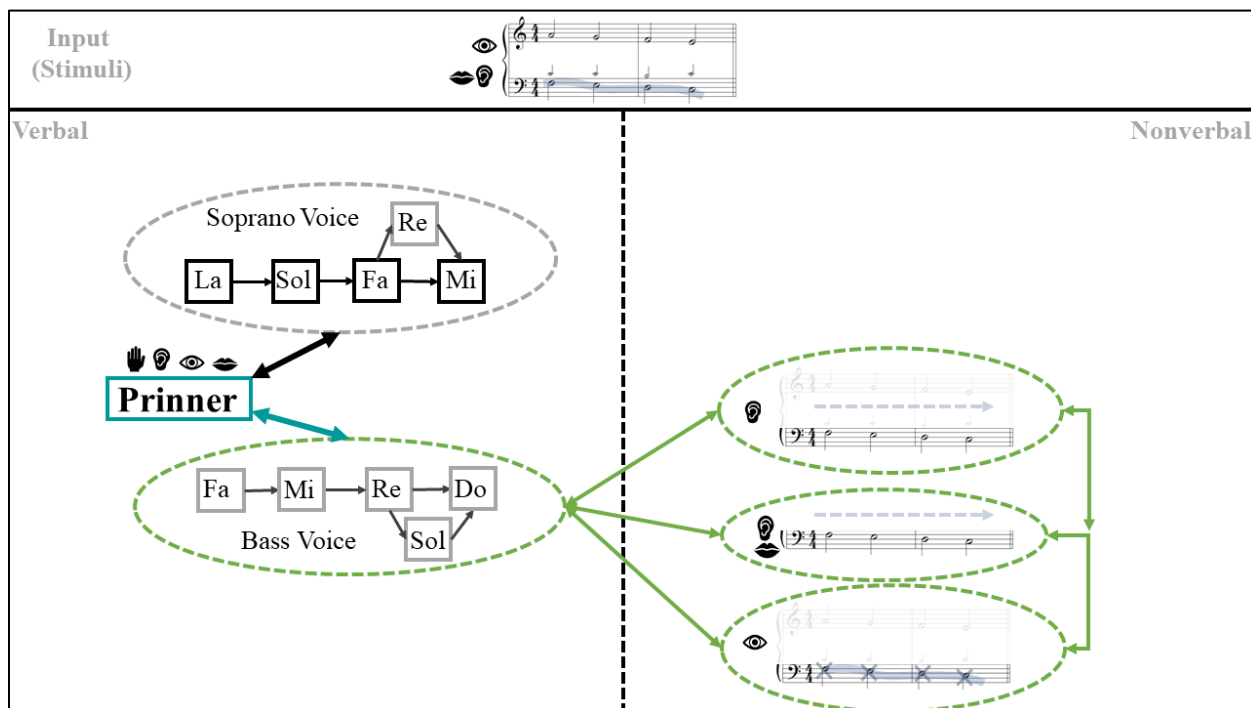


Figure 3.17. Encoding Auditory Imagen with Auditory Attention to Bassline

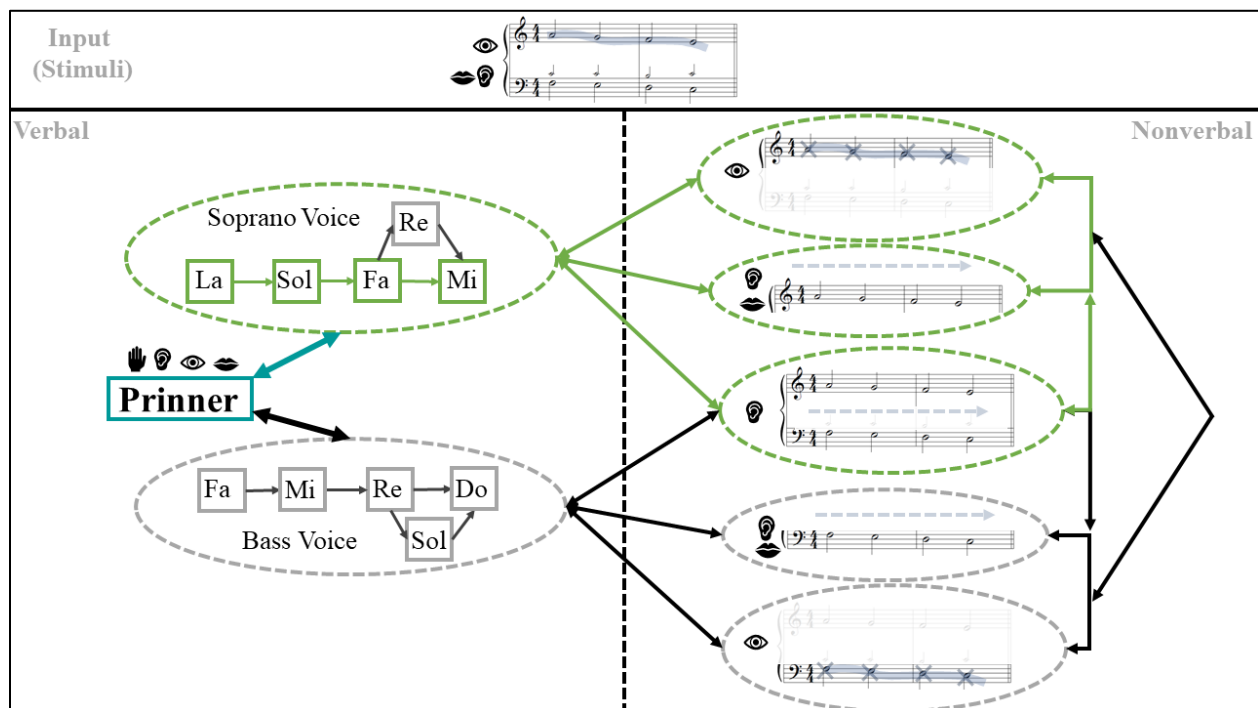


Figure 3.18. Encoding Process Repeated for the Soprano Line

*Harmony.* Harmony represents a more complex case for property simulation because harmony functions as a type of ‘background context’ for scale degrees. That is, harmony is a means of specifying the relations between different individual parts (e.g., voice leading in scale degree parts that go together), or between a focal point such as a scale degree in the bass and the pitches above it, as is the case with figured bass and Rule of the Octave (Byros 2009a, Byros 2012). “Harmony” was understood as figured bass in the pedagogical practices of the Galant era, and is still partially in use by the music theory community writ large today. As with learning intervals or counterpoint, it provides a means for encoding relationships between musical objects. For a Prinner schema, figured bass acts as a relation simulator which details intervals over a given lowest line: this is shown on the verbal side with figured bass numbers and the connection between them and the solfège for the bassline. (Numbers would be encoded as verbalizations, such as five-three, but for space reasons are shown here in symbolic format on the verbal side, as shown in Figure 3.19). On the nonverbal side, the representations directly connected to the figured bass are both visual and auditory, with the visual imagen showing visual scanning patterns beginning from the bass, then moving to the pitch above. As this scanning pattern would be common to other figured bass progressions, the association between that visual imagen and the verbalization for the figures on the verbal side is highly probabilistic. Contrastingly, the auditory imagen contains information specific to this particular example, namely the tonal position and quality of intervals present (i.e., object encoding from the solfège), such that the connection between that auditory imagen and the figured bass logogen is less probabilistic. That is, the figured bass progression used here, and its specification of the relation between pitches, applies to many other objects. The 5/3-6/3-6/3-5/3 relation does not specify a tonal position, much in the same way the interval string m2-M2-M2-M2 does not specify a tonal

location, although it may be highly suggestive of it. Thus, figured bass is a *relational* way of encoding harmony, one that requires an external contextualization via a given bassline.

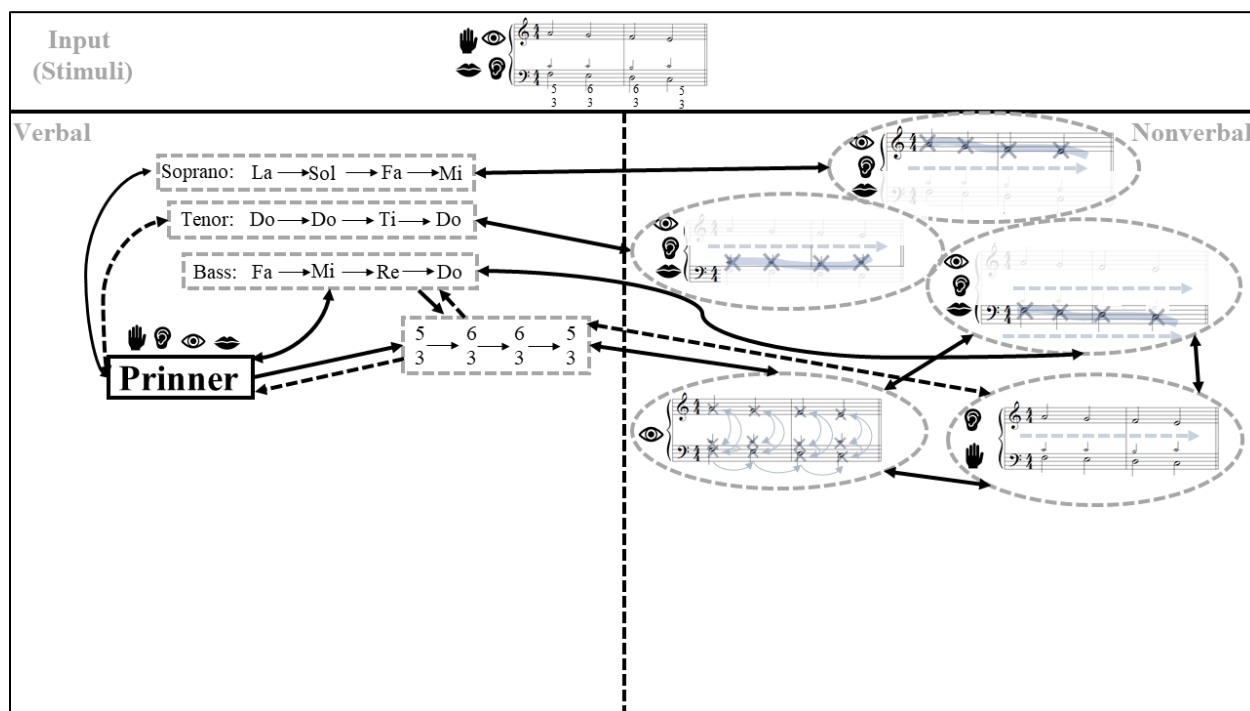


Figure 3.19. Figured Bass Harmony Simulator and Associations to Scale Degree Simulators

Another way that theorists conceptualize and encode harmony involves a combination of figured bass with roman numerals, which further specifies the tonal position of the progression (see Figure 3.20). Roman numerals provide a means of encoding harmony more as an object or feature than as a relation; while attention may be distributed across the stimulus, it is still directed towards the specific orientation and quality of each harmony. Here the roman numeral progression specifies both the relation (intervals above the bass) and the particular objects of that relation, viz. the solfège or pitches in their particular tonal position. Attention is therefore paid towards either the relations between *particular* scale degrees present in that context, or to their quality as a complete ‘object,’ such as might be the case with the IV chord. Again, using 5/3 to encode the harmony merely specifies a relation which could be seen in many different contexts.



Encoding using the roman numeral IV, however, places more attention onto the particular object, objects within the harmony, *and* their relation. Thus, harmony is a much more complex simulator that allows for access to the gestalt of a particular configuration of pitches or to the relations between particular objects such as scale degree voices and their paths of movement through time and tonal space. Given this complexity, a theorist will likely opt to encode harmony in multiple ways, such as adding in an arpeggiated vocal trace for the progression (see Figure 3.21). This encoding provides attentional focus towards the vertical relations between scale degrees, resulting in additional traces in the simulator. Importantly, when encoding in this manner, a new auditory trace is included that differs in structure from the heard auditory imagen. While the visual scanning pattern may not differ for the visual imagen, the auditory representation formed through singing an arpeggiated version of the progression is structurally different from previous traces. This type of encoding connects vastly different types of auditory representations, grounded through similarity in visual scanning patterns, such that harmony—sung as separate lines or arpeggiated parts, or played as holistic units—enables hearing events as conceptually similar despite having different auditory structures.

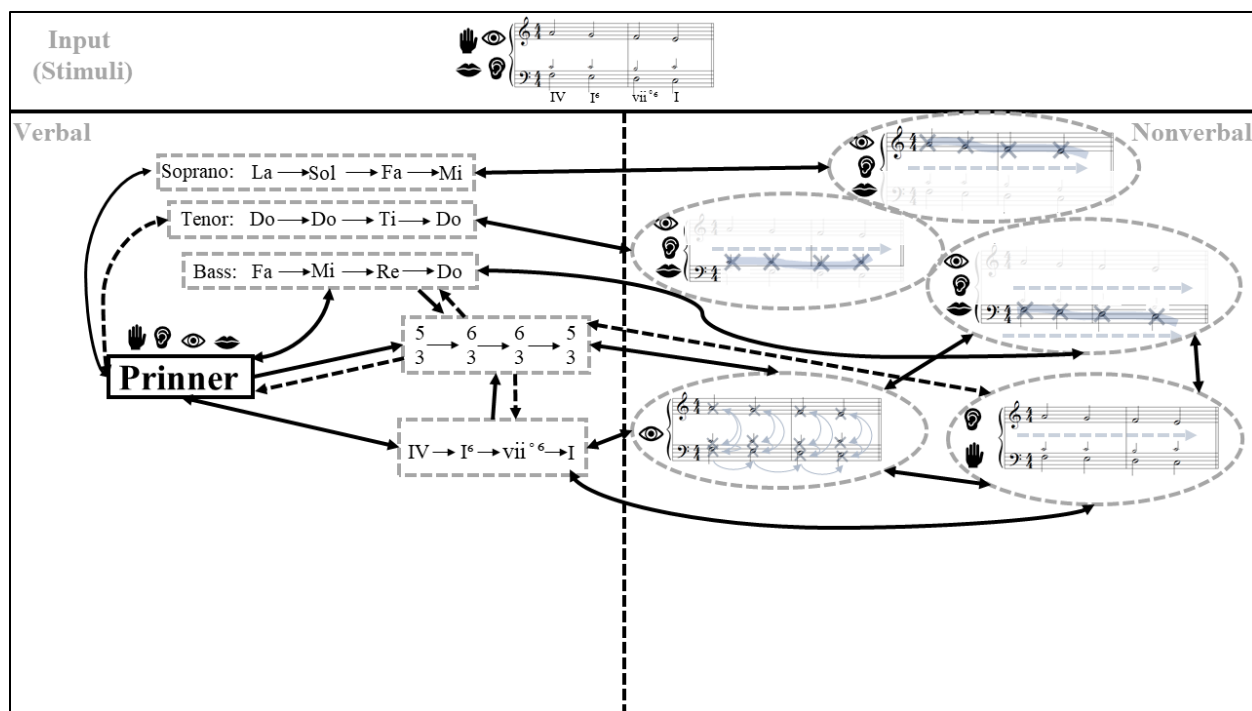


Figure 3.20. Harmony Simulator Including Roman Numerals

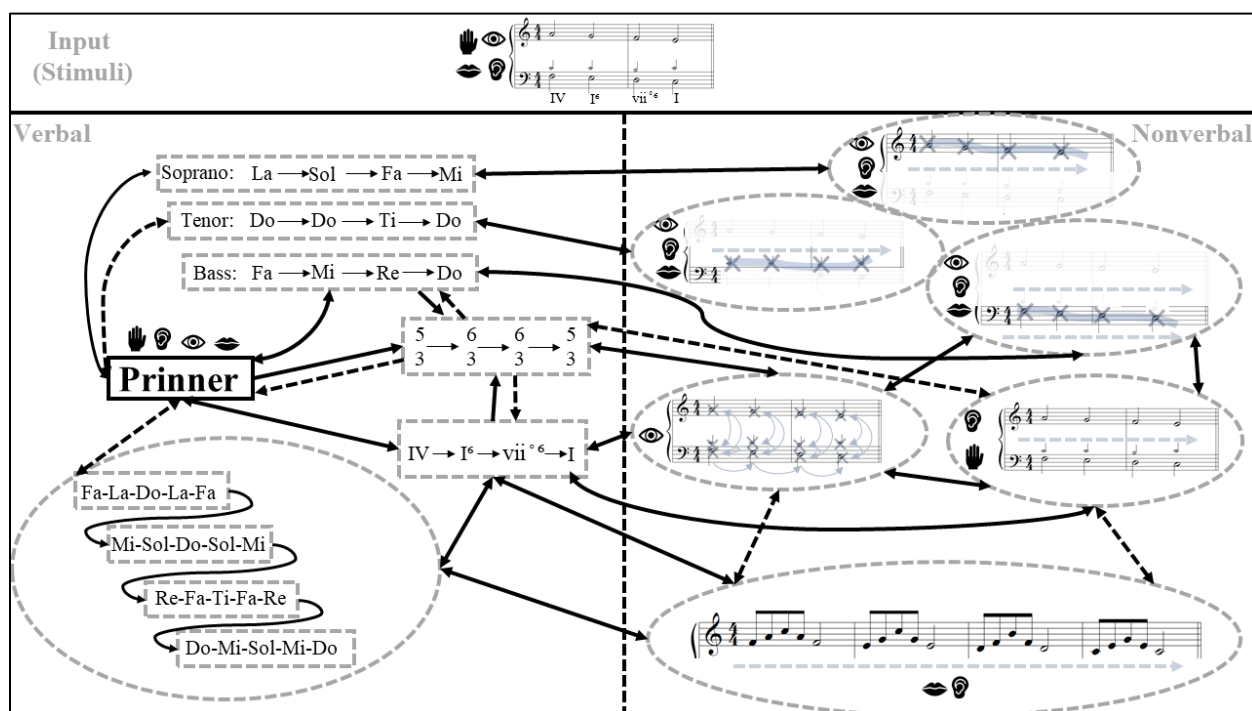


Figure 3.21. Elaborated Harmony Simulator with Additional Interactions and Traces

## Relation simulators

Here I will outline three important relation simulators used in the context of schema theory: counterpoint, harmonic function, and forms and formal function. These concepts serve very different purposes in music theoretic expertise. Counterpoint is a means of encoding relationships between lines and can be conceptualized as either relationships between particular scale degree lines (e.g., given Fa-Mi-Re-Do, La-Sol-Fa-Mi is a likely ‘response’), or for encoding simultaneous relationships between two voices (e.g., 10-10-10-10). Harmonic function is an abstract concept used to conceptualize tonal motion over time, and much like an emotional construction, is grounded in interoceptive representation (often using language of stability or tension or stability). Here, forms and formal function act as retrieval structures: they are primarily ways of creating retrieval points at different regions of scores (visual imagens), and at different levels of hierarchy in sequentially ordered auditory imagens. This is one way of contextualising the understanding of Galant schema as ‘nested’ within larger formal schema. From the current perspective, these schemata are not contained *in* one another as understood in schema theory proper, but instead arise from the act of providing multiple different encodings of the same region of a piece, which forms an association between two different bodies of simulators.

*Counterpoint.* Traditionally, counterpoint was taught as the proper combination of several voices together. Scholars have interpreted these practices as developing sensitivity to stylistic collocations from a construction grammar perspective (Gjerdingen 2020, 131). From this perspective, if one is given the melodic line “Do-Ti-La-Sol,” a proper response or collocation would be a secondary line “Mi-Re-Do-Ti;” counterpoint is a means of solidifying a relation between two melodic lines. Another way of conceptualizing counterpoint is encoding multiple lines through intervallic labelling, for example combining the just-given example as “parallel tenths.” The concept of counterpoint therefore aids in the encoding of the relative positions of each part rather than separately encoding each part. I will now demonstrate both of these types of relational encoding in a counterpoint simulator using the Prinner prototype input. This is accomplished by changing the verbal behavior and in turn, dividing attention between parts during visual and auditory attending rather than focusing attention on a single line. Here, the verbalization for parallel tenths—which specifies only the general distance between voices, and not their tonal position, or even qualities of intervallic distances—is combined with visual attending that includes fixation and saccades between voices, and auditory attention that is equally distributed across soprano and bass voices (see Figure 3.22). While the traces that the theorist encoded for scale degrees *implicitly* contain this information (i.e., nonverbal side only, verbal side *only* through the addition of verbal knowledge of counterpoint), explicitly adding an encoding for the relation provides flexibility in being able to switch attentional focus between objects and their relations, and to evaluate each on their own. This is particularly important as it allows for one to quickly and efficiently divert attentional resources during analysis and listening. It also permits more information to be ‘packed’ into the object concept for scale

degrees, as the implicit relation available on the nonverbal side now has an explicit verbal label for that relation.

Lastly, it is important to note that scale degrees and counterpoint have different conceptual affordances simply by their nature of association with different bodies of knowledge. That is, scale degrees in the order La-Sol-Fa-Mi will have a different associational network than will parallel tenths, albeit with some partial overlap.

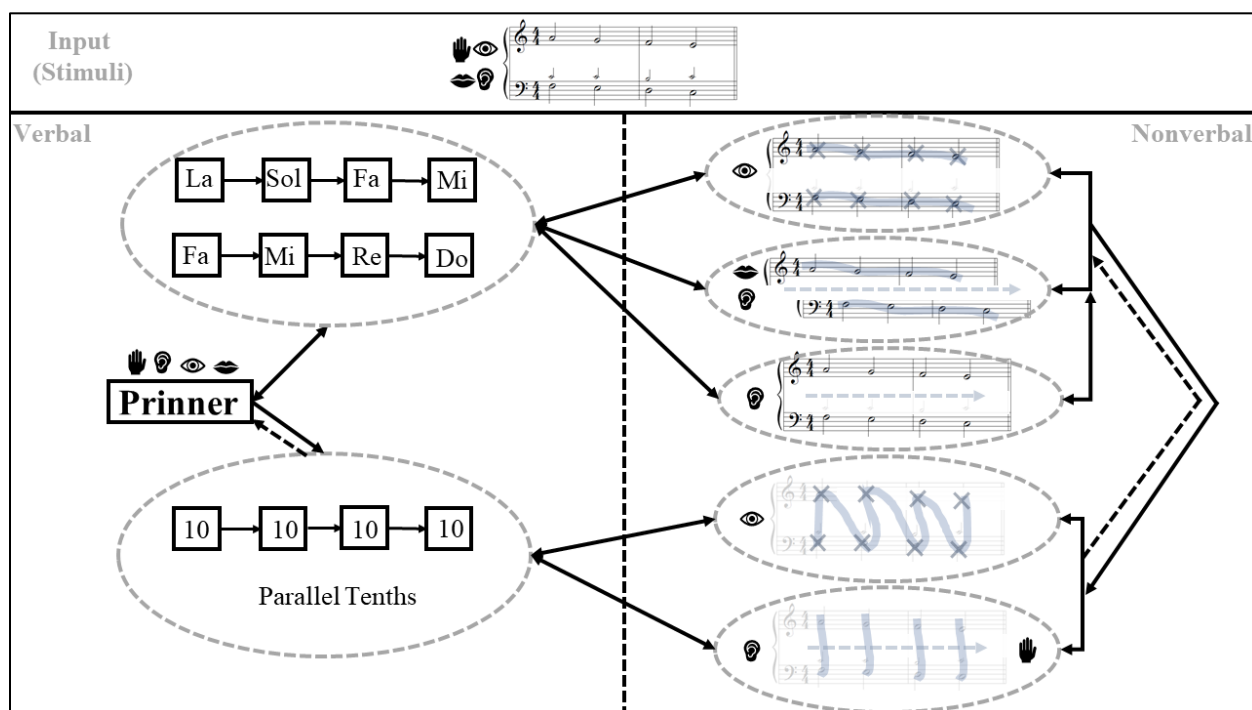


Figure 3.22. Scale Degree Property Simulator and Counterpoint Relation Simulator

*Tonality and harmonic function.* Galant schemata have been explicitly presented in opposition to previous music-theoretic frameworks that have revolved around more ‘universal’ notions of tonality and harmonic syntax as organizing principles of music (see Gjerdingen and Bourne 2015). Despite this, many modern theorists have been trained in such frameworks and apply these concepts to discussion of Galant schemata, particularly when attempting to place schemata into dialogue with other existing theories, such as formal function (see Caplin 2015). Here I will make an argument for the importance of interoceptive representation in encoding and in expertise in general, particularly as it relates to automation and increased reliance on introspection for judgements. Within the pool of music theoretic concepts available to theorists, some concepts—like tonality and harmonic function—focus attention toward introspective states rather than on features of the stimulus itself. In this way, I argue such concepts are equivalent to emotional constructions, as discussed in the previous chapter. These concepts typically involve the invocation of image schemata and other metaphors to ground conceptualization in bodily states (Lakoff 1987; Johnson 2009). These concepts include *tension* and *relaxation* (Lerdahl and Jackendoff 1983, chapter 8; Lerdahl 2001, 186) when conceptualizing tonality and pitch-space, *forces* (magnetism, inertia, gravity, Larson 1993; Larson and Hatten 2012), ‘will of the tones’ (Arndt 2011), and *function* and *discharge* (Harrison 1994) when conceptualizing scale degrees and pitch motion through time. The psychological validity of such conceptualizations, such as tension and stability, appears to be well founded, particularly as such evaluations appear to be fairly consistent among participants enculturated in similar musical styles (see Krumhansl and Toiviainen 2001).

Broadly, theorists use harmonic function to create classes or groups of objects (scale degrees, harmonies) which have a similar usage across contexts. Therefore, the objects are not

themselves the focus of the interaction; rather, the emphasis is on their motion over time and contextual usage at a given temporal location. For example, while the two progressions IV-I6-viio6-I6 and ii6-I6-V4/3-I may indeed differ in their particular features, they may be grouped together under a similar umbrella of function: PD-T-D-T. Functional labels, whether harmonic or applied to other concepts, particularly phrase structure in formal function (Caplin 1999), are important ways of classifying the relationships between perceived tonal motions and their phrase-syntactic usage. Because attention is spread across many different features in a simulator, interoceptive representations help to bind function as a relation simulator.

Recall that emotional constructions, as discussed in the previous chapter, are abstract, event-like conceptualizations that are highly dynamic context-dependent category judgements (Barett 2014, 292). Such judgements are grounded in introspectively available states of the perceiver, particularly interoceptive ones, as they play important roles in aggregating information from multiple inputs during situated conceptualization. Such relational simulators capture judgements made while attending to *multiple* features at once (Barsalou 2003b, 1181), including those involving internal states of the perceiver. As such, harmonic and tonal function will be viewed as a type of emotional construction that, when grounded to particular verbalizations and interoceptive representations, can be used to distinguish between sets of simulators using introspection. Figure 3.23 shows the hypothetical attachment of harmonic function concepts to the Prinner simulator discussed above, which contains scale degree and harmony simulators, and their associations. The verbal classification of each stage is attached on the verbal side in the order of occurrence—predominant, tonic, dominant, tonic—which is then associated with the other conceptualizations for Prinner stages (harmony, and scale degrees). Importantly, the logogen for harmonic function has a direct referential connection with an interoceptive imagen

stored in the nonverbal system, which represents the internal, sensed physiological changes that may occur when attending to harmonic function in the Prinner being encoded. Here, it is represented as a felt-sense of change in stability over time (which may include sensation of balance, or movement). As this felt-sense occurs in real time, it is strongly associated with other sequential and temporally organized imagens, particularly those in auditory and motor modalities. The stability profile for the Prinner schema gradually increases from the beginning to the end of the excerpt.

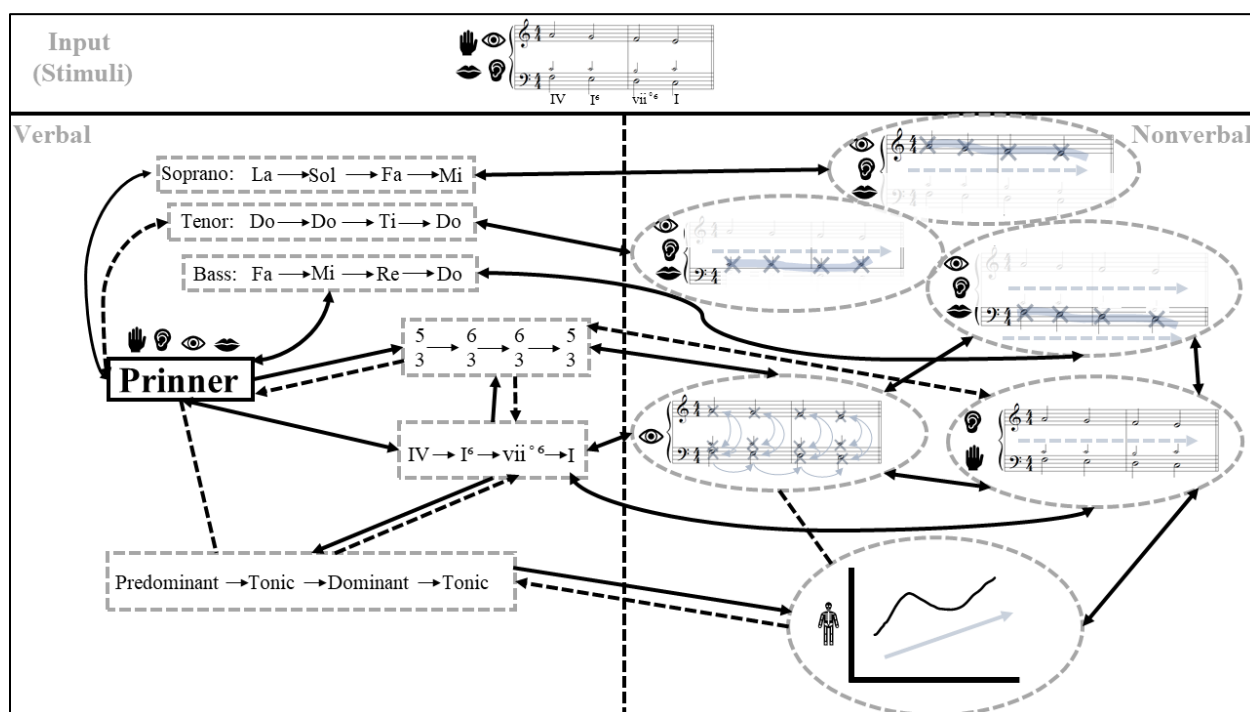


Figure 3.23. Harmonic Function added as Relation Simulator

Functional designations such as associations between verbal labels and felt interoception carry little weight on their own. Their impact comes from their association within a given set of simulators, which allows the interoceptive representation to further differentiate one instance or category from others like it. Take for example, a separate simulator constructed for the rule of the octave (ROTO), which may be very similar in structure to that of the Prinner (see Figure



3.24). Aside from differences in logogen content, what differs is the qualitative evaluation or construction of this simulator pool, signified in part by encoding tonal function. Other ways in which this simulator pool differs from that for the Prinner discussed earlier include how it is associated with other concepts, like formal functions and forms signifying its typical spatial usage (i.e., location in form). The value that harmonic function as a concept can bring to encoded simulators therefore is that, as discussed in the previous chapter, it can become increasingly more sensitive to changes in network association and activation (i.e., FoK potential), reflecting availability of different knowledge pools for each Galant schemata. Therefore, harmonic function, while outside the purview of traditional Galant schema methods, is one way in which to *capture* the relative functionality of each schema through associational connections in memory.

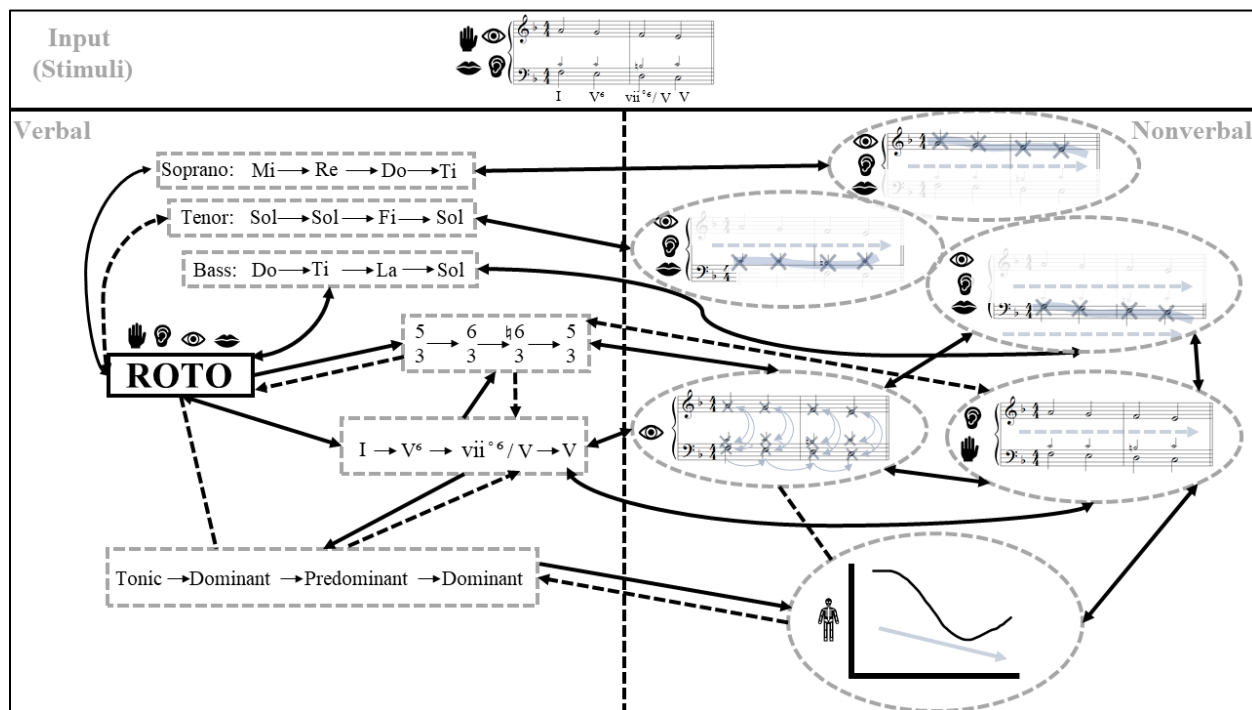


Figure 3.24. Rule of the Octave Simulator with Harmonic Function

*Forms and formal function.* Here, forms and formal functions are primarily be framed as devices for retrieval (i.e., retrieval structures in long-term working memory, see Ericsson and Kintsch 1995). In many ways, they are similar to Galant schema: formulated out of large pools of associated simulators bound together and accessible during categorization. Forms and formal functions likely include more relational simulators than property simulators relative to Galant schema. From this perspective, a simulator acquired for “continuation function” would likely contain a wide array of simulators for the properties and relations associated with that concept, such as fragmentation, harmonic rhythm, and sequence types. Here I will focus on the ways in which spatial concepts, like form, provide a way for music theoretic concepts to become associated, retrieved from memory, and easily compared across instances. I will not discuss the organization of simulators specific to forms and formal functions, but instead demonstrate their organization in memory at a larger level to understand their interaction with Galant schemata.

As a theorist interacts with a given piece of music through analysis, they apply concepts in various modes. As this occurs, simulators are similarly attached and associated in memory for that exemplar, much in the same way as was previously discussed with the Prinner prototype (see Figure 3.25) In this figure, note how the simulators that make up the Prinner concept are attached to regions of the visual and auditory imagens in memory, reflecting the activation of previous knowledge and the association of previously acquired simulators with these areas, and the potential explicit encoding of simulators to these regions. A theorist might write, sing or play a reduced variation of each Prinner type for K. 545, which would enhance the Prinner simulator base, and more strongly associate those simulators with this exemplar.

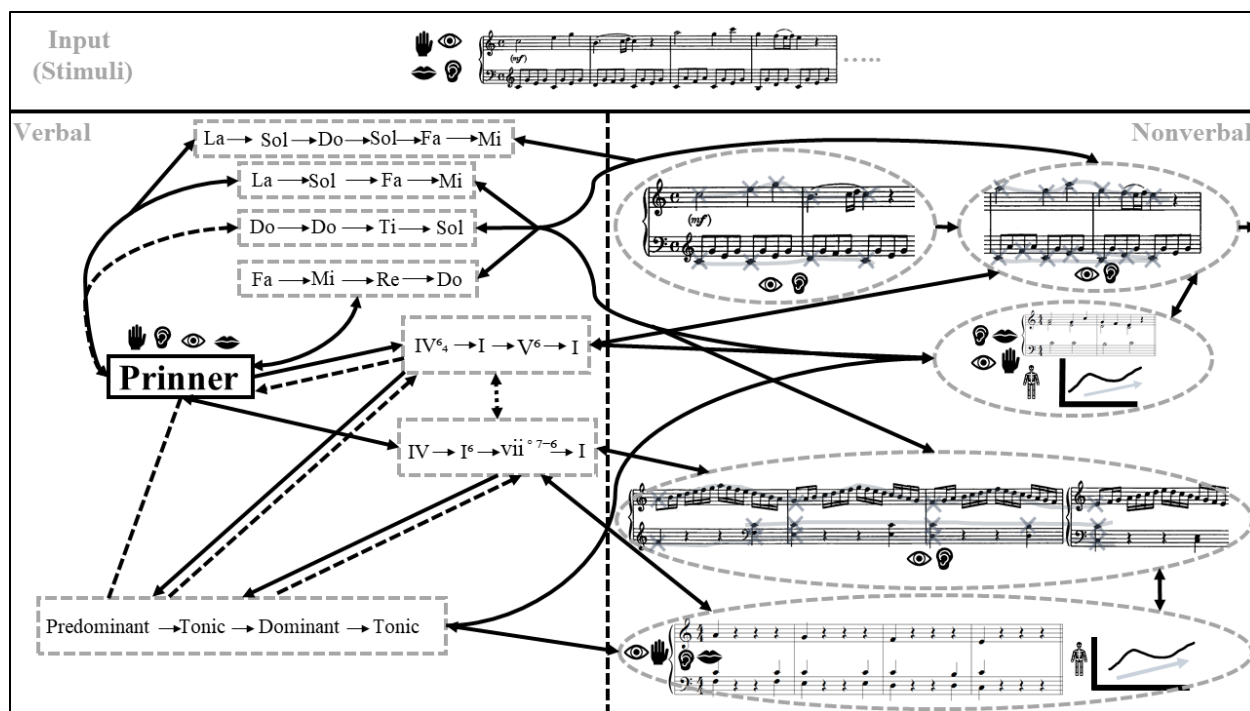


Figure 3.25. Prinner Simulators for K. 545

In the context of Sonata form, theorists will also use other concepts for formal analysis. Such concepts, shown here primarily as verbal associations (retrieval cues), are associated with various spatial locations both through vision (reinforced through labeling or writing on the score), and through recalling these labels while listening (see Figure 3.26). Each formal concept learned in isolation would have its own pool of simulators; however, each formal concept, acting as a retrieval cue for a different point, also contains any other simulators bound at those spatial regions. Therefore, in Figure 3.26 we see that both the contrasting idea and the continuation phrase have the Prinner concepts bound to them, on the verbal as well as the nonverbal sides. In this way the ‘nesting’ of different schema does not involve unimodal abstract representation. Using schemata in analyses, formal function and sonata form simulators become structured and bound to regions of the piece. This repetitive interaction of conceptualization—acquiring simulators to pieces and recalling previous knowledge attached to those simulators in

simulation—provides the large, highly interconnected network of associations required for memory skill. This will be a focus of the following chapter.

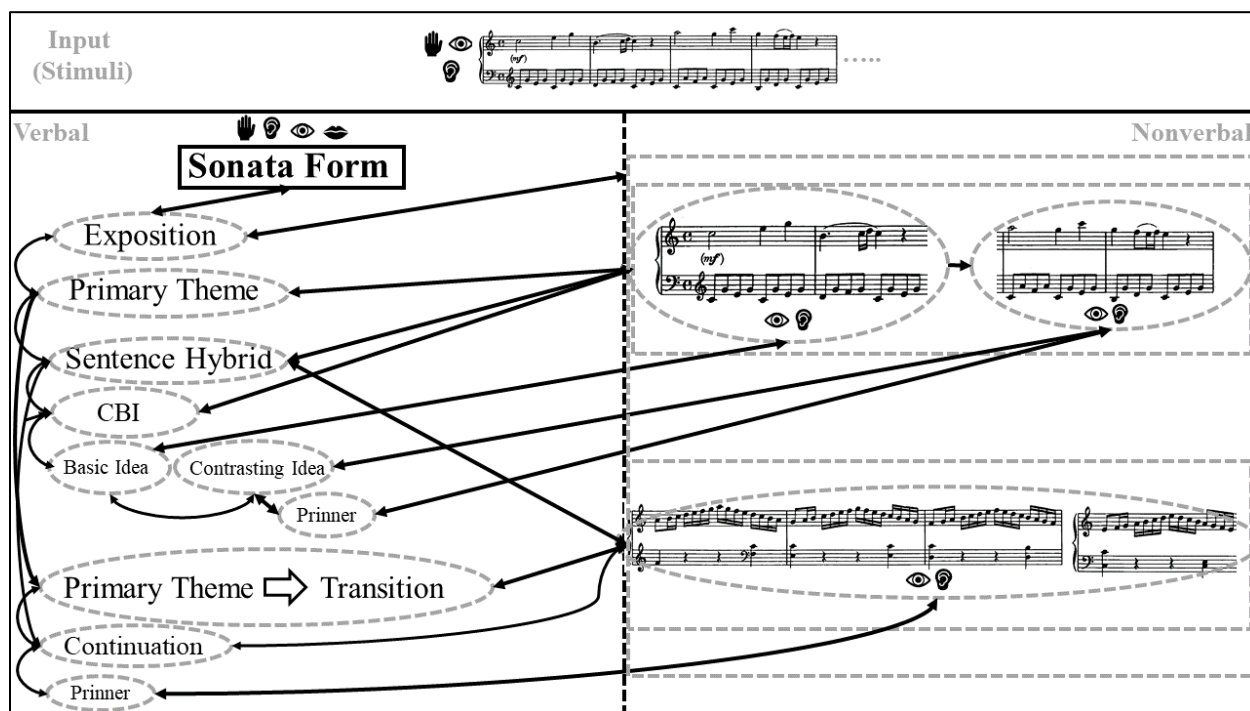


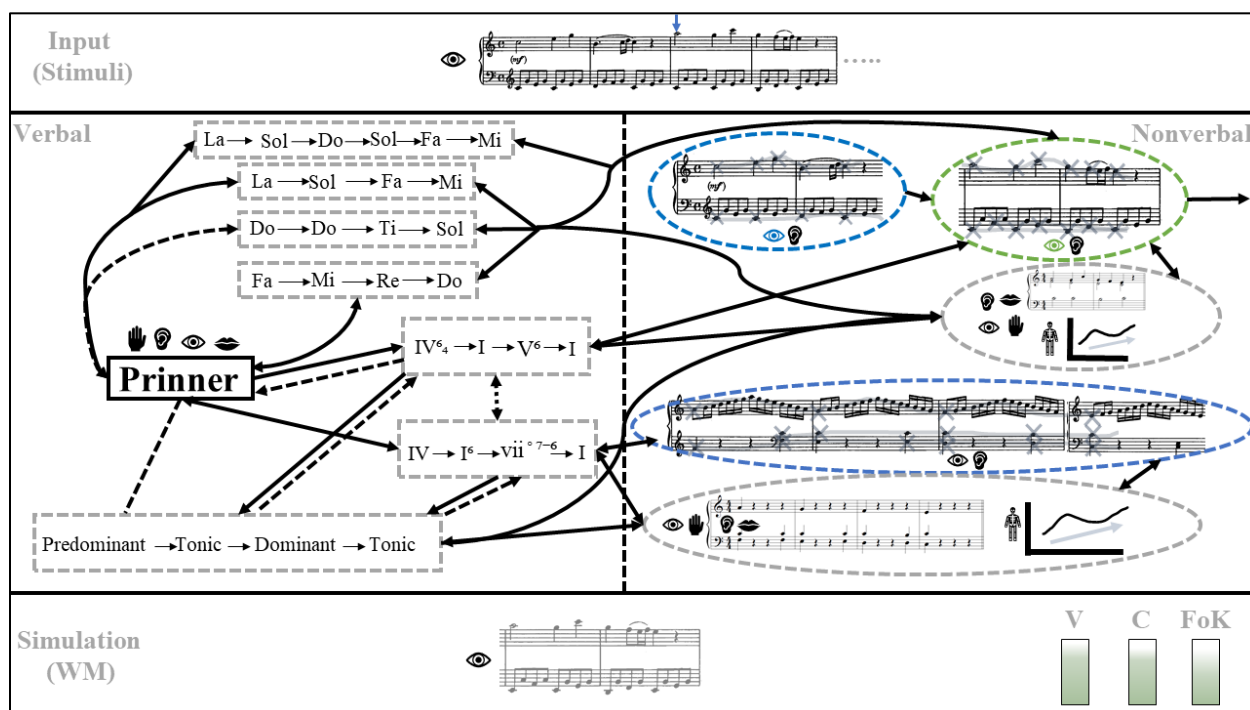
Figure 3.26. Form and Formal Function (Verbal System) as Relation Simulators for Retrieval on Nonverbal Side

### On Simulation: Flexibility in Retrieval and Maintenance

While the topic of flexibility in simulation is largely unpacked in chapter five, in order to elucidate differences in memory skill between the enculturated listeners and experts, I will briefly discuss the matter here. Recall that while an enculturated listener was able to imagine parts of exemplars and respond to categorization prompts, they were severely limited in their ability to manipulate simulations: unable to jump between sequentially organized auditory chunks, or easily access information held within a chunk. Contrastingly, a music theorist possesses much more capacity to easily perform such operations. It cannot be overemphasized how much the addition of visual information in the form of a score aids these operations. The availability of a

visual format to scan, along with associations between visual and auditory representations, allows auditory information to become somewhat synchronously available through association to visual imagens (see Figure 3.27a, b). The theorist need simply scan across to the desired location in the score (Figure 3.27a), and then instantiate the corresponding auditory representation (Figure 3.27b).

(a). Visual Scanning and Representational Activation of Visual Imagen



## (b). Associational Activation of Corresponding Auditory Imagen

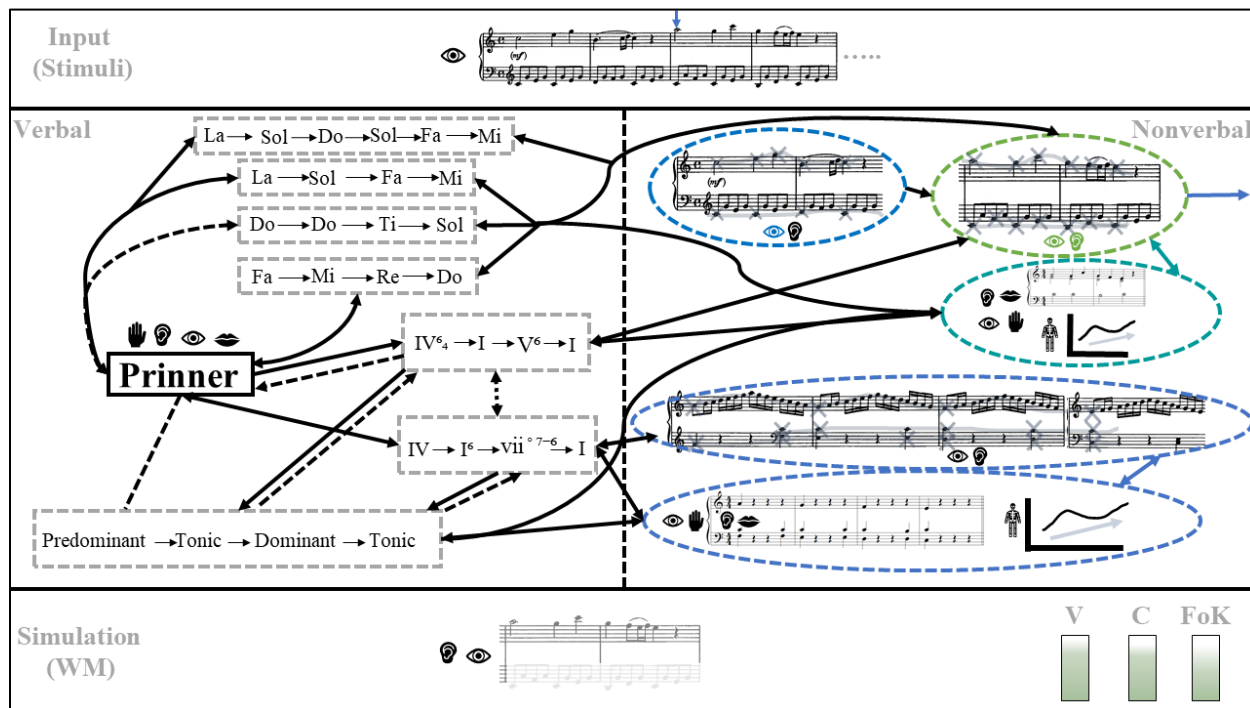
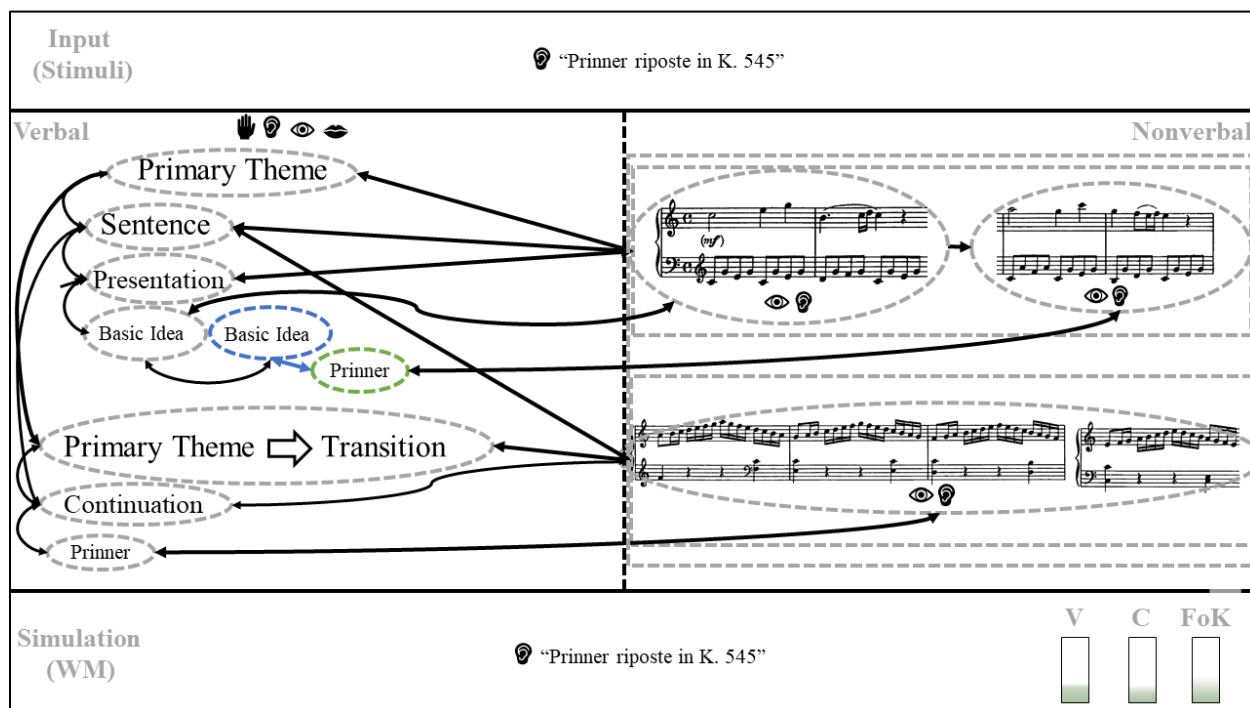


Figure 3.27. Simulation of Middle Chunk of K. 545 Through Visual Representational Activation (a) and Auditory Association (b)

Theorists can similarly use referential processing from verbal cues to access and reconstitute auditory representations during simulation (Figure 3.28a, b). They may also use simulators stored at those points to focus attention in imagery on a particular feature of interest, such as the soprano line (Figure 3.29).

## (a). Verbal Representational Activation



## (b). Referential Activation of Corresponding Nonverbal Imagens

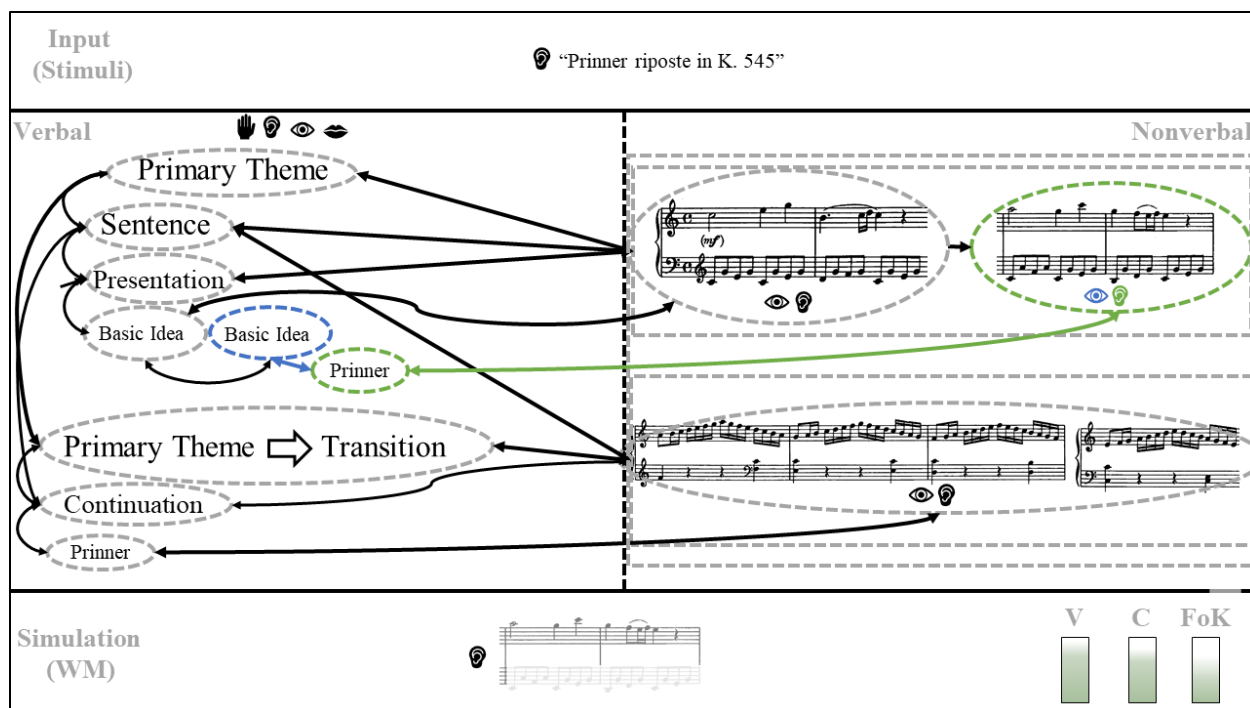


Figure 3.28. Verbal Representational Activation (a) and Referential Activation (b)

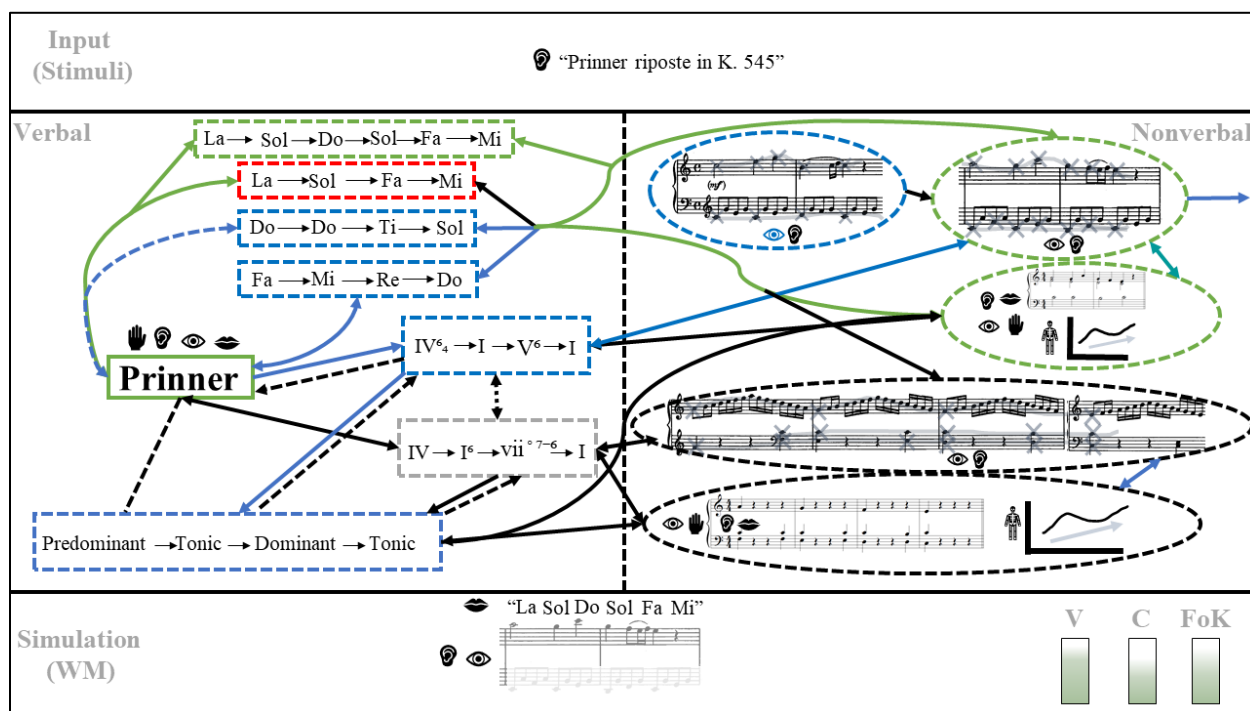
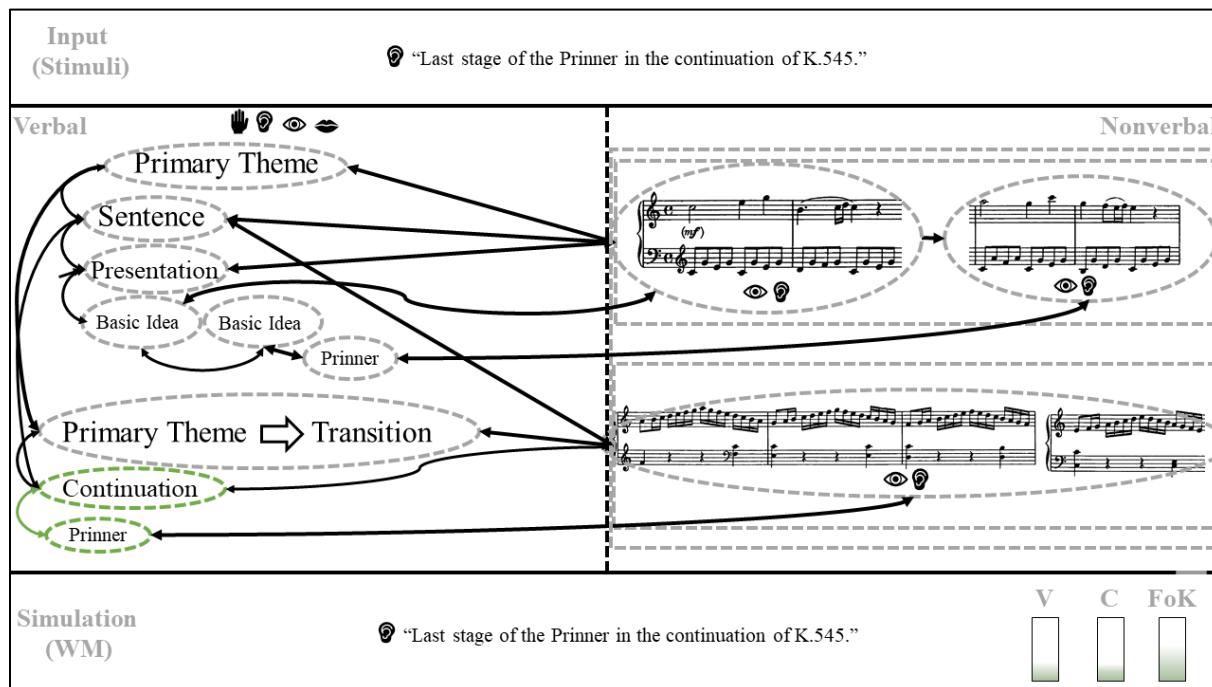


Figure 3.29. Simulation of Soprano Line in Context Using Prinner Simulator

There are also a variety of ways of accessing information held within a given chunk. A theorist may instantiate using a verbal cue, then scan across in auditory imagery to the desired location in the auditory trace (Figure 3.30a,b), or alternatively use other simulators stored at those regions to bypass or speed up the sequential cuing (see Figure 3.31). A theorist might choose to use a verbal cue for the soprano solfège in the Prinner simulator to skip ahead and cue the final stage of the Prinner without imagining in depth the entire chunk, performing a kind of representational exchange (logogen for imagen).



(a). Verbal Activation



(b). Referential Activation and Auditory Scanning in WM

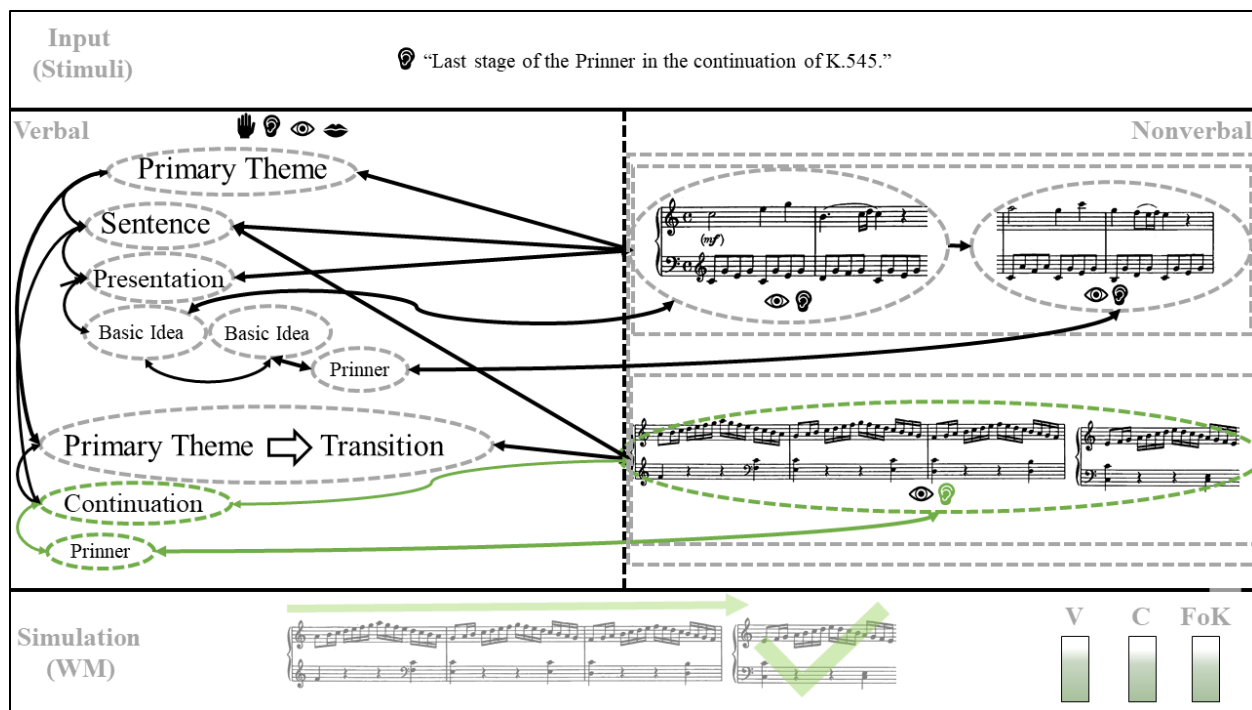


Figure 3.30. Access to End of Auditory Imagen Through Referential Activation (a) and Auditory Scanning (b)

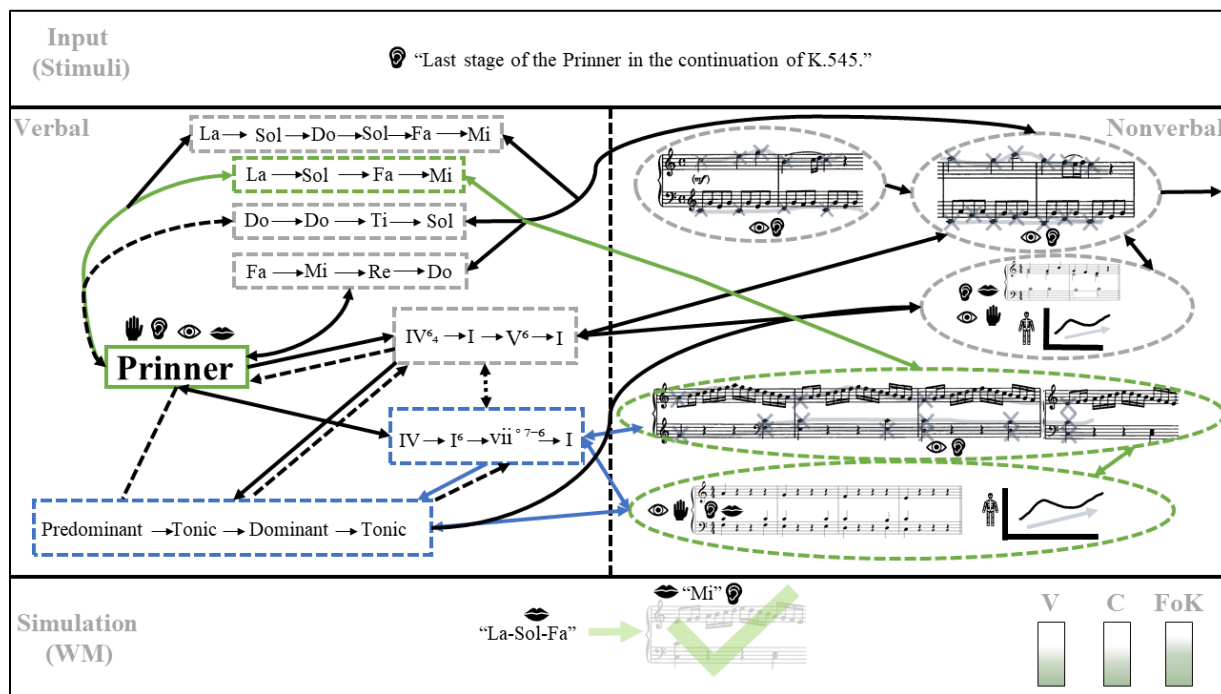


Figure 3.31. Representational Exchange Between Verbal and Nonverbal Units

Aside from advantages in retrieval for simulation, the structure of expert memory allows for powerful abstraction in situated conceptualization ability. For example, take the two shapes in Figure 3.32. It seems safe to assume that the reader has never seen Galant schema in this exact presentation before, and thus they are like ‘brand new exemplars.’ However, this likely poses little challenge for labelling them, something which the current framework is able to explain.

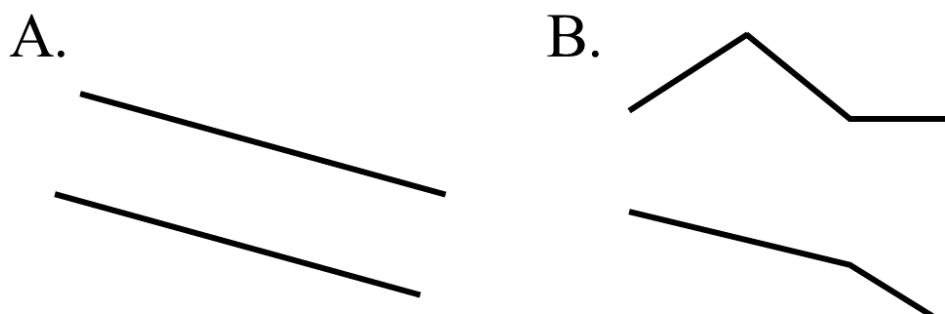


Figure 3.32. Abstract Schema Shapes

Because the mode of interaction is maintained in visual imagens (i.e., scanning patterns for voices, and their relative positions), these shapes partially activate or prime many different stored exemplars in the nonverbal system. Following this, spreading referential activation to the verbal system affords identification of the schema as a Prinner (Figure 3.33) and Romanesca (Figure 3.34). Further categorization behaviors are available as these abstract symbols can also prime and activate other representations held in the simulators by association. For example, it would be quite easy to simply now imagine the sound of a Prinner bassline in ‘the abstract’ (Figure 3.35) or imagine a familiar exemplar of a known Prinner (Figure 3.36). These examples demonstrate the power and flexibility of the simulator networks described above.

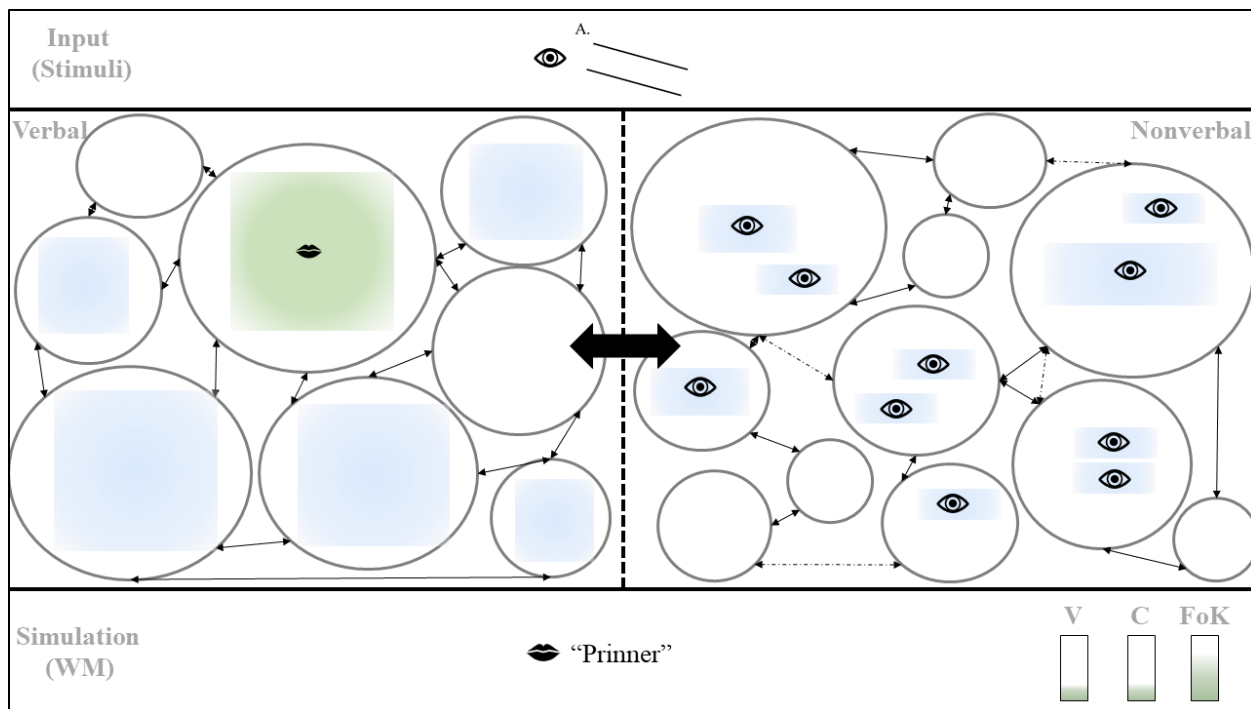


Figure 3.33. Visual Priming and Referential Activation in Prinner Simulators

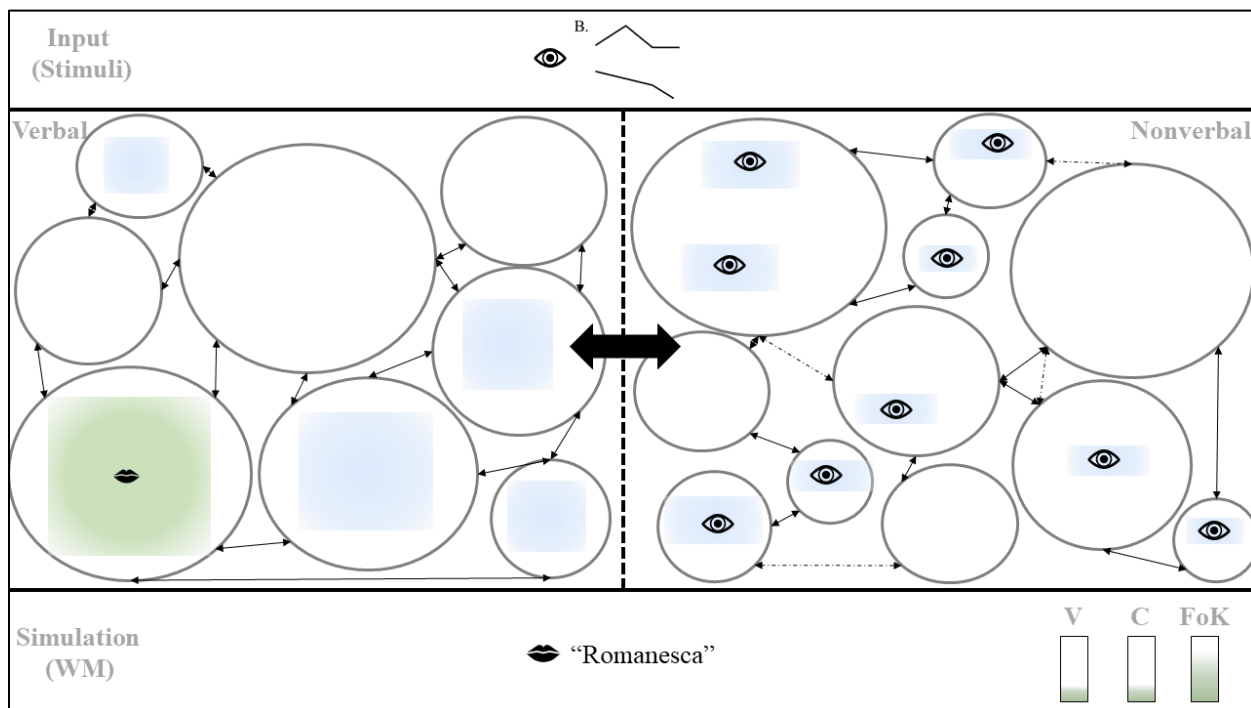


Figure 3.34. Visual Priming and Referential Activation in Romanesca Simulators

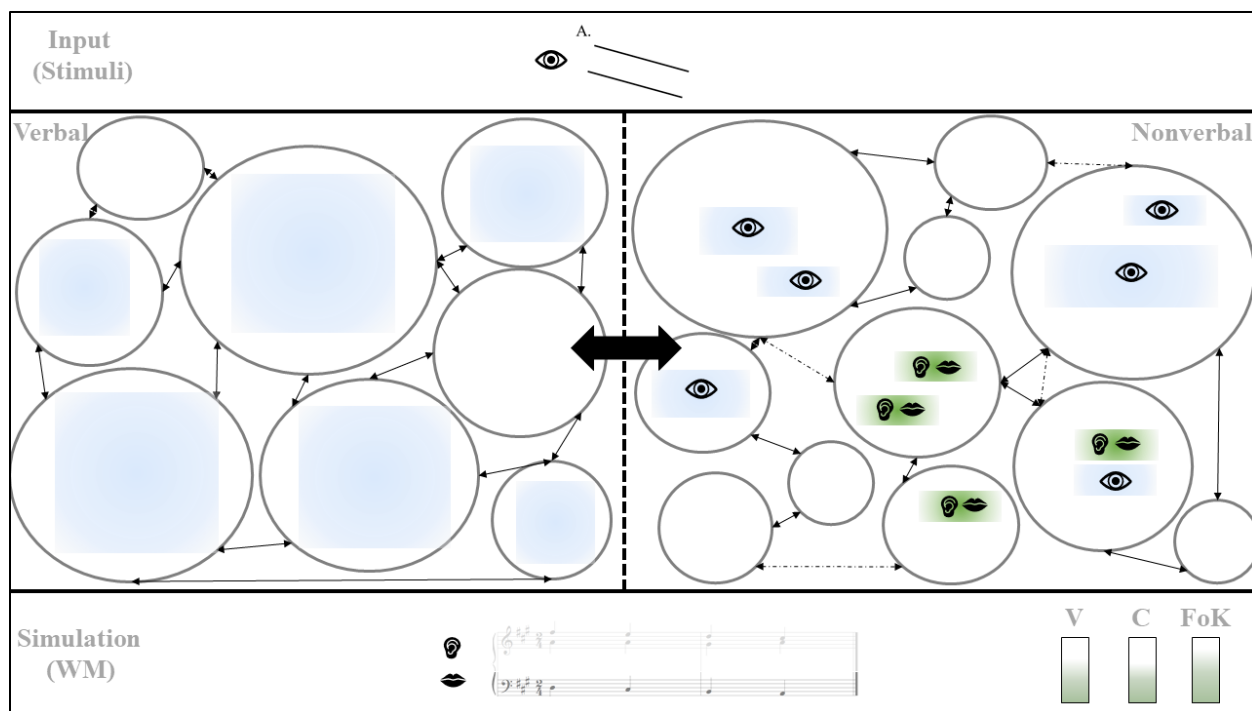


Figure 3.35. Associational Activation and Auditory Simulation of Prinner Baseline

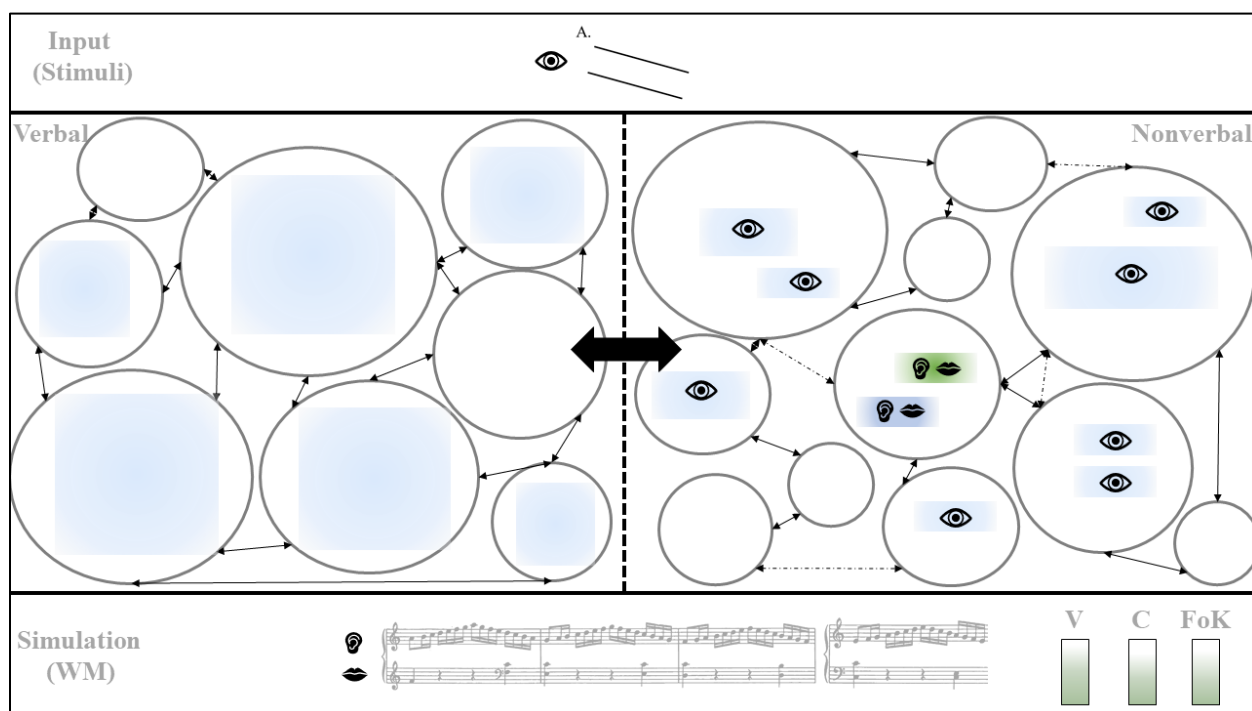


Figure 3.36. Selection and Simulation of Familiar Prinner Exemplar

## Summary and Conclusions

In this chapter I have outlined the differences in representation and simulation ability between two groups: encultured listeners and Galant schema experts in music theory. While encultured representation garnered from auditory statistical learning affords many cognitive functions needed in listening, the modally and system-limited nature of such loose holistic representation means that simulation is quite restricted. Contrastingly, as theorists explicitly acquire Galant schemata categories using music theoretic concepts—which actively encode properties and relations into LTM—simulation ability is much more flexible. Now that I have used the DCT framework to outline the basic organization of a Prinner schema in a hypothetical music theory expert, I will use this framework to create a hypothetical developmental trajectory for the Prinner category as it is acquired over time by a hypothetical modern learner.

## Chapter 4

# Galant Schema Acquisition as Memory Expertise: Acquiring Eighteenth Century Hearing

In this chapter, I adapt the memory expertise framework developed in Gates (2021) to construct a hypothetical developmental trajectory for Galant schema acquisition by a modern music theorist. In the first portion of the chapter, I will overview the long-term working model developed by Ericsson and Kintsch (1995), and the adaptation of this framework in Gates (2021). Here I suggest that the ‘Loop’ in music-theoretic expertise is an iterative process that focuses on expert memory acquisition. I then provide an account of Galant schema acquisition, firstly by positioning traditional training methods within Neapolitan and Parisian conservatories as a form of memory expertise, and secondly, by sketching out a developmental trajectory for Galant schemata (specifically, the Prinner) in a modern context using the framework developed in chapters 2 and 3.

### Memory Expertise Defined: Ericsson and Kintsch (1995) Long-Term Working Memory

The long-term working-memory model (hereafter LTWM) of memory expertise (Ericsson and Kintsch 1995; Ericsson 2018) evolved out of extensive research into expert performance in domains like chess in order to explain how these experts were able to bypass known memory constraints.<sup>58</sup> The field has since grown to explore many different domains,

---

<sup>58</sup> These are primarily related to the functionality and limitations of working memory and long-term memory. Information held in working memory is temporarily stored and therefore accessible very quickly. Due to its

including sports, medical diagnosis, and memorized musical performance. The LTWM model suggests that in order to acquire an expert level of performance in a given discipline, one must acquire domain-specific memory skill and domain-specific expertise.

The LTWM model proposes that through many hours of deliberate practice, an expert acquires a set of knowledge structures and retrieval cues which forms the basis of an acquired memory skill (Ericsson and Kintsch 1995). Such acquired skill is believed to consist of three components. The first is a large body of domain specific knowledge. The second is a set of retrieval cues for this domain specific knowledge arranged in the form of a stable hierarchical structure called a retrieval structure (see Figure 4.1).<sup>59</sup> The last component is LTWM itself, in which encoding and retrieving information using a retrieval structure is cultivated and rapidly sped up with deliberate practice, eventually making the rate of information storage and retrieval from LTM comparable to that of WM (Ericsson 1985, 194).<sup>60</sup>

---

temporary nature, it is known to be susceptible to interference effects, such as retroactive interference (a process by which information in working memory is overwritten by new incoming information and/or processing). Long-term memory is not temporary, and therefore is not prone to these sorts of effects. However, processing speeds of long-term memory for retrieval and encoding are extremely slow (over 5–10 seconds). Ericsson and colleagues have found that certain memory experts are not susceptible to interference effects while performing tasks that require working memory, suggesting they instead use long-term memory at comparable retrieval and encoding speeds to working memory (see Ericsson and Kintsch 1995; Ericsson and Roring 2007).

<sup>59</sup> These retrieval cues are presumed to be organized hierarchically, both spatially and sequentially in memory, and are used to efficiently recall any encoded information associated with them. For example, such a structure was used by a subject called “SF” to memorize and recall a series of 30 digits. SF was reported to use a mnemonic coding scheme at retrieval cue level 1 (e.g., digits 3596 encoded as 3 mins 59.6 seconds), along with a spatial encoding scheme (relative positions of such cues into groups) at level 2 in the retrieval structure (Ericsson and Kintsch 1995, 217).

<sup>60</sup> Contrasted with chunking theory, which says that the size and complexity chunks in working memory increase with expertise, Ericsson’s memory skill theory helps to explain how experts circumvent known limitations of working memory, namely retrospective interference (the wiping of information in working memory), as well as capacity constraints (like Miller’s magic number, see Ericsson and Kintsch 1995, 215).



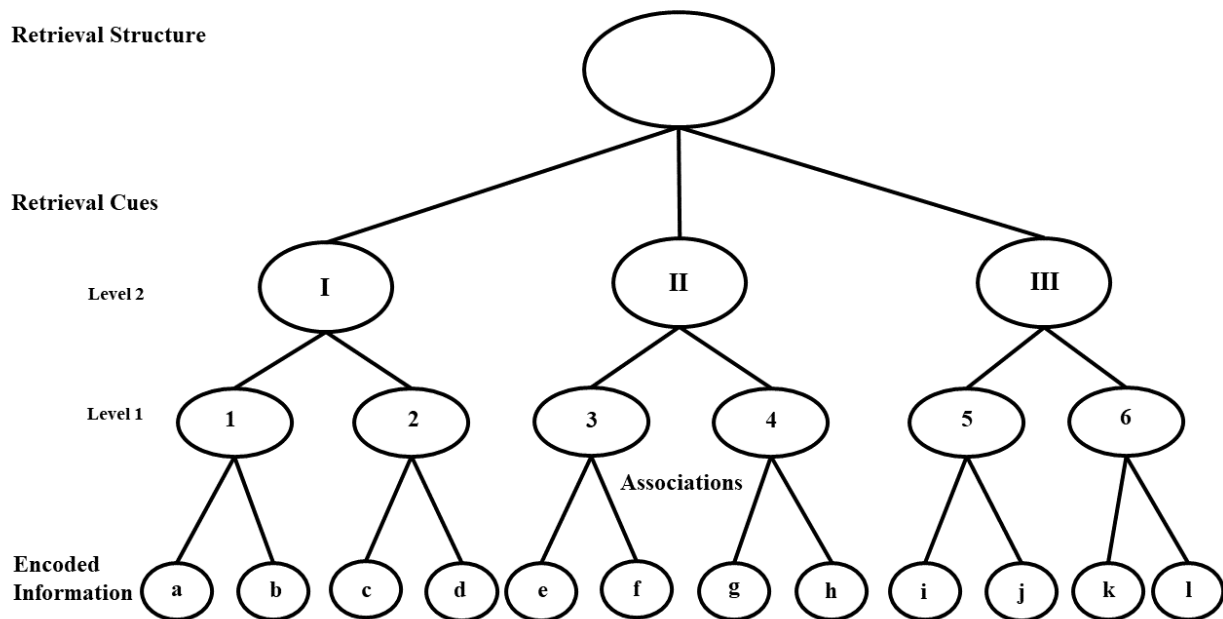


Figure 4.1. Sample Retrieval Structure (from Gates 2021)

Acquired memory skill is specifically tailored to meet the demands imposed by a task or set of tasks, such that expertise in a given domain does not generalize (Ericsson and Roring 2007). Therefore, an expert's acquired memory skill (i.e., their knowledge and retrieval structures) will vary drastically based on their domain of expertise. Each type of memory skill therefore presents a set of processing benefits and deficits. For example, some domains may have a high demand for retrieval accuracy (e.g., memorized musical performance or delivering a memorized speech), while others put more of a premium on the encoding and processing of new information in order to adjust potential future actions (e.g., chess, sports, medical diagnosis).

Recent neuroimaging research has verified much of Ericsson and Kintsch's (1995) LTWM model, including its implied developmental trajectory. Guida et al. (2012) completed a meta-analysis of experimental neurological data comparing novices to experts in various domains. Their findings reveal a hypothetical developmental trajectory of functional brain

reorganization gained with expertise, which supports many of Ericsson's claims. The authors suggest a gradual two-step process in the acquisition of expertise and related brain changes (Guida et al. 2012, 221–44). Recall that the first stage of expertise acquisition in Ericsson's model includes chunk creation and storage of relevant information in LTM. As novices gain relevant knowledge structures in LTM, and these structures become more efficiently processed, working memory becomes less burdened, resulting in a reduction of brain activity (Guida et al. 2012, 235–236). The second phase of Ericsson's LTWM model is the speeding up of memory skill with practice, resulting in long-term retrieval being nearly equal in speed to that of working memory. This second phase results in functional reorganization of the brain (Guida et al. 2012, 236), reflecting acquired memory expertise. This trajectory was applied to musical imagery acquisition in the aural skills classroom in Gates (2021), to which I will now turn. While this original work was more narrowly focused on musical imagery, I will show that it can easily be adapted to the current context.

### Dual-Coding and Situated Simulation as Memory Expertise: Insights from Gates (2021)

Gates (2021) examined pedagogical practices for imagery development in North American aural skills and developed a hypothetical model for imagery development in the aural skills classroom by integrating these insights with research in the cognitive sciences. In this article, I proposed that imagery development is a form of LTWM. In particular, I argued that the functions of a wide range of pedagogical activities—content acquisition, impacting imagery quality (e.g., tonal imagery), imagery cue development, and lastly, increasing perceived cognitive awareness and metacognition—map onto stages of LTWM acquisition (see Figure

4.2). Content acquisition and imagery quality activities map onto the first stage of LTWM acquisition—chunk acquisition and semantic encoding—while the methods for acquiring cueing imagery map onto the construction of retrieval cues phase. Lastly, activities that are purported to aid in metacognitive awareness and the freeing of mental resources (‘doing things with imagery’), represent the final stage of LTWM acquisition, where memory encoding, storage and retrieval are sped up, stabilized, and made more automatic. These stages are in turn mapped onto neurological changes observed with increasing memory skill. Finally, I proposed that such changes would correlate with subjectively observable differences in imagery quality; an increase in perceived vividness in imagery quality during the initial stages of LTWM acquisition, and an increase in perceived control over imagery when LTWM is acquired. In the current section, I will expand upon this work to extend the framework to category learning and simulation expertise in music theory.

Aural Skills Imagery Development Category	LTWM Model (Ericsson and Kintsch 1995)	LTWM Function	Neurological Change (Guida <i>et al.</i> 2012)	Relevant Imagery Properties
Content Acquisition	LTM Encoding (Semantic/Meaningful Encoding)	Chunk Formation	Reduction in Brain Activity	Vividness
Imagery Quality (e.g. Tonal Imagery)		Chunk Relationships and Associations		
Methods for Image Generation	Cues in Retrieval Structure	Cues for Retrieval and Maintenance	Functional Reorganization	Control
Doing Things with Imagery	Acquiring LTWM	Speeding up LTM Encoding and Storage		

Figure 4.2. Aural Skills Activities as LTWM Development (from Gates 2021)

### The “Loop” as LTWM Acquisition

In this section, I expand this framework to account for memory expertise acquisition in music theory, using Galant schema theory as a case study. Recall Rogers’ (2004) description of the purpose of music theory pedagogy discussed in chapter 1: a loop between thinking and listening. Rogers proposes that ‘thinking’ (categorical knowledge acquired in the theory

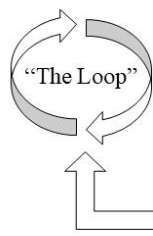
classroom) and ‘listening’ (sonic content acquired in aural skills) are two sides of the loop, developed in an intricately reciprocal manner. The link between the two activities—analysis—is where theoretic knowledge, both verbal and nonverbal, is brought online in context so that the analyst may come to a greater understanding of a piece of music. This process, repeated iteratively, represents the ‘loop’ in music theory expertise, and is easily mapped onto the LTWM trajectory (see Figure 4.3). Here, ear training forms the basis for encoding and storing chunks (primarily nonverbal) into memory, while mind training can be understood as the acquisition of a diverse set of verbal retrieval cues for these nonverbal chunks.<sup>61</sup> When these knowledge sets are deployed together in the context of analysis, they need to be fluently (re)encoded, stored, and retrieved, demonstrating the acquisition of LTWM skill.

From a DCT perspective, each stage shows representation acquisition and fluency in different processing types. Ear training, or chunk formation, involves the acquisition and association of nonverbal imagen representations, including fluency in representational and associational processing. Mind training—the acquisition of retrieval cues—involves the acquisition and association of logogens, including within-system associational processing, and importantly, cross-system referential connections to nonverbal representations. During analysis, one gains fluency in representational, associational, and referential processing, such that representations in differing modalities (auditory, visual, motor) and formats (verbal/nonverbal) can be fluently accessed and exchanged for one another. From the DIPSS perspective used here, the initial stages of LTWM acquisition can be viewed as the acquisition of a large simulator base

---

<sup>61</sup> In practice, this boundary is far fuzzier than it is presented here. I do not mean to suggest that verbal encoding (‘mind training’) includes no semantic, meaningful encoding. The addition of verbal codes to a nonverbal base helps encourage larger more meaningful connections in memory that may not exist from nonverbal processing alone. Similarly, verbal codes are often acquired in the context of score analysis and part writing, so both visual and motor imagens are acquired during ‘mind training.’

in category learning, while acquiring skill in such LTWM ability is viewed as gaining fluency in simulation ability and categorization skill. As in Gates (2021), such phases are understood to have probable effects on subjective imagery properties—vividness and control—through the LTWM acquisition process.



LTWM Model (Ericsson and Kintsch 1995)	LTWM Function	Rogers (2004)	Paivio (2007)	Barsalou (2003a) DIPSS	Relevant Imagery Properties
LTM Encoding (Semantic/Meaningful Encoding)	Chunk Formation Chunk Relationships and Associations	Ear Training (Listening)	Imagen Acquisition (Representational and Associational Connections)	Simulator Acquisition	Vividness
Cues in Retrieval Structure	Cues for Retrieval and Maintenance	Mind Training (Thinking)	Logogen Acquisition (Associational and Referential Connections)		
Acquiring LTWM	Speeding up LTM Encoding, Storage and Retrieval	Analysis (The Link)	Fluency in Representational, Associational and Referential Processing	Fluency in Simulation	Control

Figure 4.3. LTWM Development Updated for Music Theory Expertise in the Current Framework

## Memory Skill and Category Learning: Historical and Modern Pedagogical Approaches Compared

Now that I have outlined my approach to music theory LTWM acquisition, I will situate it within the context of Galant schema acquisition. To do this, I will position and compare historical training techniques for Galant schemata as discussed by scholars with learning techniques available to modern North American theorists. By analysing traditional training methods, many of which are used—albeit in diluted formats—by modern musicians, it is possible to examine the functions of various training activities from a cognitive perspective. As learning Galant schemata in a modern context is much less systematic than historical conservatory training, comparisons of modern with traditional methods can help to shed light on the effect of shifting the goal of training toward a music theoretic analytic context rather than practical musicianship.

## **Schemata Learning as Memory Skill Acquisition: A Historical Perspective**

In this section, I demonstrate that traditional training pedagogies, particularly within the Naples and Paris Conservatory traditions, were designed to systematically develop domain specific memory skills or LTWM. Previous work has shown that many of these pedagogical practices revolved around the implicit learning of Galant schemata patterns (Gjerdingen 2020; Barawganath 2020). The learning of category labels for these patterns appears to rarely have been a part of the training process, which emphasised implicit procedural learning over explicit declarative learning. Rather, training revolved around the acquisition of meaningful patterns (schema) and their statistical likelihood of collocated patterns or combinations (Gjerdingen and Bourne 2015).

In the first section which follows, I discuss the functions of Galant pedagogical training types in the context of LTWM development. Here I argue that each pillar of Galant training—*solfeggio*, *partimenti*, and *counterpoint*—progressively shifts from encoding and storage related activities (simulator acquisition) to those focused more on retrieval practice (simulation acquisition). While each pillar of training may be considered a separate type of domain-specific memory expertise (i.e., each goes through LTWM stages individually), together they provide highly distributed simulators for Galant categories stored across modes and systems. In the second section, I provide more detail regarding the so-called ‘semantic’ or meaningful encoding that occurs in traditional Galant pedagogies. Here I argue that simulators are acquired through extended, gradual memory elaboration. Essentially, the simplified ‘models’ or ‘prototypes’ acquired early in training play vital roles in carving out space in memory onto which subsequent exemplars are attached, aiding in elaborated category learning. Similarly, such simplified models provide a foundational basis for category interactions—that is, so-called ‘habit responses’

(Meyer 1956) in controlled conditions. These early learned models become frequently modified exemplars used during active category simulation, essentially transforming them into a type of prototype representation that helps to bind perceptually dissimilar exemplars into probabilistic simulator pools.

### Training Types and their Functions

Like scholars who emphasize the extent to which such training practices were focused largely on building memory skill (see Gjerdingen 2020, 83), here I propose that the pedagogical traditions in Naples and Paris conservatories largely revolved around developing domain-specific memory skill, or LTWM. Such memory skills were geared toward providing students with the ability to rapidly compose and improvise, in addition to training instrumental and vocal musicians (Gjerdingen 2020, 191). Scholars have noted that historical training regimens revolved around three primary activities or core pillars—solfeggio, partimenti and written composition or counterpoint (Sanguinetti 2012, 42; Baragwanath 2020, 288). I argue here that such training practices reflect LTWM acquisition. These training activities gradually progress from encoding and storage of musical patterns such as solfeggio and partimenti rules, to retrieval practice using advanced partimenti fugues, counterpoint, and composition. Thus, early stages of traditional training methods are primarily focused on implicit learning of category representation (simulator acquisition), while later stages of training focus more on explicit retrieval practice (simulation practice). Each training pillar—solfeggio, partimenti, and counterpoint—operates within the same repetitive ‘loop,’ designed to develop fluent memory skill, progressing from encoding to retrieval practice in an iterative fashion. On a larger scale, as students progressed through the different training domains (solfeggio to partimenti and eventually counterpoint training), their activities became increasingly more focused on associational encoding and retrieval of prior

representations rather than on encoding and storage of simulators in category representations (see Figure 4.4). Each form of practice produced its own domain specific expertise. However, because knowledge overlapped between the domains, learners were able to create category representations for schemata that were distributed across systems and modalities. Therefore, such experts' schema category knowledge was a highly elaborated and distributed set of simulator representations, probabilistically associated in LTM and easily accessible for use in different activities and contexts (LTWM).

LTWM Model (Ericsson and Kintsch 1995)	LTWM Function	Traditional Training Activity	Paivio (2007)	Barsalou (2003a)
LTM Encoding (Semantic/Meaningful Encoding)	Chunk Formation	Solfeggio Partimenti	Imagen and Logogen Acquisition (Representational and Associational Activation)	Simulator Acquisition: Property Simulators
	Chunk Relationships and Associations			
Cues in Retrieval Structure	Cues for Retrieval and Maintenance	Counterpoint	Associational and Referential Activation	Simulator Acquisition: Relation Simulators
Acquiring LTWM	Speeding up LTM Encoding, Storage and Retrieval	Advanced Composition and Partimenti	Fluency in Representational, Associational and Referential Processing	Fluency in Simulation

Storage and Encoding  
↓  
Retrieval

Figure 4.4. Traditional Training Activities as LTWM Acquisition



*Solfeggio.* The first type of training activity that all students undertook was singing or solfeggio practice. As the church was one of the largest training institutions for musicians and required large numbers of singers for services and festivals, the most common type of training was that which prepared students for service in the church. Solfeggio practice was largely focused on developing musical literacy, on imparting fundamental aspects of musical vocabulary and style, which in turn provided a means for conceptualization of pitch and pitch relationships (Gjerdingen 2020, 100). From the perspective of the current project, and in the context of training practices that continued to train maestros, solfeggio practice was a means of developing property simulators for single lines (melodies). While such training was focused primarily on explicit acquisition of such property simulators, aspects of other property and relation simulators—such as basslines, harmony, and counterpoint—were acquired implicitly because melodies were presented in context and not in isolation. As part of the larger agenda for those lucky enough to receive training as composers, solfeggio functioned as an important first step for acquiring relevant melodic chunks, preparing students for partimenti and counterpoint. As was said in Naples, “Whoever can sing, can play” (Baragwanath 2020, 54; see also van Tour 2015, 86).

Baragwanath (2020) discusses several taxonomies of solfeggio and their didactic purposes in 18<sup>th</sup>-century apprenticeship training. Here, I interpret these different taxonomies within the context of building domain-specific memory skill (see Figure 4.5). Students, particularly in the Neapolitan Conservatory tradition, would often begin spoken solfeggio practice—*solfeggio parlato*—before engaging in singing or musical production of any kind (Baragwanath 2020, 14; Gjerdingen 2020, 103). This was a means for students to gain familiarity and fluency with musical notation and syllable reading, involving a long acquisition and

associational period for visual imagens and logogens and the referential connections between them. After this phase, students would progress to type 1 solfeggio, using unaccompanied, single-line exercises designed to teach ‘musical fundamentals’ such as hexachords and rhythm, in both older and newer notation styles (Baragwanath 2020, 249-250). Acquiring these individual lines solidified their auditory, visual, vocal and verbal representations and associations.

LTWM Model (Ericsson and Kintsch 1995)	LTWM Function	Solfeggio Training Type (Baragwanath 2020)	Paivio (2007)	Barsalou (2003a)
LTM Encoding (Semantic/Meaningful Encoding)	Chunk Formation Chunk Relationships and Associations	<i>Solfeggio Parlato</i> (Verbal Solfeggio) Type 1 Type 2	Imagen and Logogen Acquisition (Representational and Associational Activation)	Simulator Acquisition: Property Simulators
Cues in Retrieval Structure	Cues for Retrieval and Maintenance	Type 3	Associational and Referential Activation	Simulator Acquisition: Relation Simulators
Acquiring LTWM	Speeding up LTM Encoding, Storage and Retrieval	Type 3 Type 4	Fluency in Representational, Associational and Referential Processing	Fluency in Simulation

Storage and Encoding  
↓  
Retrieval

Figure 4.5. Solfeggio Training Types from Baragwanath (2020) as LTWM Acquisition

After mastering musical fundamentals, students would progress to type 2 solfeggio, which entailed learning single lines and their application in imitative style, which was vitally important for musical performance in the context of the Church (Baragwanath 2020, 252). This was a means of re-encoding and solidifying single melodic lines learned in type 1 solfeggio in a new context, and provided relational encoding of these lines. Type 2 solfeggio was often identified as a precursor to the study of counterpoint and composition (Baragwanath 2020, 255). Type 3 solfeggio, the most well-known type today, was only undertaken by those in conservatories as part of the progression to professional singer (of secular works, such as opera), and for those fortunate enough to be progressing toward maestro training through partimenti and counterpoint (Baragwanath 2020, 266-267). These solfeggi were often full-length pieces, containing highly elaborated melodies and keyboard accompaniment (figured or unfigured


basses), and often used fast-slow-fast groupings to provide knowledge about compositional practice of set pieces (*ibid.*, 278). These were the most melodically elaborated of the solfeggio types, focusing on the acquisition of such elaborations in popular and secular styles. These pieces allowed students to explicitly expand their representations for melodies, while also supporting implicit learning of basslines and harmonies needed as a precursor to partimenti and composition. In this way, type 3 solfeggio provided students with explicit representation and association of multimodal images and logogens for melodic content, while providing auditory and visual image representations for heard basslines (either played or sung by their teacher) and harmonies. Type 4 solfeggio, the final and most uncommonly used kind, resembled a combined version of type 2 and type 3 solfeggio, employing imitative fugal-like textures with keyboard accompaniment (Baragwanath 2020, 246). The progression through solfeggio of types 3 and 4 shows an increasing focus on re-encoding, reinforcing, and storing patterns learned in types 1 and 2 and on associating these with their co-occurring, collocated basslines and keyboard harmony (Baragwanath 2020, 159).

In summary, learning solfeggio was a means to provide early learners with representations of relevant musical chunks in a melodic context, and their associations in contexts with basslines and harmonies. This focus on the acquisition of relevant musical chunks, in both sacred and secular contexts, prepared musicians for work in the church as vocalists. For those few fortunate enough to study in a conservatory, such as those in Naples or Paris, learning solfeggio was a vital step in musical literacy needed for progression to partimenti and counterpoint. Overall, solfeggio training focused on the acquisition of fundamental musical vocabulary in the form of melodic simulators. Next, students moved onto the explicit acquisition of basslines and harmony through partimenti training.

*Partimenti, counterpoint, and composition.* Many scholars have noted the intricate link between partimenti, counterpoint, and composition training (Sanguinetti 2012, van Tour 2015). Counterpoint training in particular was viewed as a necessary step towards free composition (van Tour 2015, 200; Byros 2015), though the training which provided the transition from counterpoint to free composition is virtually unrecorded. I will discuss training in both counterpoint and free composition due to their interconnected nature. Within conservatory training in both Naples and Paris, the goal of training was to develop the ability to rapidly compose. This is evidenced by contest pieces in the Paris Conservatory, which Gjerdingen (2020) argues functioned explicitly as a probe of memory (see Chapter 14). For such examinations, students were locked in a room for several hours, armed only with a table, chair, blank staves and ink. They did not have access to a keyboard instrument, and were therefore required to "...imagine, correct, and evaluate a composition entirely in his or her own mind" (Gjerdingen 2020, 192). Neapolitan conservatories tested students yearly to ensure that they were skilled enough to remain at the school (van Tour 2015, 80). Training in both partimenti and counterpoint provided requisite skills for such examinations. Here I will demonstrate that training in each discipline progressed from encoding to retrieval activities (see Figure 4.6). Both partimenti and counterpoint training typically began with short, controlled, and concise exercises, and progressed to longer, more complex ones, which I will discuss in turn.

LTWM Model (Ericsson and Kintsch 1995)	LTWM Function	Partimenti Training Type	Counterpoint Training Type	Paivio (2007)	Barsalou (2003a)
LTM Encoding (Semantic/Meaningful Encoding)	Chunk Formation	Rules	Two-, Three- and Four-Part Exercises	Imagen and Logogen Acquisition (Representational and Associational Activation)	Simulator Acquisition: Property Simulators
	Chunk Relationships and Associations				
Cues in Retrieval Structure	Cues for Retrieval and Maintenance	Longer Partimento (Prescribed Conditions, Diminution)	Two-, Three- and Four-Part Dispozizioni	Associational and Referential Activation	Simulator Acquisition: Relation Simulators
Acquiring LTWM	Speeding up LTM Encoding, Storage and Retrieval	Partimenti Fugue	Fugue	Fluency in Representational, Associational and Referential Processing	Fluency in Simulation

Storage and Encoding



Retrieval

Figure 4.6. Partimenti and Counterpoint Training Types as LTWM Acquisition

Typically started after solfeggio, partimenti training was used to teach students bass lines and harmonies at the keyboard, as well as their usage in the context of full pieces. Such training helped students acquire simulators for bass lines, harmonies, and the association between them. It also allowed students to utilize their solfeggio knowledge in a new context, as partimenti realizations required the addition of melodic lines. Therefore, partimenti training functioned to encode new simulators (bass, harmony) and recall previous knowledge (soprano lines), all within motor and auditory modalities via the keyboard. Gjerdingen (2020) highlights the salience of partimenti training for memory by making a direct comparison between partimenti and modern-day jazz lead sheets. He suggests that the purpose of both of these is to serve as a memory aid, helping musicians recall the sound of the composition or something similar in its genre (p. 114). However,

...whereas a lead sheet represents a known composition, a partimento only provides a thread that leads through the phrases, sequences, and cadences of an unknown composition, something that the performer will improvise at the keyboard or write down in a multivoice score. As with a lead sheet, it is the

content of the performer's memory that determines success. One first had to learn a vocabulary of musical patterns that one could recall when prompted by the partimento. The richer those memories, the richer the realization of a partimento (Gjerdingen 2020, 114).

Partimenti are sparser than a lead sheet, and were often novel to the performer (i.e., never seen before), and therefore relied more heavily on prior memory representations and recall for successful musical realization.

In some traditions, the degree of relatedness between partimenti and counterpoint training was very high, as with the school of Durante, which focused on voice leading and the relation of other lines and harmonies above a given bass (van Tour 2015, 127). In other traditions, such as the school of Leo, there was more emphasis on invertible counterpoint (i.e., writing lines above and underneath a given line) which were often used as models for counterpoint and had more connections with solfeggio (van Tour 2015, 172). Within the Durante and related traditions there were different types of exercises. Firstly, there were so-called 'Rules,' shorter models used to teach a particular pattern; such teaching rules ensured encoding of simulators for bass lines and harmonies of various patterns. While the specific orderings of such teaching may have varied across schools, the contents were very similar. These rule types included basic axioms such as consonance/dissonance and cadences, rule of the octave, suspension, bass motions (*bassi del moti* or sequences) and scale mutations or modulations (Sanguinetti 2012, 100).

Once a particular class of rules had been learned, a student moved quickly to application in context using longer partimenti that resembled real pieces. Within the Durante tradition, such partimenti were called 'prescribed conditions' which focused on

the applied application of a particular rule or focused on the solving of a particular compositional problem such as using syncopations (*alla zoppa*), retrograde movement (*chancherziato*), and *ostinato* (van Tour 2015, 132-133). Such exercises were often realized multiple times using varied repetition or *diminuti* practice. This latter involved the direct application of variation technique, and was also applied to counterpoint (ibid., 134; Gjerdingen 2020, 260). *Diminuti* involved varying bass and/or melody, using increased rhythmic activity, making changes to inner voices, and using complimentary rhythms (Sanguinetti 2012, 185). The final type of application in partimenti training was the partimenti fugue, which involved detecting points of imitation. Some points imitation was marked in the score, while others were left to be discovered by the student (Sanguinetti 2012, 194-195). Partimenti fugue was typically undertaken *after* counterpoint training had begun, often concurrently with partimenti training. For example, the content of Fenaroli's fourth partimenti book, which included several partimenti fugue, was meant to be undertaken after counterpoint training had been underway, applying knowledge of counterpoint within the domain of keyboard performance (van Tour 2015, 162). The partimenti fugue represents the culmination of memory skill within the domain of partimenti training.

Counterpoint, often studied at the same time as partimenti, featured a similar trajectory of training exercises, particularly within schools such as Durante, where counterpoint and partimenti were deeply interconnected pedagogically (see Figure 4.6 above). Much like in partimenti training, students would begin with shorter exercises, learning how to write two-, three- and eventually four-voice counterpoint over a given bass. This was a means of re-encoding the skills learned previously, such as soprano lines in solfeggio or bass lines in

partimenti, in a new context where singing or playing was not used; thus, students needed to relearn and recall prior information in a new context. Students then applied what they had learned contextually, using longer pieces called *disposizioni*, which were partimenti basses intended for counterpoint realization. Within the Durante school in particular, partimenti basses were *explicitly* used as models: a student would be given a partimenti bass and they would be expected to compose multiple counterpoints for it, much like the *diminuti* practice in partimento training (van Tour 2015, 25; Gjerdingen 2020, 138). The final type of counterpoint exercise undertaken by students was the fugue, which was viewed as a necessary step toward free composition, albeit a step that is not well documented in existing historical texts (van Tour 2015, 200).

#### Gradual Memory Elaboration Through Progressive Variation

Having shown how the pedagogical activities within the Neapolitan and Parisian Conservatory traditions align with a trajectory for LTWM acquisition, I now turn to particulars of memory encoding. In this section I will show that traditional Galant pedagogies revolved around creating extended, highly elaborated memory episodes in order to represent schema categories. I will show an important role for schema ‘prototypes’ (often referred to as *models* or *movimenti* in partimenti training) in ‘carving out’ a pool of memory episodes for the acquisition of subsequent category knowledge. These early acquired ‘schema prototypes’ establish an initial base of episodic memory pools for each category for later recall, making them frequently revised exemplars (Barsalou 1990). By interacting initially with such paired down, simplified schema models, early memory episodes establish a clear basis for habitual interaction with each category and its features. These episodes then carry forward to future interactions with more complex



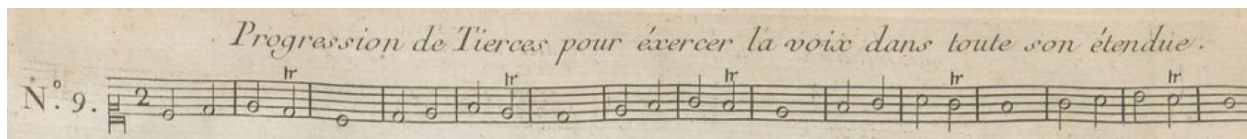
exemplars. This is a claim similar to made by Ashley (2020), who argues that prototypes within the discipline of music theory function as heuristics that guide analytic interaction and discovery.

Such gradual memory elaboration is sometimes replicated in experimental settings within the perceptual and category learning domain. These gradual variation procedures, called transfer along a continuum (e.g., Lawrence 1952; Mackintosh and Little 1970) or easy-to-hard learning (Wisniewski et al., 2017) offer several benefits to both memory and attention. Progressive variation in stimuli ordering positively affects perceptual and category learning by facilitating attention to relevant stimuli dimensions (Chandrasekaran *et al.* 2016; McCandliss *et al.* 2002; McClelland, Fiez and McCandliss 2002; Iverson, Hazan and Bannister 2005; Church *et al.* 2013; Roark and Holt 2018), as well as facilitating incremental representation-based learning (Wisniewski *et al.* 2019).

Such gradual elaboration is inherent in traditional Parisian and Neapolitan pedagogies, where ‘prototypes’ learned early on play important roles throughout expertise acquisition. In each domain of expertise, simple or ‘feature reduced’ models are taught first, followed by progressively more complex or embellished examples. Within solfeggio training, students would begin with Type 1 solfeggi, which are unaccompanied lessons for beginners focusing on *canti firmi* and simple scales and leaps before progressing to more complex and embellished exercises (Baragwanath 2020, 249). In the French solfeggio tradition, a similar approach was adopted: students would begin by learning a simple model, and then would learn increasingly complex exemplars (see Figure 4.7a, b, c). Similarly, when advancing to partimenti practice, rules were learned first, followed by application in longer partimenti, leading eventually to partimenti fugue. The practice of *partimenti diminuti* exemplifies the approach to gradual memory

elaboration: each partimento that was learned in multiple formats, contributing to a highly elaborate memory episode for each exemplar encountered (see Figure 4.8).

(a). Third leaps for exercising the voice in the whole range (p. 3)



(b). Mix of half and whole notes (p. 4)

This musical exercise, labeled N° 12, is titled 'Une blanche pour chaque tems, une ronde pour la mesure entière'. It is written on a grand staff (treble and bass clefs) in 2/4 time. The exercise features a mix of half and whole notes. The upper staff contains half notes, and the lower staff contains whole notes. The tempo is marked 'Moderato'. The exercise includes various ornaments and slurs. The key signature is one sharp (F#).

(c). Mixture of quarter and half notes (p. 6)

This musical exercise, labeled N° 13, is titled 'Deux noirs ou une blanche pour chaque tems'. It is written on a grand staff (treble and bass clefs) in 2/4 time. The exercise features a mixture of quarter and half notes. The upper staff contains quarter notes, and the lower staff contains half notes. The tempo is marked 'Andantino'. The exercise includes various ornaments and slurs. The key signature is one sharp (F#).

Figure 4.7. Three similar introductory solfeggio exercises from Levesque and Bèche (1779)

## No. 1

Style 1

Style 2

Style 3

Figure 4.8. Durante *Partimenti diminuti*, three ways of embellishing the ascending half step in the bass from the leading tone to its tonic (Durante and Gjerdingen, “Partimenti Diminuiti.” From *Monuments of Partimenti*, <https://partimenti.org/>)

There is evidence which suggests that these early models may have been important to conceptualization throughout training and beyond. Gjerdingen (2020) demonstrates that in fine arts training within the *École des Beaux-Arts*, artists would explicitly work from a prototypical sketch forward to increasing detail (see Figure 4.9, p. 277). A similar approach was used in the Paris Conservatory by Bazin and others, where completion of contest pieces required the production of rapid solutions to compositional problems (Gjerdingen 2020, 299). Within the Italian solfeggio tradition, the *amen* and *appoggiatura* rules guided students’ attention toward pitches ‘hidden’ within elaborated melodic lines. When there were multiple interpretations

available for these lines, different solmizations helped students distinguish between the melody's different conventional uses (Baragwanath 2020, 134).

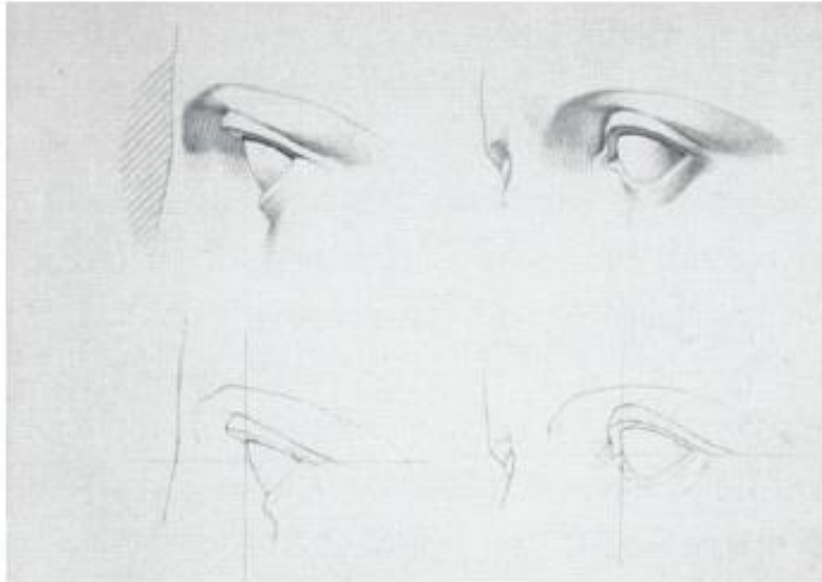


Figure 4.9. Sketch and refinement of two eyes by Bernard Julien (Gjerdigen 2020, 277)

### **Modern North American Music Theory Training: A General Account of Contemporary Galant Schemata Acquisition**

For modern musicians, learning Galant schemata is quite a bit less of a regimented affair than were the training practices in eighteenth-century conservatories. Modern North American musical training differs quite substantially from the eighteenth-century apprenticeship model, and similarly, the goals for modern training differ. Modern theorists may benefit from some of the same domain-specific expertise as Galant musicians in composition, partimenti realization, and related pursuits; however, they do not require the same level of procedural fluency to accomplish typical modern goals of theory and analysis such as schema identification,

hypothesis generation, and causal inference.<sup>62</sup> There are some similarities between the conceptual and habitual tools.—solfège, harmony, counterpoint, keyboard playing, and singing—used by modern theorists and eighteenth century musicians, even though the training practices, timeline, and pedagogical goals differ substantially. Van Tour (2015) notes similarities between the two traditions:

As a long-serving teacher at the Gotland School of Music Composition I have some familiarity with the instruction of composition. Although our students today, of course, write music of a very different kind, the training in some of the basic prerequisites in music theory today is nevertheless quite similar today to that of the late eighteenth century (Van Tour 2015, 24).

Some scholars, however, claim that training in North American universities—particularly as related to more contemporary theories of harmony—is problematic as it has removed a vital element of craft from training, instead focusing on abstract conceptualizations of tonality which removes vital aspects of style from the study of music (Gjerdingen 2020, 323). Most theorists therefore learn Galant schemata through independent study as schemata are not taught as a part of core curriculum.<sup>63</sup> As the level of experience or enculturation in eighteenth-century style differs greatly among those in the schools of music, the level of familiarity with Galant schemata also likely varies widely. Those with moderate familiarity with eighteenth century repertoire—likely learned through undergraduate or graduate level training in music theory—will have memory representations oriented around the cognitive tools used to interact with this repertoire,

---

<sup>62</sup> As such, the current framework holds that the memory expertise (representations and simulation abilities) likely differs significantly between modern music theorists and those trained in the eighteenth century.

<sup>63</sup> Some music theory textbooks have begun introducing Galant Schemata, such as *Open Music Theory* (<http://openmusictheory.com/schemataOpensAndCloses>). However, the material remains entirely introductory, and merely distills information from *Music in the Galant Style*.

including scale degrees, harmony, and counterpoint. However, because Galant schemata are not explicitly taught, any simulators which have been acquired for these schemata will be loosely organized compared to those gained through more enculturation or explicit training. Therefore, for this population, really learning Galant schemata entails a re-alignment of simulators to represent these categories. The probabilistic connections between existing simulators will be modified and strengthened, while new traces, also probabilistically associated, will be added. Over time, simulators in memory will become more tightly bound, creating more structured representations for Galant schema categories.<sup>64</sup>

In addition to the seminal *Music in the Galant Style* (Gjerdingen 2007), there are now resources available to study traditional methods including partimenti (Sanguinetti 2012; van Tour 2015), harmony and counterpoint (IJzerman 2015), and solfeggio (Baragwanath 2020). Gjerdingen (2020) is also a superb resource for many different training practices and contains recommendations for texts and treatises for those interested in studying schemata, including not only the texts above, but also traditional training treatises including *Traité d'accompagnement au piano* by Émile Durand (1892), *Cours complet d'harmonie* by Augustin Savard (1860), *Cours d'harmonie* by François Bazin (1857), *87 Leçons d'harmonie* by Theodore Dubois (1891), *Traité de la fugue* by André Gedalge (1901), *Solfèges d'Italie avec la basse chiffrée* edited by Levesque and Bèche (1772), and *Solfèges du Conservatoire* edited by Edouard Batiste (1865), all of which are available either online (IMSLP) or for purchase. For younger students, Gjerdingen recommends study of practical partimenti texts, particularly the partimenti handbook by

---

<sup>64</sup> The process of re-alignment will likely look different for different learners. For those with a high degree of enculturation in eighteenth-century music, explicitly learning Galant schemata will primarily entail the addition of verbal traces and their referential connections to imagens stored in the nonverbal system. This addition of logogens may modify the base of existing imagens in the nonverbal system (i.e., chunk size, revision of existing traces enhancing features or regions denoted by the associated logogen, etc.).

Giovanni Furno, the modern version of which contains *regole* with clear explanations in both English and Italian (Gjerdingen 2020, 325). This should be followed by study of Fenaroli, Durante and Leo, the texts of which can be found on the *Monuments of Partimenti* website (Gjerdingen 2020, 326).

I have already discussed above how many modern music theory practices (e.g., The Loop) are a type of memory expertise. Here, I will theorize about the processes by which modern theorists both acquire new simulators and re-align existing simulators to represent Galant schemata categories and will demonstrate important roles for schema ‘prototypes.’ Such prototypes provide the means for developing gradual memory elaboration and are actively recalled during simulation through attention and verbalization. In the first section below, I will outline the role that schemata prototypes play in modern acquisition for theorists. Here, the study of simplified models provides a basis for interaction with schemata categories, which helps to modify the probabilistic connections between existing simulators (i.e., re-aligning them). In the second section, I will demonstrate how LTWM is acquired through analysis, one of the means to ensure that schemata simulators come online at the right time and in the correct order. Re-iterating this process helps to develop probabilistic connections between simulator pools *across* different schemata categories, developing so-called ‘top-down’ sensitivity to schemata orderings.

Given the unitary approach to memory organization in the current framework (i.e., no strict separation between semantic and episodic, declarative and procedural memory), each ‘session’ of interaction with Galant schemata is referred to as a memory episode. Categorical ‘generalization’ occurs when the many memory episodes that are acquired begin to blur together (Sadoski and Paivio 2014, 68, see Figure 4.10). When category knowledge is later recalled or

simulated, large portions of previously encoded memory episodes activate at the same time (functionally ‘semantic’ access), providing ‘abstracted’ category knowledge.<sup>65</sup>

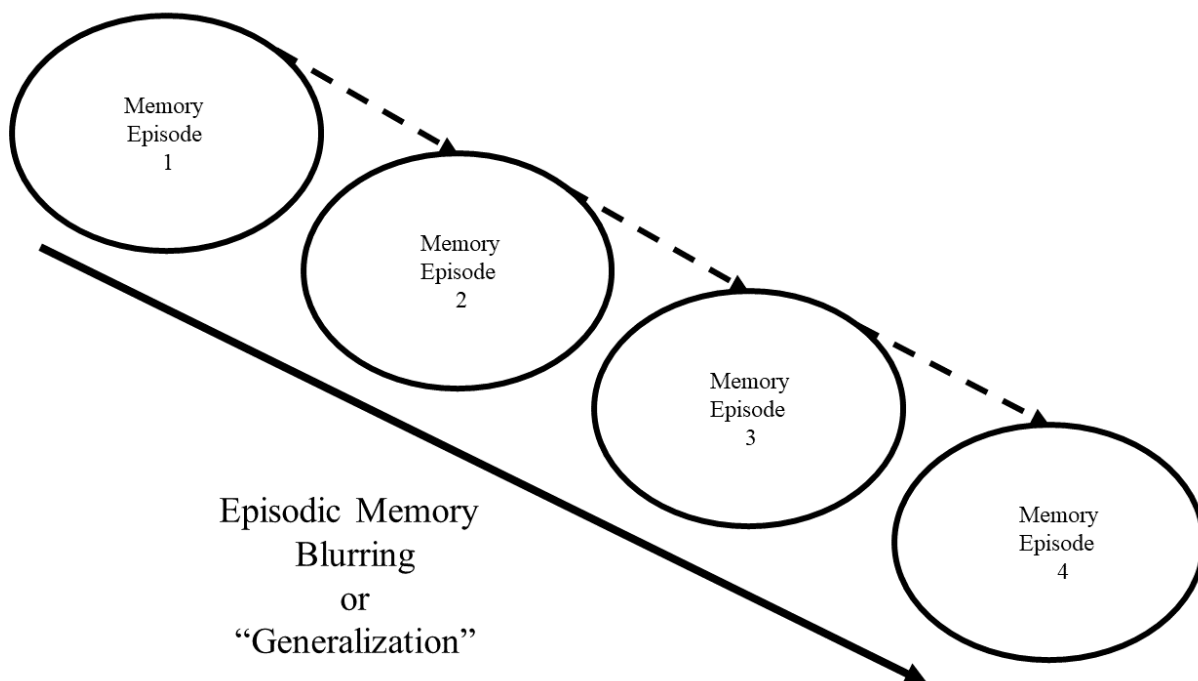


Figure 4.10. Episodic Memory Blurring (“Generalization”) Over the Course of Several Memory Episodes

#### Re-aligning Simulators through Schemata ‘Prototypes’ and Gradual Memory Elaboration

Similar to training practices in the Neapolitan and Parisian traditions, North American music theory training relies on the use of scaled back or simplified models as initial exemplars in the learning process. Below is an example from Laitz (2016), which shows the gradual variation or embellishment of a melodic line (Figure 4.11). Such simplified models serve as foundation for the acquisition of simulators early on in learning, and similarly, are actively recalled (at least

<sup>65</sup> This is differentiated from the recall of a particular episode: for example, the difference between recalling the episodic memory pool for ‘the summer quarter where I learned Galant schemata categories’ versus recalling a larger pool of memory episodes for the Romanesca schemata to complete a given task (e.g., identification in score analysis).



partially) during subsequent encounters with the category. Within the discipline of music theory, such active recall of simplified models is most evidenced by the practice of creating score reductions (see Figure 4.12), a practice common between Schenkerian-inspired North American music theory pedagogical techniques and contemporary Galant schema theory (Rabinovich 2013).

A. B. LN P P

C: I

C. LN P P D. LN P

I V I V I V I I V I V I V I I V I

I ————— V I  
PAC

E. LN P

I V I V I V I —————

Figure 4.11. Example 6.2 From Laitz (2016, 189) showing Stages of Embellishment

1 5

*p*

i ————— V V<sup>7</sup> ————— i

Figure 4.12. Score and Reduction of Beethoven's Piano Sonata in D minor, "Tempest," op. 31, no. 2, Allegretto, from Laitz (2016, 192-193)

It is likely that those trained within North American institutions already possess a fair number of category simulators for melody and basslines, as well as common harmonizations and voice arrangements (i.e., counterpoint). However, since this training is geared around abstract harmonic syntax, such simulators will not be highly structured (i.e., probabilistically arranged) around Galant schemata categories. Therefore, when learning Galant schemata, a modern music theorist with only moderate levels of familiarity with eighteenth-century repertoire will essentially work to re-align pools of simulators to represent schema categories.<sup>66</sup> Note that many theorists refer to such simplified models as 'prototypes,' which include both simplified score-based representations (see Figure 4.13), as well as summary-like presentations of schemata

<sup>66</sup> Again, to reiterate, for those with a high level of enculturation in eighteenth-century style, the explicit learning of Galant schemata categories primarily involves the addition of logogen representations, verbal associations, and referential connections across to existing exemplars in the nonverbal system. If these learners also add in new types of activities (e.g., partimenti realization), then schemata learning also involves the addition of new imagen representations in different modalities.

features using language and symbols (see Figure 4.14).

### Partimenti Prototypes

**Opening**

Romanesca                      Do-Re-Mi                      Sol-Fa-Mi                      Meyer

5 6 5 6 5 6 5 5 5 6 5 5 6 6 5

5 4 3

Figure 4.13. Galant Schemata Prototypes from Open Music Theory

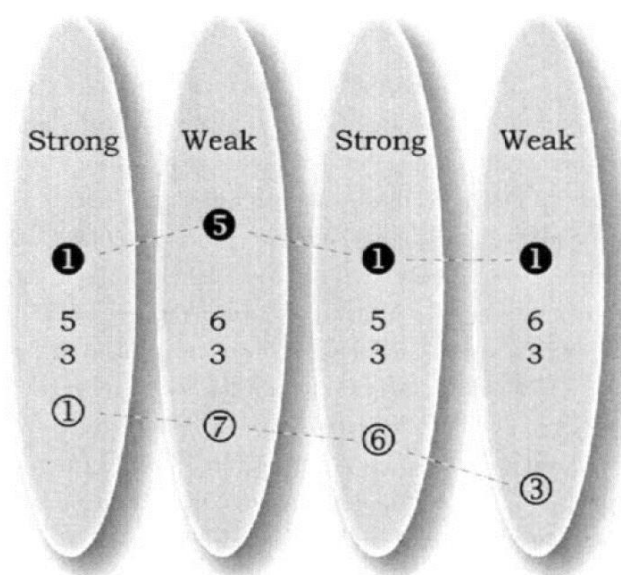


Figure 4.14. Romanesca Prototype from the Prototype Appendix (Gjerdingen 2007, 454).

In the view taken here, there is no fundamental difference between such schema ‘prototype’ presentations and what theorists call ‘exemplars,’ a term used to describe a so-called ‘real’ example of a schema from composed music. Here, schema ‘prototypes’ and ‘exemplars’ are considered to be schema exemplars, as both are stored in memory episodes. Each acquired episode entails encoding and retrieval of imagens and logogens, with selective and co-operative

activation of the verbal and nonverbal systems over time. The difference between studying or interacting with a so-called ‘prototype’ or summary representation is that the interaction is more streamlined. Attention can easily be allocated to relevant features of the presented stimuli: when visual search is facilitated because the visual scene is less cluttered, auditory attention can similarly be allocated efficiently. This ensures that features and relations are encoded and associated in memory. When more difficult exemplars are presented, more WM capacity is required for category identification, which in turn requires more memory expertise and fluency in processing (i.e., automatic activation) of simulators stored in long-term memory. Much like traditional conservatory training, gradual elaboration of memory pools from simplified prototypes to more complex exemplars is encouraged in Galant schemata texts.<sup>67</sup> Theorists however do recognize the limitations of learning only so-called ‘prototypes’ and variations of these, which do is not all that is needed to develop expertise. Baragwanath (2020, 158) notes with regards to solfeggio acquisition, more than learning prototypes is necessary:

For a modern musician seeking to explain this process, it is tempting to assume that solfeggio singers kept the unadorned syllables in mind as some sort of schema prototype and subjected them to processes of variation to generate different types of melodic material. This may well be true of the early stages of training and of straightforward elaborations and repetitive figurations. A *do-re-mi*, for instance, could be filtered through the process of adding leaps of a third with passing notes and appoggiaturas to give rise to a viable cantilena. But such simplistic type-token

---

<sup>67</sup> For example, in *Music in the Galant Style* (Gjerdingen 2007), each schema type is designated its own chapter. Within each chapter, each schema is presented ‘out of context’ in its most simplified form (i.e., ‘prototype), and is then followed by several ‘exemplars’ (i.e., real pieces). While some contextualizing occurs in each schema chapter, contextual training (in the form of full-piece analyses) are generally left until *after* the schema (in isolation) has been overviewed. This is evidenced by the fact that the final chapters of *Music in the Galant Style* (21-29) mostly pertain to analyses of complete works, allowing learners to put to work conceptual knowledge acquired in previous chapters, in the context of ‘real exemplars.’

thinking seems better suited to producing neat reductive analyses than learning how to “speak” Galant melody. Conjuring up sophisticated musical discourses in real time, as performers and composers are known to have done, requires more than a mental store of schema prototypes and a toolbox of variation techniques. Like spoken language, it demands an extensive vocabulary of real words and phrases that can be instantaneously adapted to suit any situation.

As I will show below, learners are still required to learn longer exemplars (i.e., full pieces) in order to acquire representations for typical schema orderings and their relationship to other non-adjacent probabilities (e.g., form). However, the study of schemata ‘out-of-context’ still provides benefits to feature and relation encoding for those categories, in the same way that studying individual words (e.g., spellings, conjugations, etc.) is a necessary, but not comprehensive, part of attaining literacy in language.

Consider a modern music theorist, perhaps a graduate student, learning Galant schemata for the first-time using *Music in the Galant Style* (Gjerdingen 2007) as their resource. I will outline four hypothetical memory/learning episodes and resulting representations for such a learner. As ever, each memory episode will include sequentially ordered chunks for each activity completed in that episode. One such memory episode, diagrammed in Figure 4.15, shows a learning sequence for our hypothetical theorist as they interact with the appendix and some examples in *Music in the Galant Style*.

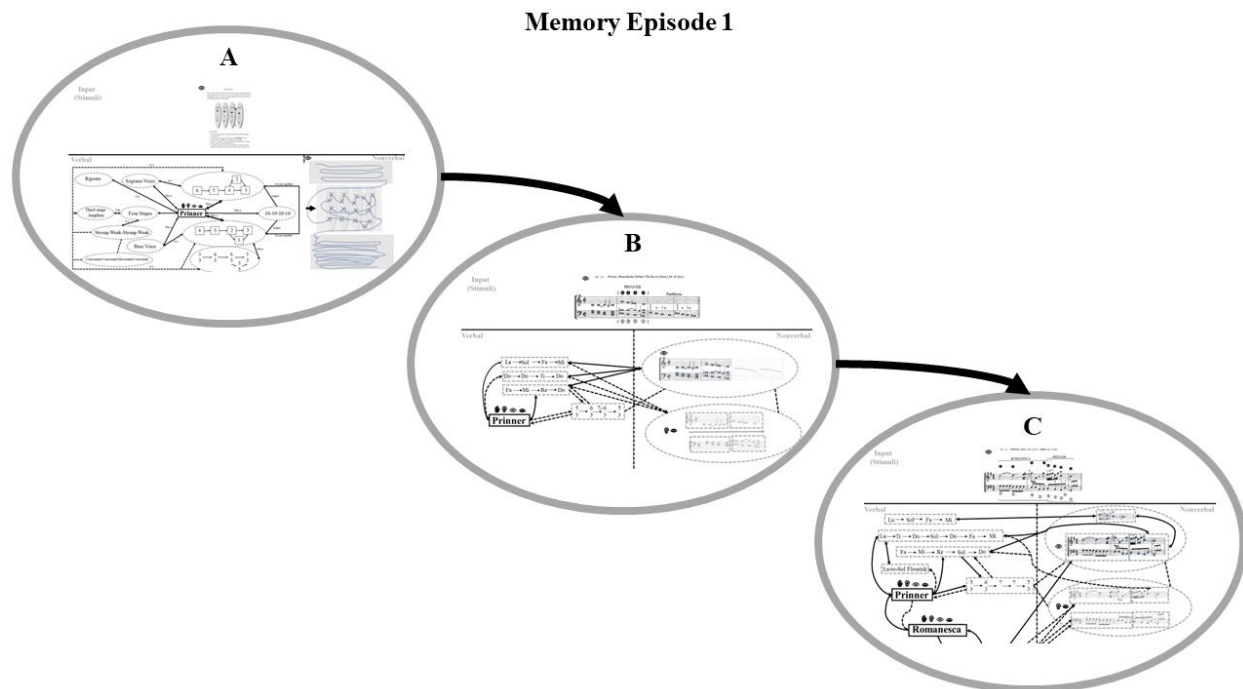


Figure 4.15. Three Learning Interactions with Music in the Galant Style (2007) in Memory Episode 1

In the first part of the learning process, the learner examines the prototype for the schema in question (see Figure 4.16), which primarily involves logogen acquisition, but also imagen acquisition on the nonverbal side, representing the spatial interaction with the text (i.e., where the text is located in the visual field).

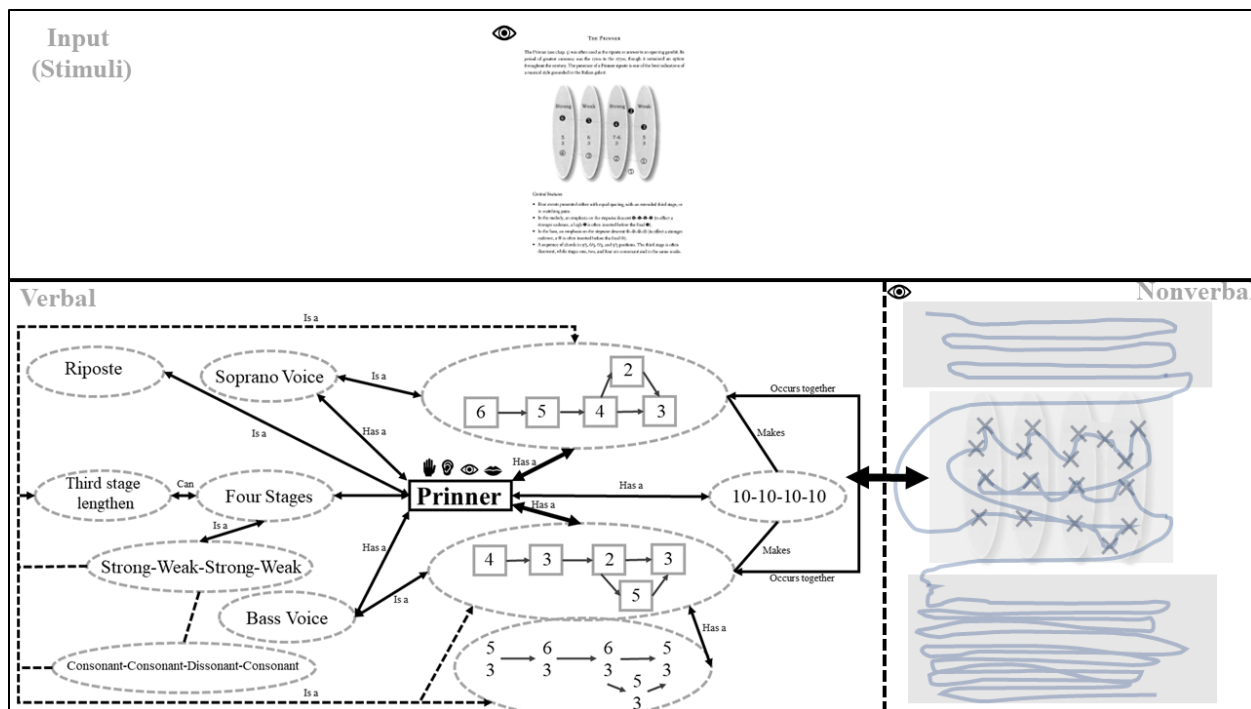


Figure 4.16. Visual and Verbal Interaction with Prinner Prototype from Gjerdingen (2007)

Following this, assume that the learner interacts with the Prinner chapter examples. They begin with the ‘prototype’ in score form (see Figure 4.17), which involves logogen encoding and recall for scale-degrees and figured bass,<sup>68</sup> along with encoding the visual imagen of the score in the nonverbal system, which is in turn accompanied by a degraded auditory imagen formed through subvocalization of the excerpt. Following this, the learner interacts with the example in the text, which contains both Romanesca and Prinner schemas. As with their interaction with the initial Prinner score ‘prototype,’ their interaction here involves logogen encoding and recall, as well as imagen encoding of the score and subvocalized bass and soprano voice for both Romanesca and Prinner categories. As a result, pools of simulators are encoded for both category types, as well as a relation between them, shown on the verbal side as a more unidirectional probability

<sup>68</sup> Not shown here is the fact that scale degrees in numeric form would also be shown. The assumption with the current figure is that the learner saw the scale degree numbers, but recoded them, using subvocalization, into moveable-do solfège.

(Romanesca→Prinner), and on the nonverbal side as sequentially organized chunked versions of the simulators for each category in visual and auditory modalities. Note that the Prinner contains a new ‘feature,’ the La-to-Sol flourish, which is encoded the verbal system as the label and solfège sequence (La-Ti-Do-Sol). Similarly, the bassline differs from the initial prototype, with the addition of Sol to form an imperfect authentic cadence.

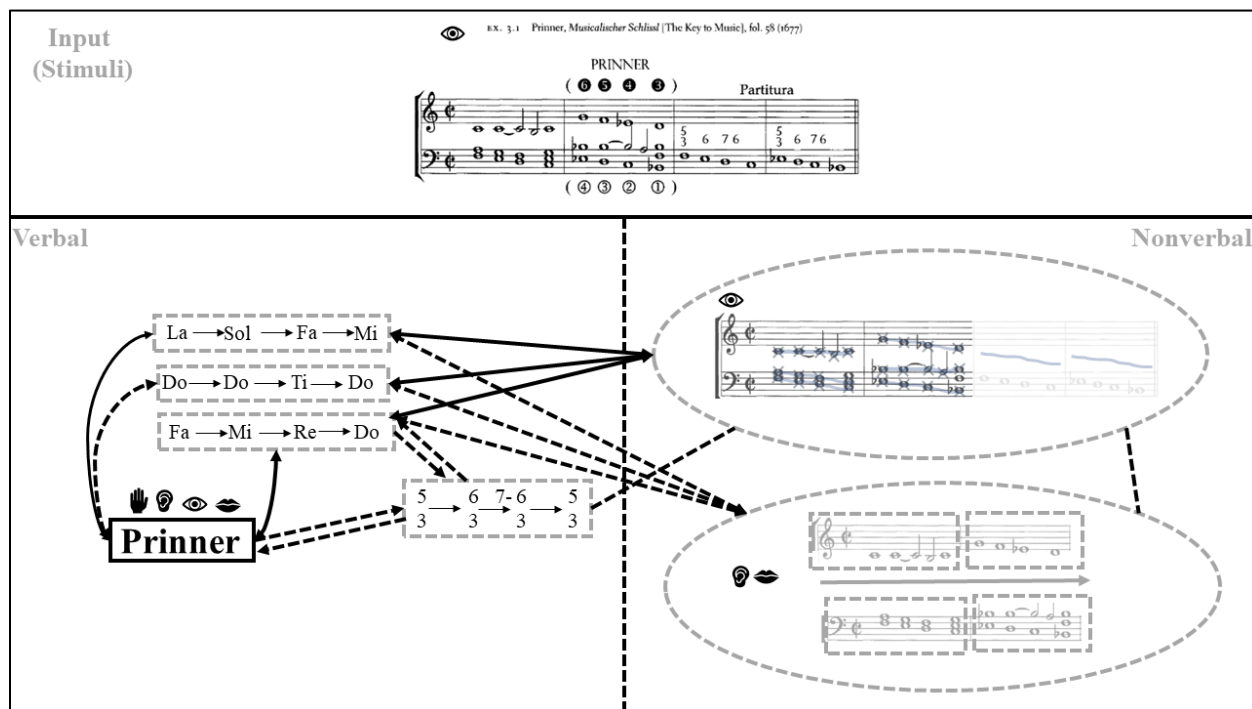


Figure 4.17. Interaction with Example 3.1 from Gjerdingen (2007)



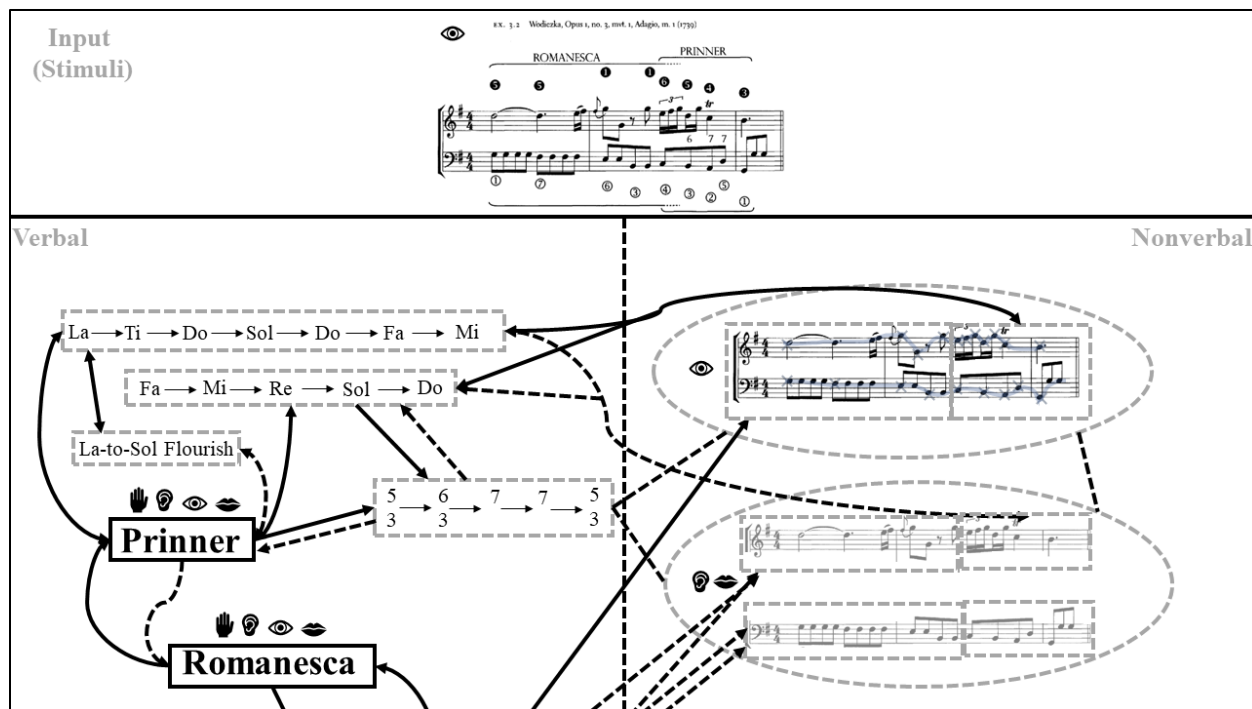


Figure 4.18. Interaction with Example 3.2 from Gjerdingen (2007)

Simple temporal proximity of these memory chunks (A, B, C) in the first memory episode helps to associate together the information encountered so far. However, temporal proximity is not enough to ensure that the information stored in each chunk within the episode is tightly associated, or even retained over time; part of each episode will inevitably fade over time if not reactivated. Given that each example in each chunk is *vastly* different from others in terms of surface similarities, it is unlikely that simply presenting an example like the Romanesca-Prinner example would activate traces in the Prinner prototype pool through representational activation alone. To encode information in newly acquired episodes in a manner that increases the likelihood of associational and referential processing, and activates previously encoded traces in episodic pools, requires simulating aspects of previously acquired ‘prototypes’ when interacting with a new exemplar to create memory traces that are more similar to previous interactions. This is most concretely accomplished by creating a reduction of an exemplar to

resemble the prototype. In the absence of actively creating such a reduction, the reduction can be *simulated* through adjusting visual attention and subvocalizing only the solfège and pitches that pertain to the ‘prototype,’ much like the *amen* practices in traditional solfeggio practice. The addition of such a trace is shown in Figure 4.19, where the learner creates a new trace by fixating visual attention on the primary stages of the Prinner schema while skipping over the other embellishing tones, including referential connections to the corresponding solmization.<sup>69</sup> The addition of this interaction increases the likelihood of reactivating traces with similar structure acquired previously. These include verbal association between different “La-Sol-Fa-Mi” logogens, visual attention and fixation in imagens, and referential connections between them (see Figure 4.20). Such co-activation of traces across chunks helps to integrate category knowledge acquired within and between memory episodes.

---

<sup>69</sup> The interaction may also simply revise the existing visual trace such that those fixation points are enhanced compared to the rest of the trace. A new set of traces could also be created, particularly because the verbal behaviors are differentiated. A similar effect is achieved in either case.

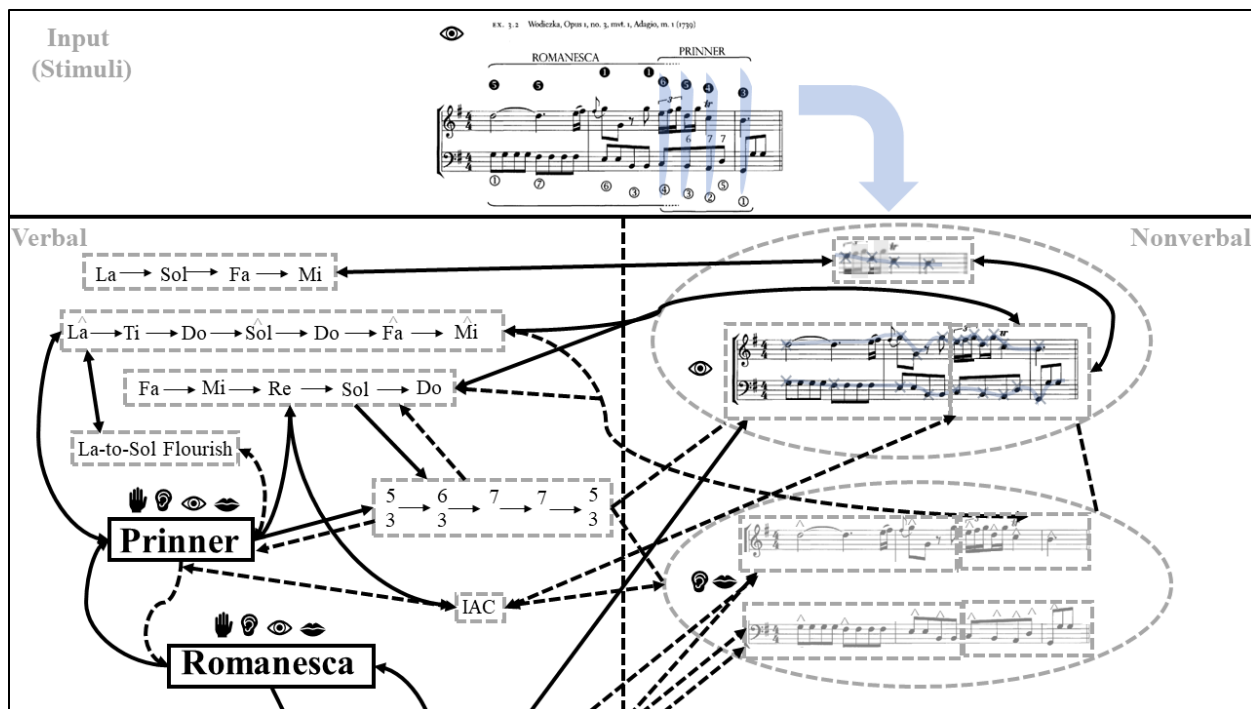


Figure 4.19. Visual Fixation on Prinner Prototype Stages and Association with Prototypical Prinner Logogen (La-Sol-Fa-Mi)

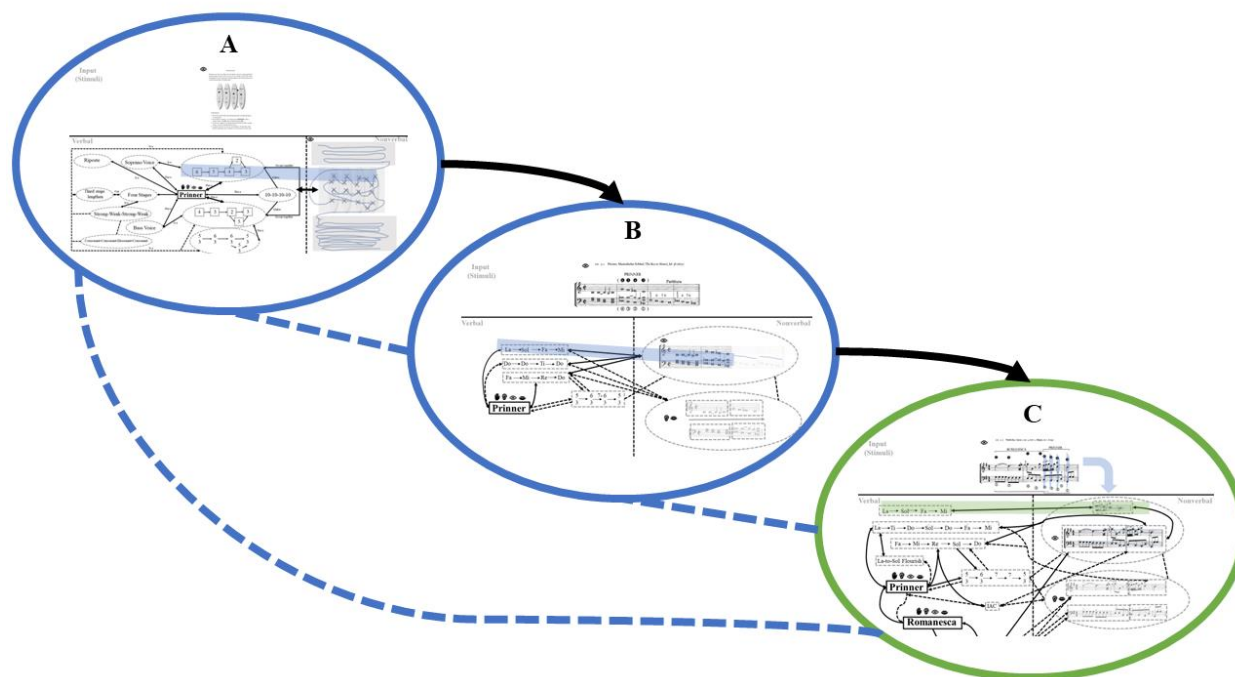


Figure 4.20. Addition of La-Sol-Fa-Mi Logogen and Visual Fixation Activates Similar Interactions in Previous Memory Chunks

Subsequent memory episodes may also be structurally quite different, containing different re-activations or new encodings of imagens and logogens in different modalities and in different orders. Consider a hypothetical second learning episode, completed sometime after the first. The learner completes three different activities (see Figure 4.21): they review the summary information previously learned (A), create a score prototype by hand utilizing this information (B), and finally perform this newly created example on the piano (C).

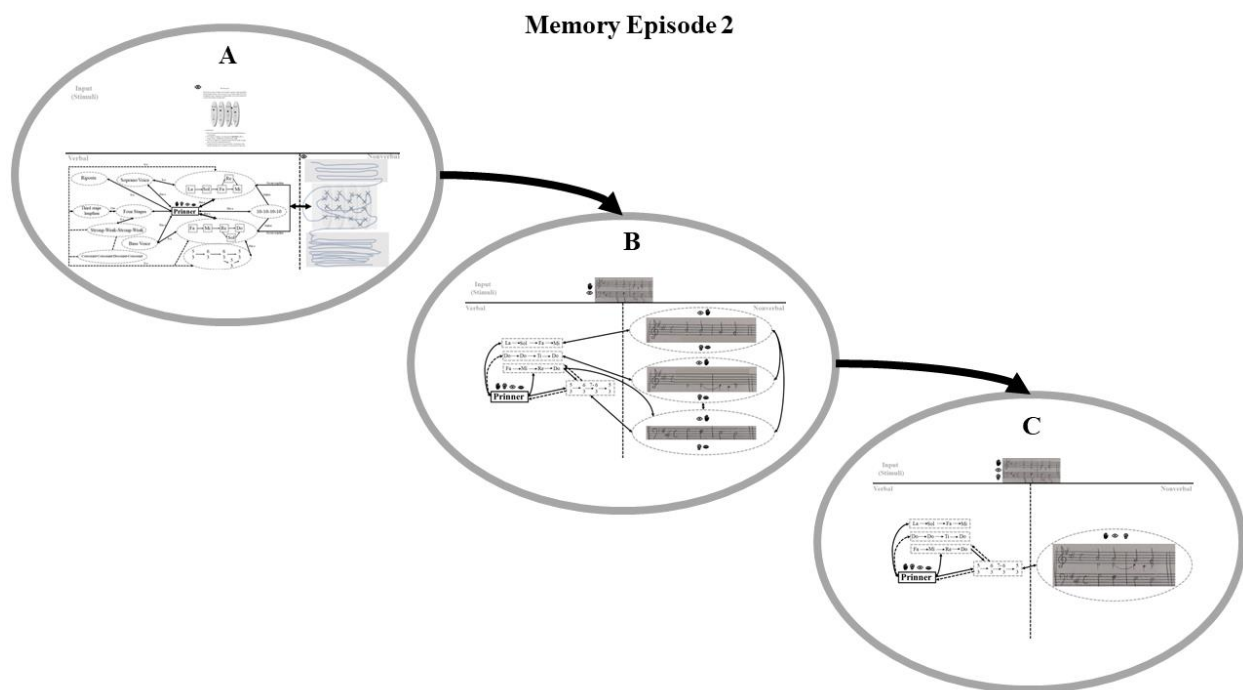


Figure 4.21. Memory Episode 2 Involving Review and Application of Previously Learned Material

After the initial review of Prinner features has been completed (which re-activates some of the traces from the previous episode), the learner uses the verbal information to guide creation of a simple exemplar in score format (see Figure 4.22). Each individual line, bass, melody, and alto voice, are recalled and instantiated in the new context, concurrently using solfège to direct the writing. Lastly, the learner plays the completed excerpt on the piano, which creates new

associated auditory, motor, and visual traces, which have direct referential connections across to the verbal system but do not actively form a connection to the solfège syllables (see Figure 4.23).

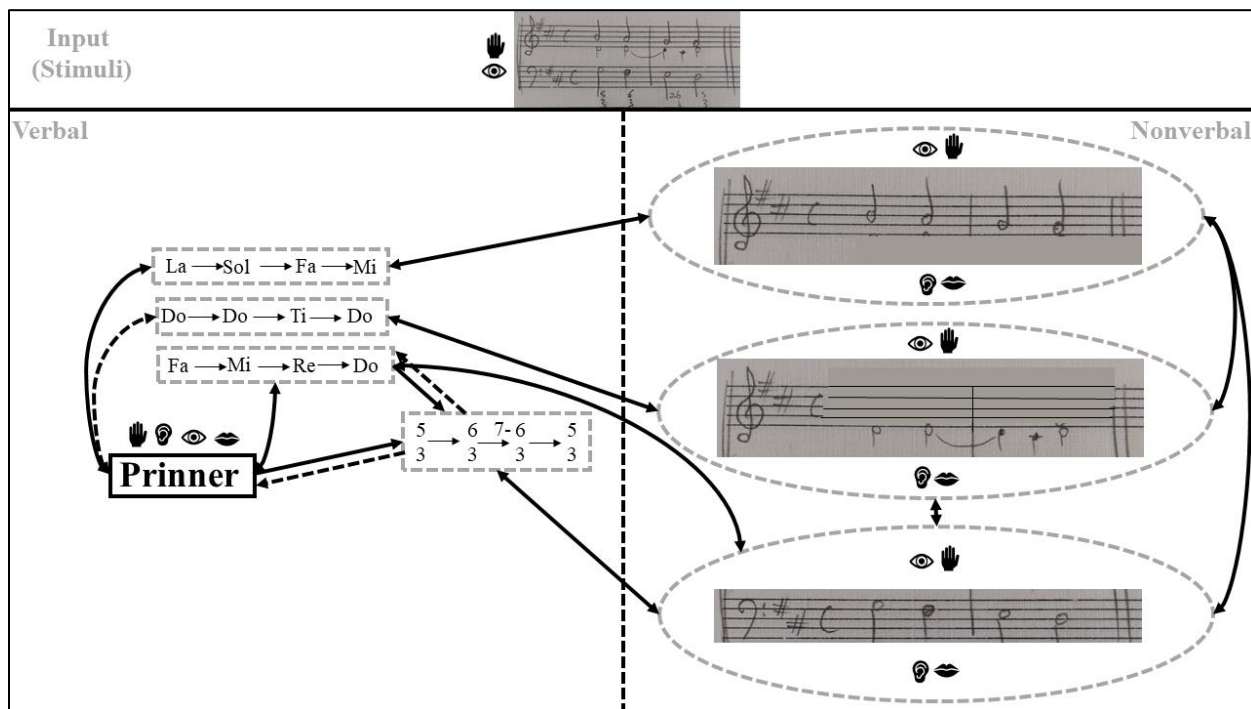


Figure 4.22. Use of Stored Prinner Logogens to Direct Writing Out of Prinner Schema

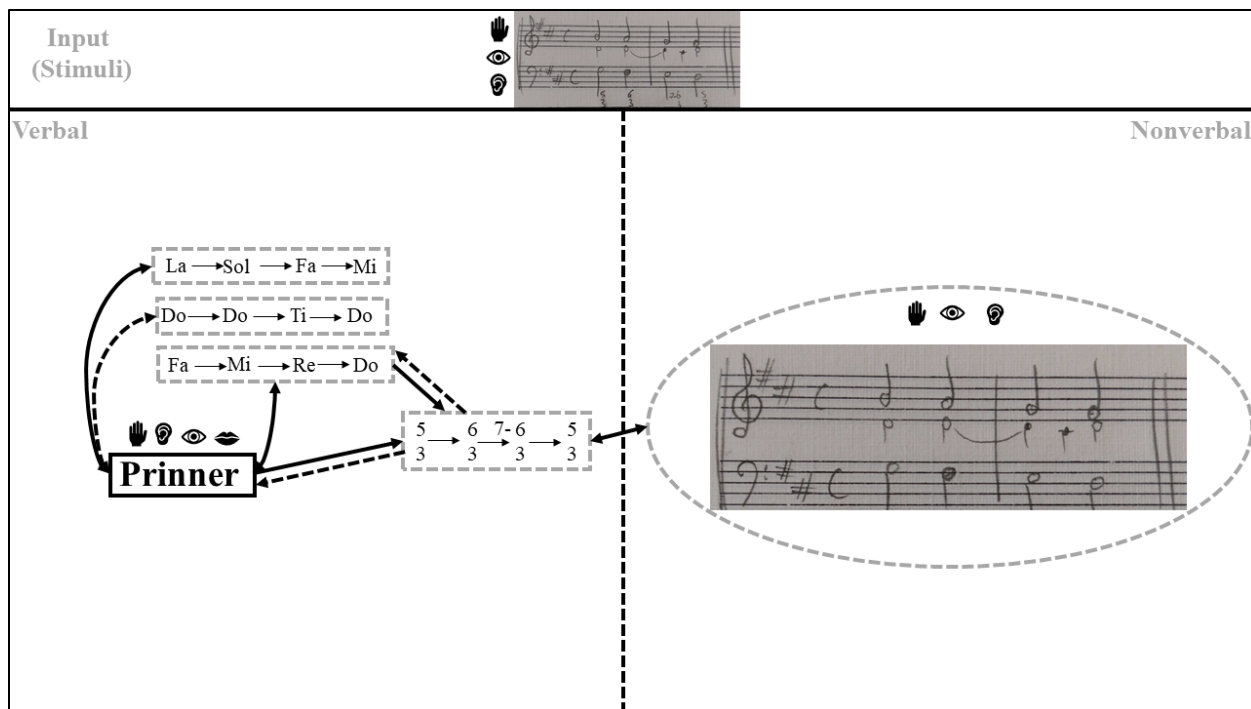


Figure 4.23. Memory Trace for Newly Created Prinner in Auditory, Visual and Motor Modalities (Piano Performance)

Subsequent learning episodes may result in a significantly different set of memory traces within and across verbal systems. With each new interaction, new traces are added and some of the previously acquired traces are re-activated through representational, associational, or referential activation both within and across verbal and nonverbal systems. As one final example, consider two additional episodes in an initial learning phase for Galant schemata. The learner begins by reviewing the previously created exemplars from the last learning session, recalling previously encoded traces (e.g., piano playing) and perhaps adding in new traces for verbalized or sung interactions (Figure 4.24, A). Following this, the learner inputs the exemplars into a computer notation program, creating a completely new set of traces (Figure 4.24, B).

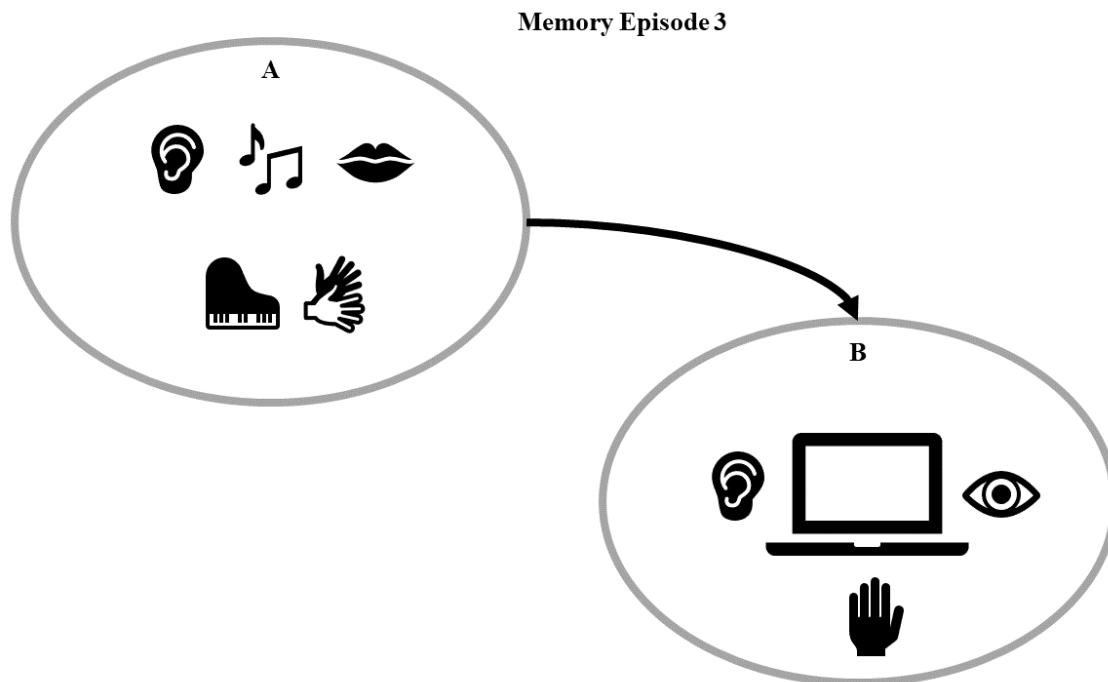


Figure 4.24. Memory Episode 4 Involving Recall and Creation of New Traces

The next memory episode involves listening to the created examples at random to improve schemata identification, primarily reactivating prior traces (Figure 4.25). Over time, some part of each episode may be entirely forgotten, while others are strengthened or revised, either through re-activation or through the addition and consolidation of similar traces across system and mode. Through this process, the initial episodic components of the memory representations blur together to form generalized knowledge (Figure 4.26).

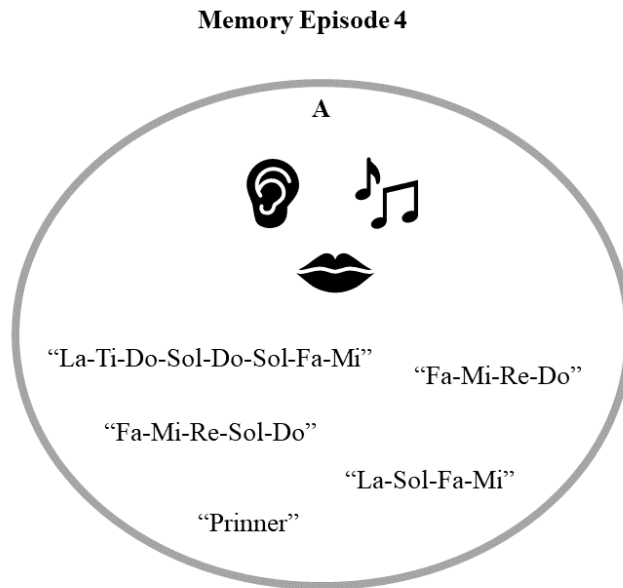


Figure 4.25. Schema Recall Practice in the Auditory Modality

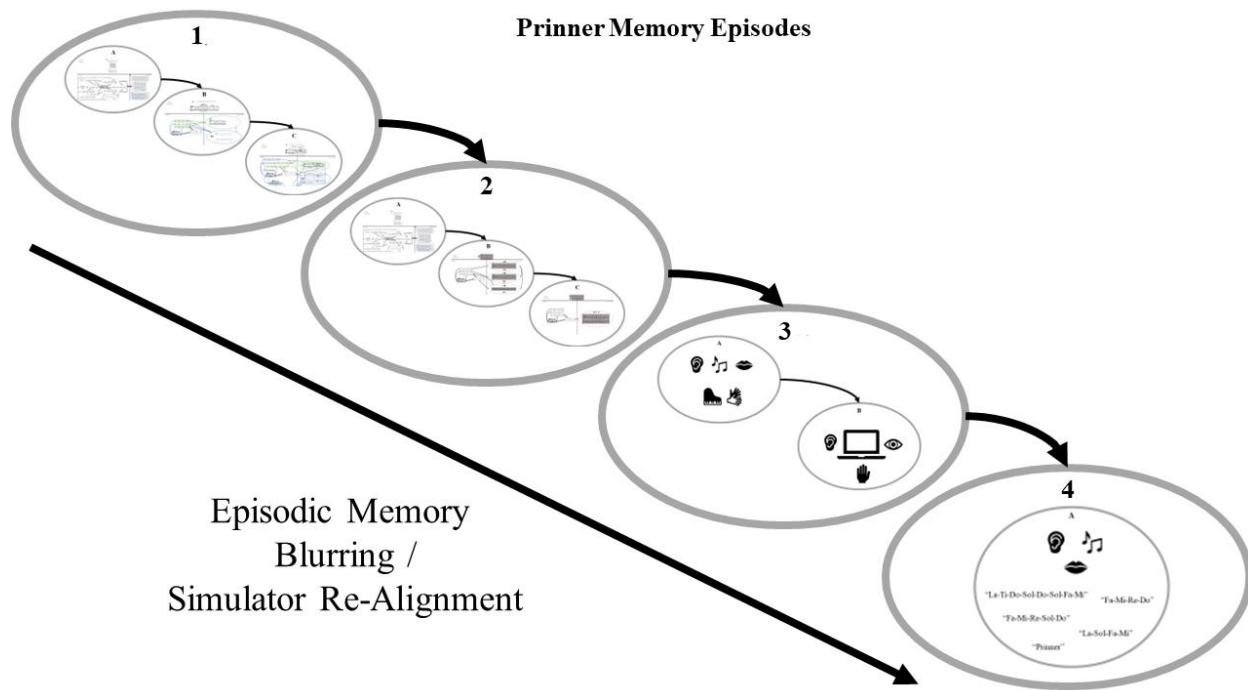


Figure 4.26. Episodic Memory Blurring Across the Four Memory Episodes



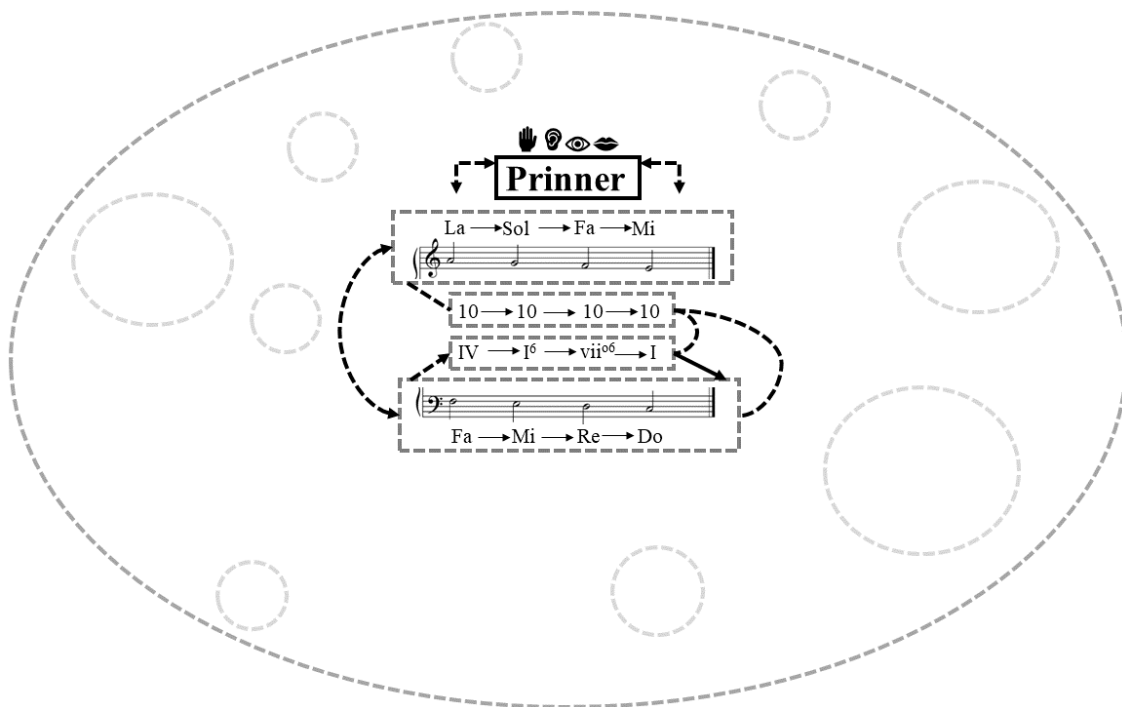
This process is iterated, with Figure 4.27 illustrating the process of simulator alignment on a larger scale. Initially, there may be relatively few memory episodes where simulators are associated in a category interaction (Figure 4.27a), as there are relatively few memory pools within the Prinner category.<sup>70</sup> The low probability between simulators for the Prinner category is shown by the dashed lines between simulator types (melody, bass, harmony, counterpoint). As more interactions with the Prinner category occur, more memory episodes are added, which re-activates parts of previously acquired episodes (Figure 4.27b). This helps to increase the probability between simulators and may result in the addition of a consistent interoceptive response. This is illustrated as a stability interpretation but could also be experienced as increased vividness and FOK in working memory. Once exemplar saturation has been reached,<sup>71</sup> probabilities between simulators is highly probabilistic (Figure 4.27c).

#### (a). Novice Prinner Representation

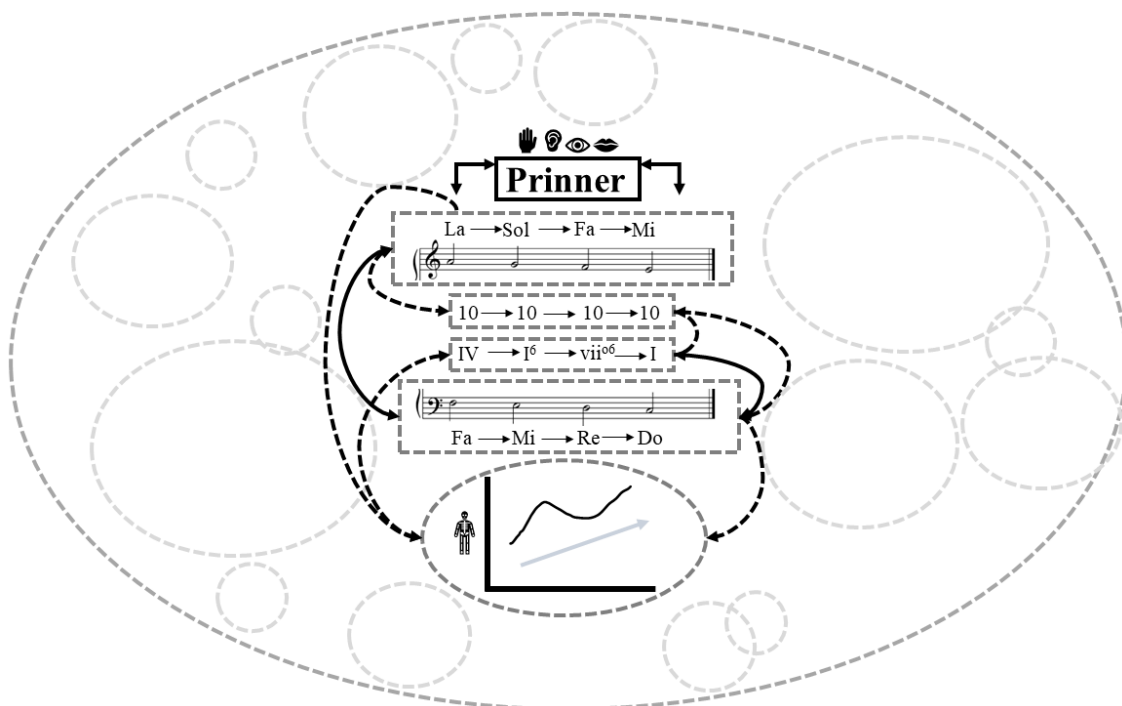
---

<sup>70</sup> The representation of the Prinner traces here is purely symbolic and does not represent how the categories would be stored neurally. Episodic memory traces would contain multiple types of category information, such that activation of Prinner category knowledge would result from partial activation of these traces; traces would not be centralized in the manner presented here.

<sup>71</sup> The process of reaching the required number of exemplars for expert fluency (Abbot-Smith and Tomasello 2006).



(b) Intermediate Prinner Representation



(c). Expert Prinner Representation

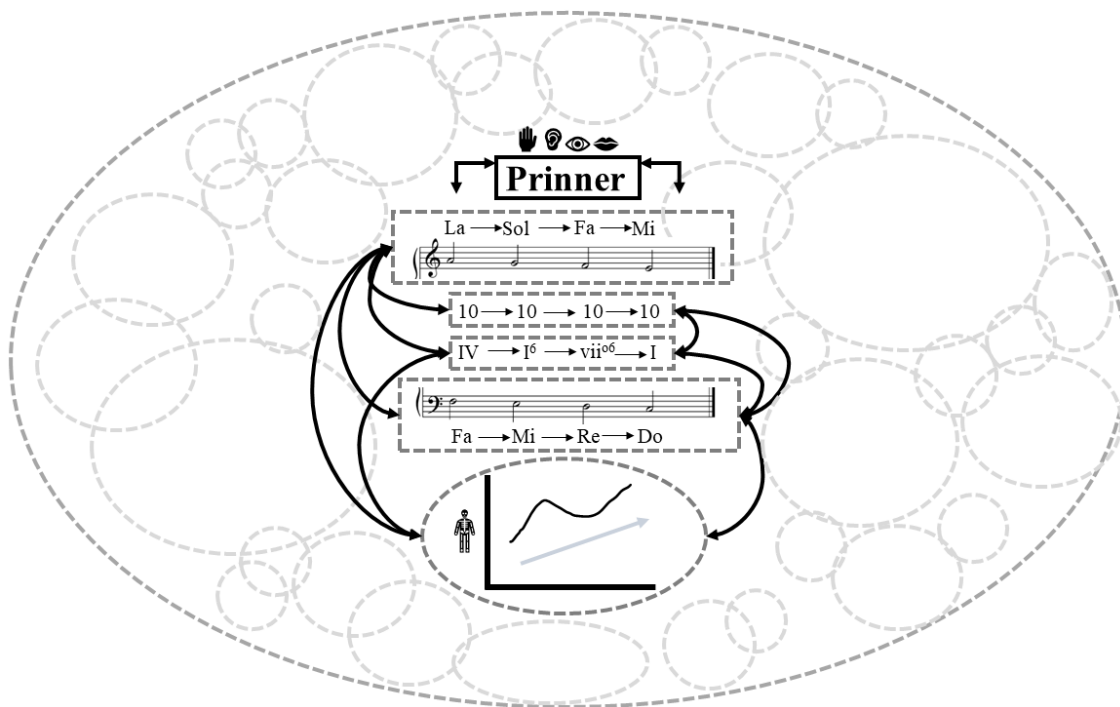


Figure 4.27. Category-Level Pinner Traces for Novice (a), Intermediate (b) and Expert (c) Levels

A similar process occurs for different sub-categories of Pringers, including sequential and cadential types (Caplin 2015, see Figure 4.28). Different regions develop with different types and probabilities between simulators: cadential Pringers, for example, use particular bass and harmony arrangements, have a higher likelihood of other features, such as the high  $\hat{2}$  motion in the soprano. As each pool of simulators is acquired and becomes more highly structured, it may develop related but differentiated interoceptive responses. As features and relations are more fluently and repeatedly processed, pools are better distinguished from one another.

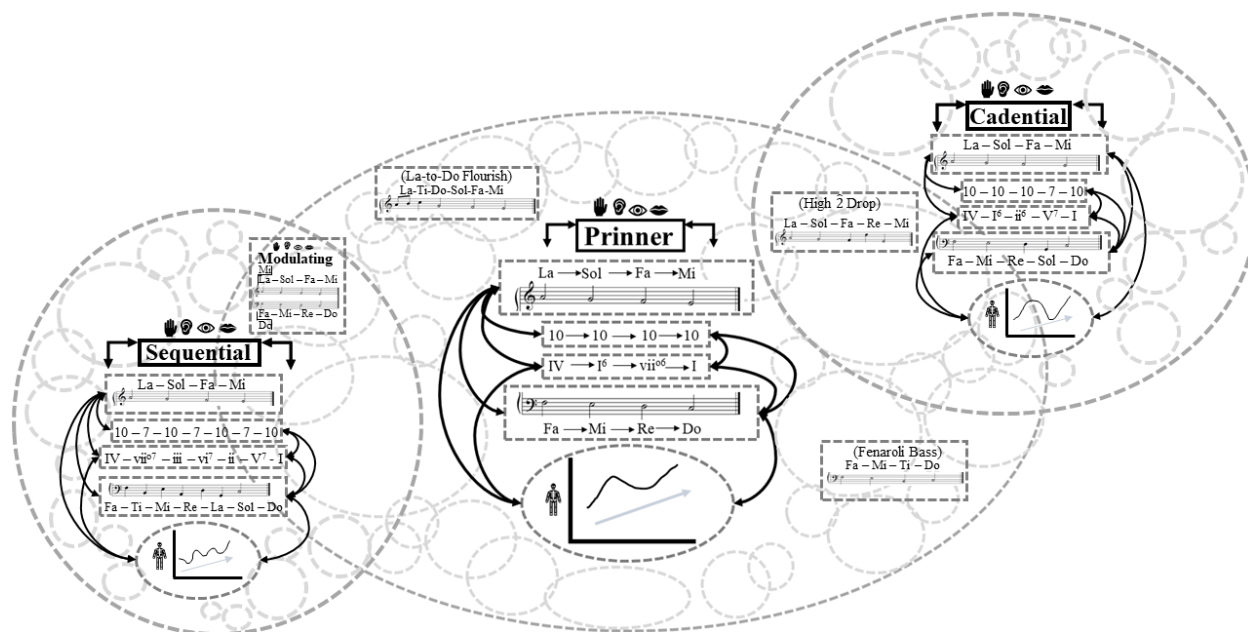


Figure 4.28. Category Level Representation for Prinner Subtypes

### Acquiring LTWM Through Analysis

Unlike students in Parisian and Neapolitan conservatories—whose schemata expertise was distributed across several different domains (solfège, partimenti performance, composition)—modern theorists’ memory expertise is primarily acquired in the domain of musical analysis. The process of studying complete compositions and performing explicit categorization of various sorts entails active simulation, where simulators for schemata categories are brought online in the right order, and at the right time. This process focuses primarily on recall but also involves new encoding, particularly in cases where schema presentation is unfamiliar and the analyst has to process and store such exemplars into memory. The use of multiple music theoretic concepts in the process of analysis, such as phrase and formal analysis, helps to form associations between features and concepts, developing distributional learning, or sensitivity to the probability of different schema categories over time.

To demonstrate this, I will illustrate a hypothetical analytical process for learning the schemata in K. 545 as outlined in *Music in the Galant Style* (Gjerdingen 2007).<sup>72</sup> The intended goal for the sessions is to hear the presented formal analysis. The process is essentially identical to learning individual schema: each practice session (here, called ‘analysis session’) creates memory traces for a series of interactions, which results in pools of memory episodes. Here they will all relate to different interactions with the same piece over time. Once multiple memory episodes have been acquired, these begin to merge into generalized knowledge about a particular piece of music. I will outline four different analysis sessions (see Figure 4.29): the first where the analyst reviews the analysis from Gjerdingen (2007) through visual inspection, the second where the analyst completes a formal analysis on the score and through listening, a third session where the analysis from the first two episodes is recalled and applied in score analysis, and a final session where the analyst recalls the form and schemata while listening.

---

<sup>72</sup> The process is based loosely off the author’s personal experience learning Galant schemata.

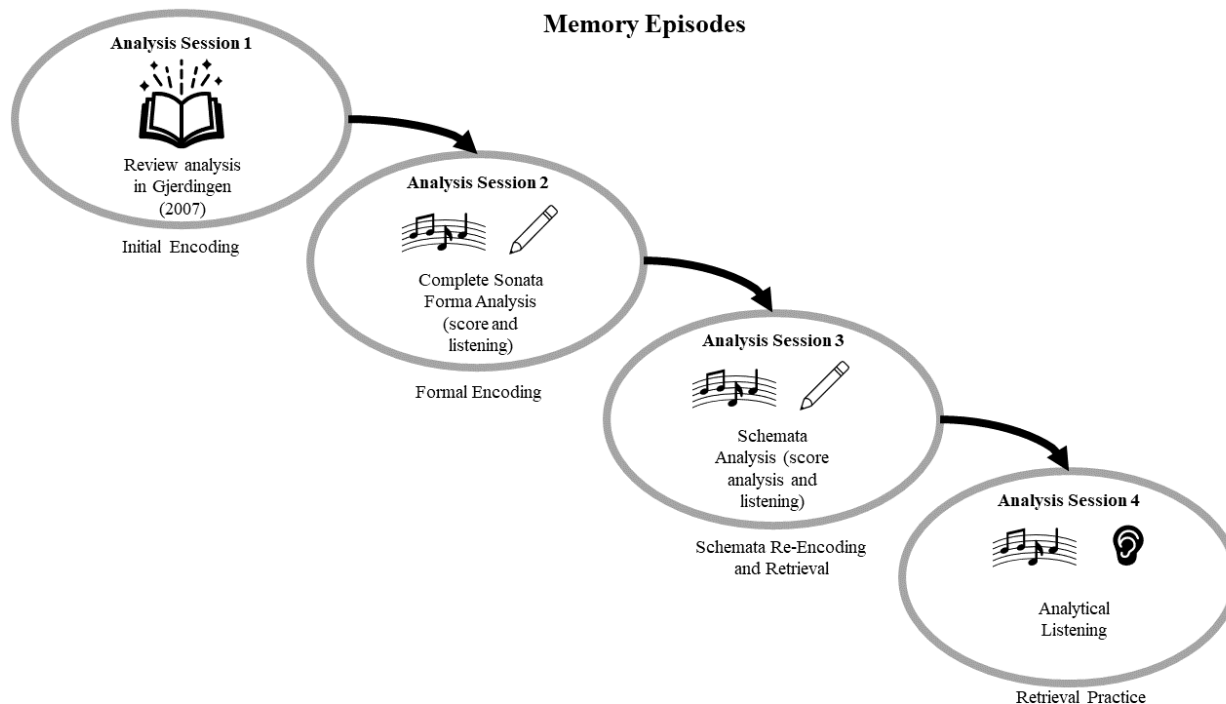


Figure 4.29. Four Analysis Sessions for K. 545

In the first session, the theorist reviews the analysis of K. 545 from Gjerdingen (2007). This involves two distinct interactions, one with the schemata in list format (*ibid.*, 364, see Figure 4.30), and one with the annotated score (*ibid.*, p. 365-368, see Figure 4.31). These interactions use the verbal system (recall), along with visual and verbal encoding of the score. In the first interaction with the schemata list, the theorist constructs a memory trace that involves the acquisition of verbal logogens, organized sequentially, whose spatial layout is encoded in the nonverbal system (Figure 4.30). This is followed by reviewing the schemata analysis presented in score format (Figure 4.31), encoding a similar sequentially ordered set of logogens in the verbal system, associated spatially with the score *imagen* stored on the nonverbal side. Also encoded are the interactions that the theorist has with each schema during visual inspection, such as inspection of scale degree labels and their configuration (e.g., counterpoint). Since the theorist is familiar with the piece, this encoding process may also include partial recall of auditory

imagens of the piece during visual inspection, which may weakly encode some auditory features of the stimulus into the episode.

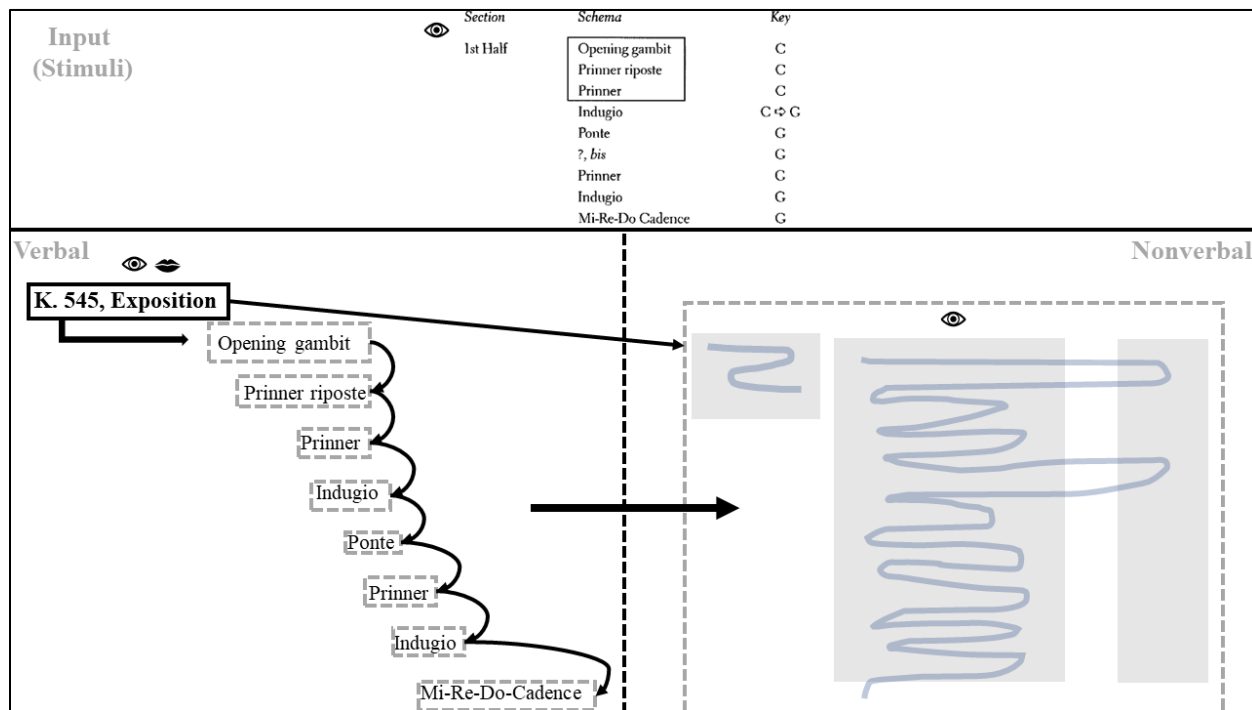


Figure 4.30. Review of K. 545 Schema Chart from Gjerdingen (2007)

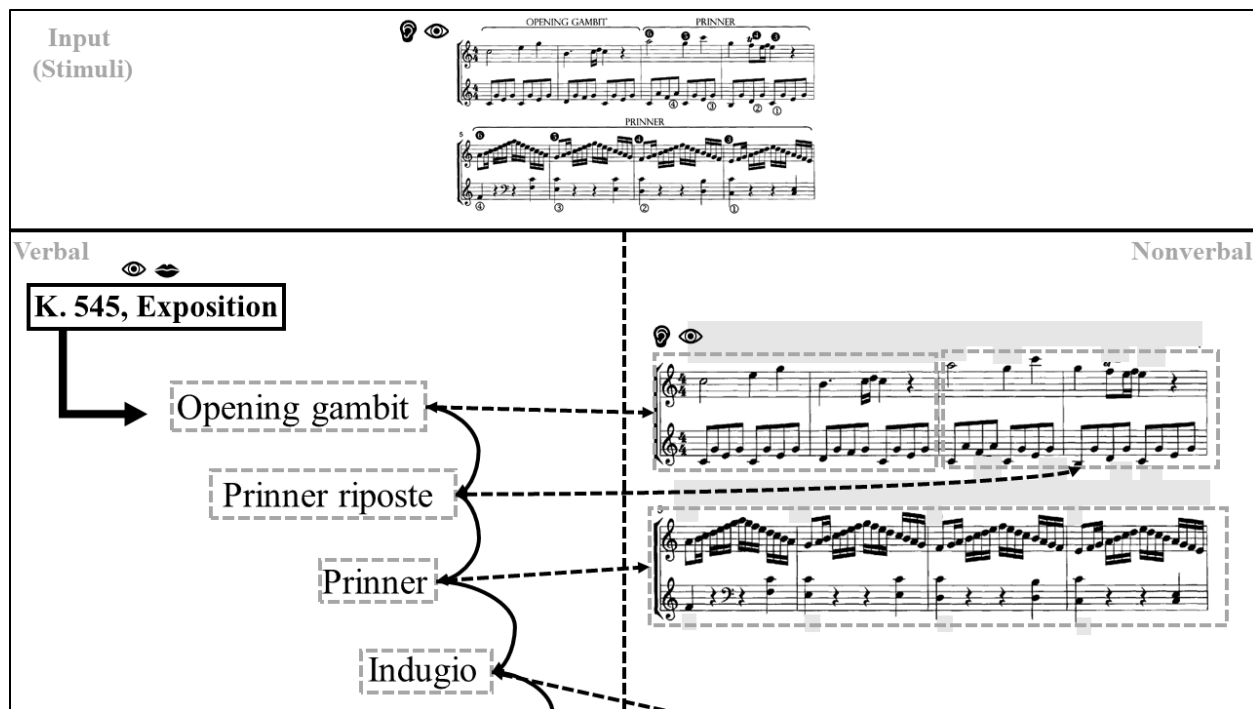


Figure 4.31. Review of K. 545 Score Analysis in Gjerdingen (2007)

The second analytical session, undertaken sometime after the first, involves recall and application of existing formal knowledge through score analysis and listening (see Figures 4.32 and 4.33). Such interactions create new traces and re-activate existing traces, re-aligning and associating them with previous knowledge structures.



Exposition  
 (P) *(mf)*

BI — CI — CI

P-TR — Continuation

(TC) → *(cresc.)*

Overlaid arrival I:HC MC CF

(S) *(p)*

The image shows a handwritten musical score analysis for K. 545, consisting of five systems of piano notation. The first system is labeled 'Exposition' and '(P)', with a dynamic marking of '(mf)'. It features a treble and bass staff with a melodic line in the treble and a rhythmic accompaniment in the bass. Above the staff, there are handwritten annotations: 'BI' with a bracket over the first two measures, and 'CI' with brackets over the last two measures. The second system is labeled 'P-TR' and 'Continuation', showing a more complex melodic line in the treble staff. The third system is labeled '(TC) →' and '(cresc.)', indicating a crescendo. The fourth system is labeled 'Overlaid arrival' and 'I:HC MC CF', with a circled 'I:HC MC' and a dynamic marking of '(p)'. The fifth system is labeled '(S)' and '(p)', with a circled 'S' and a dynamic marking of '(p)'. The score includes various musical notations such as notes, rests, and dynamic markings, along with handwritten annotations and brackets indicating structural divisions.

Figure 4.32. Formal Score Analysis of K. 545 (by hand)

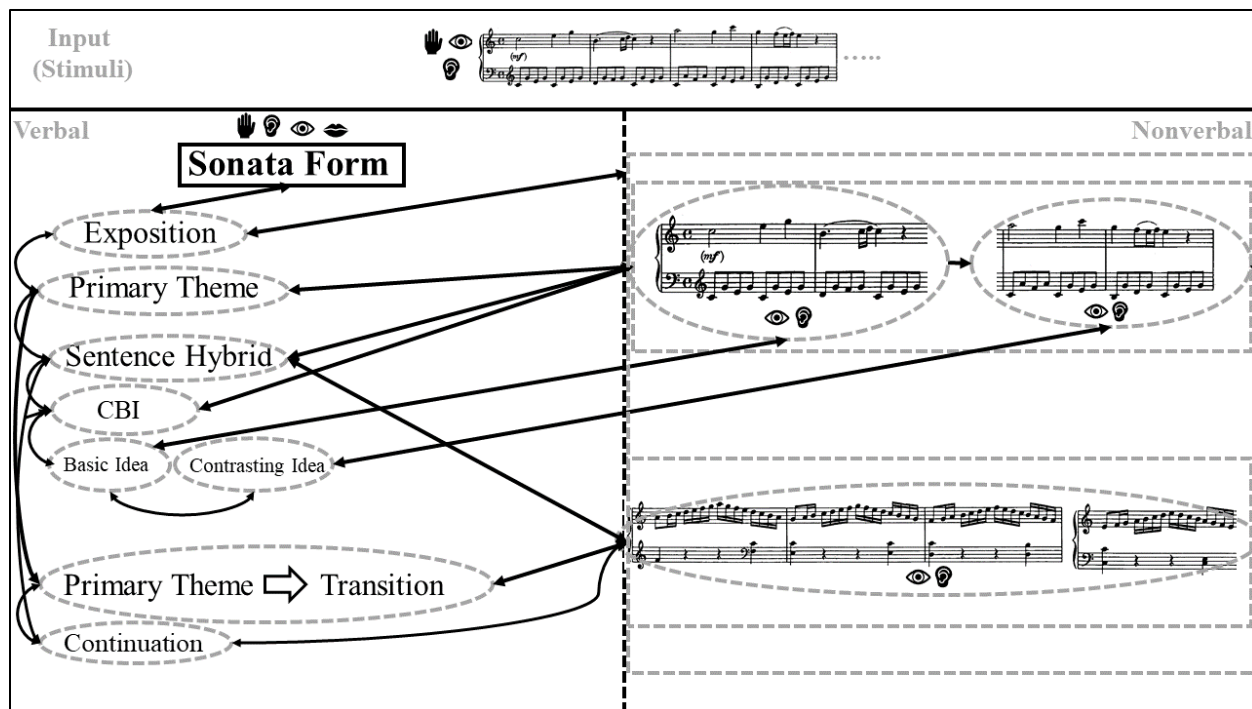


Figure 4.33. Resulting Memory Traces from K. 545 Score Analysis

In the third analysis session, the theorist attempts to amalgamate the formal and schemata analyses by actively trying to recall the schema, using the previously-used score (see Figure 4.34). By reviewing the formal analysis while actively recalling the schemata, the analyst creates an elaborated memory trace that recalls previous schemata (session 1) and formal traces (session 2) while further elaborating these with the addition of motor (score annotation), auditory (listening), and vocal traces within and across verbal and nonverbal systems (see Figure 4.35). In this figure, attentional highlighting is shown in the LTM memory trace, representing enhancement of some parts of the trace relative to others. Attention is distributed across the bass-soprano co-occurrence points in the Prinner, while for the Indugio and Ponte schemas, attention is focused on the bassline. As a result, simulators for those features stored in LTM will be activated through attention: counterpoint, soprano, and bass line for the Prinner, and bass line for the Indugio and Ponte.

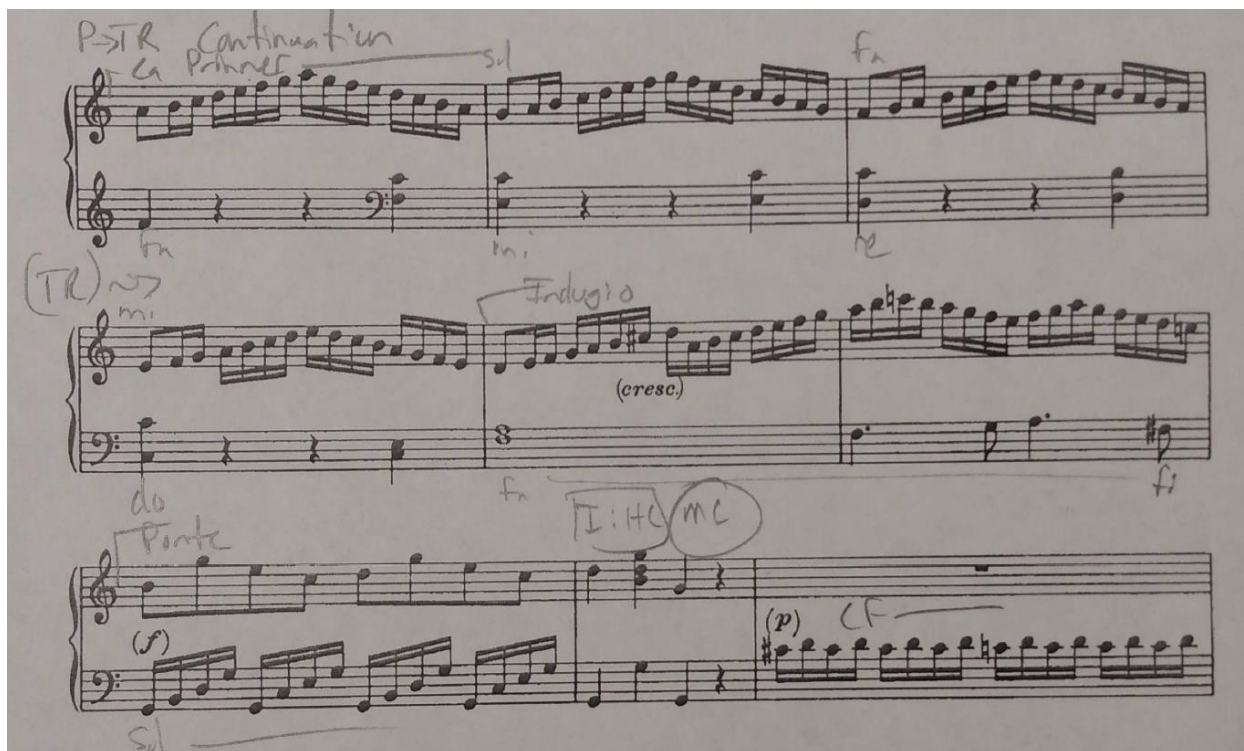


Figure 4.34. Addition of Schema Labels onto Formal Analysis of K. 545

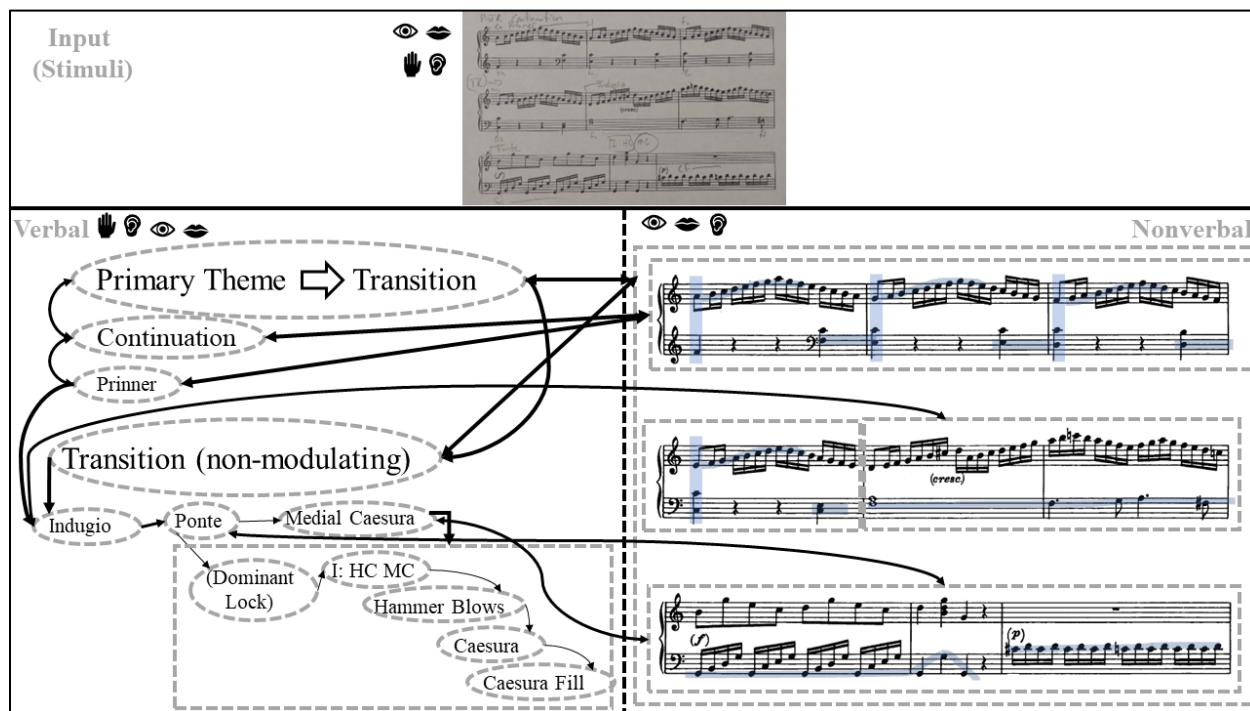
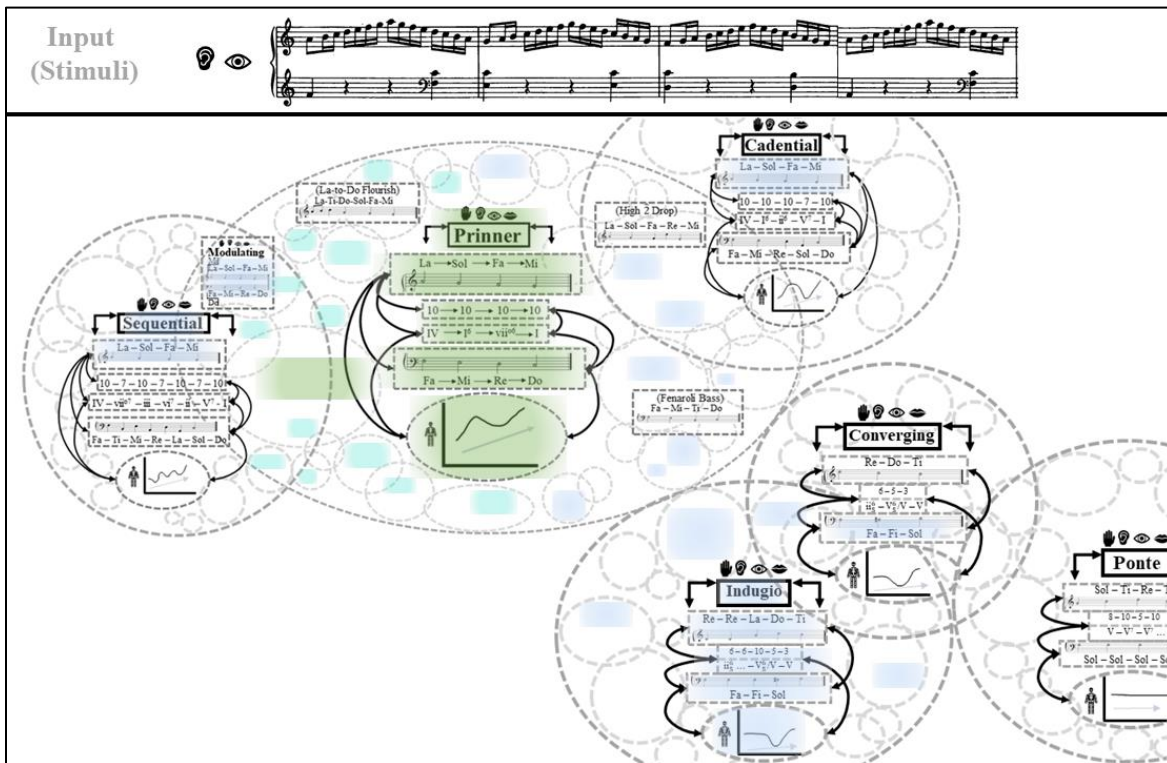


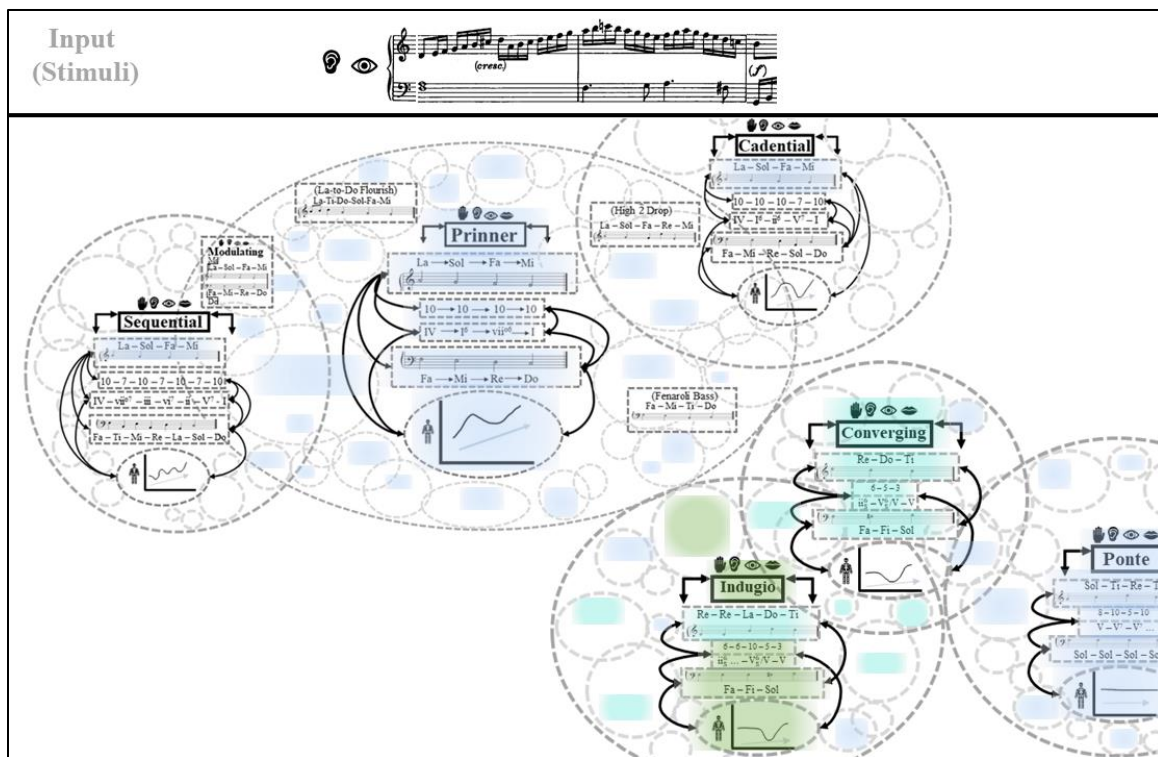
Figure 4.35. Resulting LTM Trace from Addition of Prinner Analyses in K. 545

Figure 4.36a,b,c shows the representational activation of simulator pools for the transition of K. 545 over time. The activation of multiple Prinner simulators involves representational activation, as well as associational and referential activation to traces that may be primed (in blue) but not completely active (Figure 4.36a). This may also include priming of pools for schemata which typically follow the Prinner, such as the Indugio (which will indeed come online in the next part of the transition, as shown in Figure 4.36b). This activation primes the next likely schema, the Ponte, which comes online during the medial caesura (see Figure 4.36c). This sequential activation of various simulators stored in different schema pools establishes distributional or 'top-down' probabilities between schemata. As simulator pools activate over time, connections between simulators are modified to reflect this pattern of activation. Note that there are no separate representations for patterns of schema; rather, this information is contained in the probability of sequential activation of different pools over time.

(a). Prinner Activation and Indugio/Converging Priming



(b). Indugio Activation and Ponte Priming



## (c). Ponte Activation

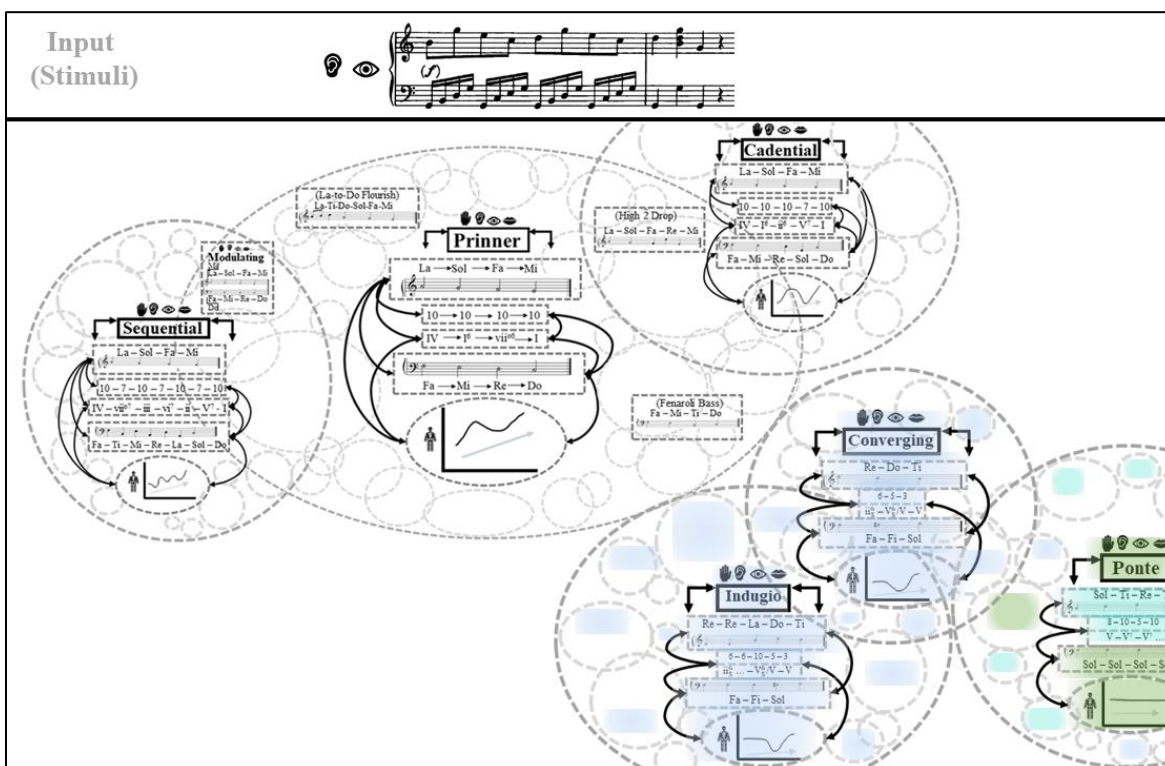


Figure 4.36. Activation of Schema Pools Over Time in the Transition of K. 545 Showing Prinner (a), Indugio (b) and Ponte(c) Activation

However, there may be instances where simulators do not come online through representational activation in listening or score viewing, and so these interactions do not facilitate a ‘pop-out’ effect for schema detection. This may be particularly true if the encountered instance is dissimilar from stored exemplars. As an example, the sequential Prinner in the S space of K. 545 is metrically offset, and the voicing of the soprano line buries the typical Prinner voicing inside the texture; therefore, attention is more drawn to the Sol-Fa-Mi-Re and Mi-Re-Do-Ti lines in the soprano and bass voices, respectively. While the sequential aspects of the Prinner may be primed through harmonizations or general bass motion, other simulators in this pool may not activate (see Figure 4.37).

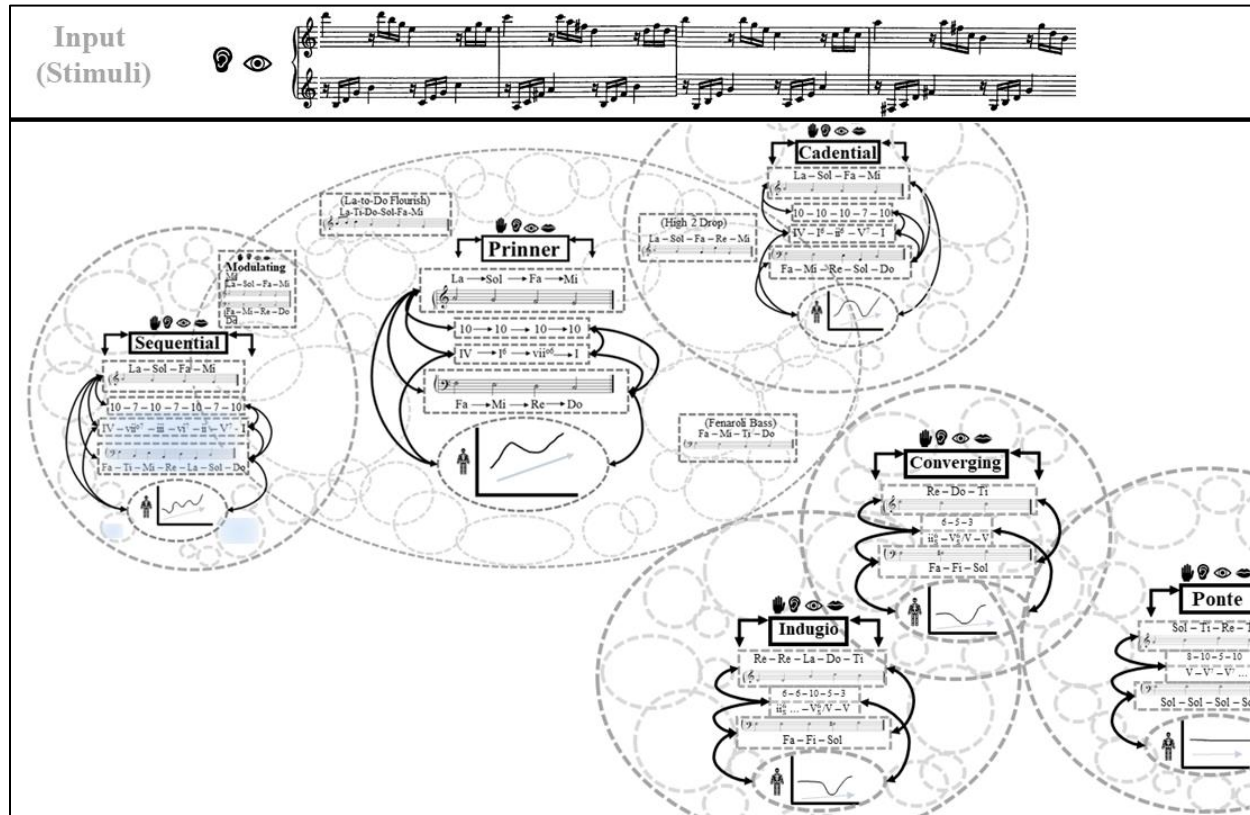


Figure 4.37. Lack of Representational Activation for Sequential Prinner in K. 545

To compensate for this, the theorist may need to deliberately bring the schema online, exerting more effort to ‘hear’ it. This can easily be accomplished by bootstrapping a categorization decision through associational and referential processing. To do this, the theorist might create an additional trace of this portion of K. 545 that contains sung solfège for the buried soprano line, paired along with the exemplar (see Figure 4.38). The addition of this trace helps to bring online previous logogen and sung traces for the soprano line and to direct visual and auditory attention to the hidden line. This in turn modifies the existing exemplar trace, enhancing regions replicated by vocalization. Singing the soprano line and viewing the hidden Prinner stages in the texture therefore allows the analyst to ‘hear’ the partly-hidden Prinner: the addition of vocalization (solfège logogens and sung imagens) produces spreading activation, activating

more simulator traces and priming others, making the *experience* of Prinner here more ‘obvious’ (see Figure 4.39).<sup>73</sup> Note that this example may still seem less ‘Prinner-esque’ compared to other exemplars that afford a pop-out effect through representational activation alone. The ‘hearing’ of Prinner in this context may seem less convincing than previously encountered Prinner in K. 545, but this exemplar is now more probabilistically bound to the existing pool of Prinner simulators, and subsequently encountered versions of this exemplar may be more likely to show the pop-out effect. The more times this process is repeated, the more convincing a hearing of Prinner hearing will be.

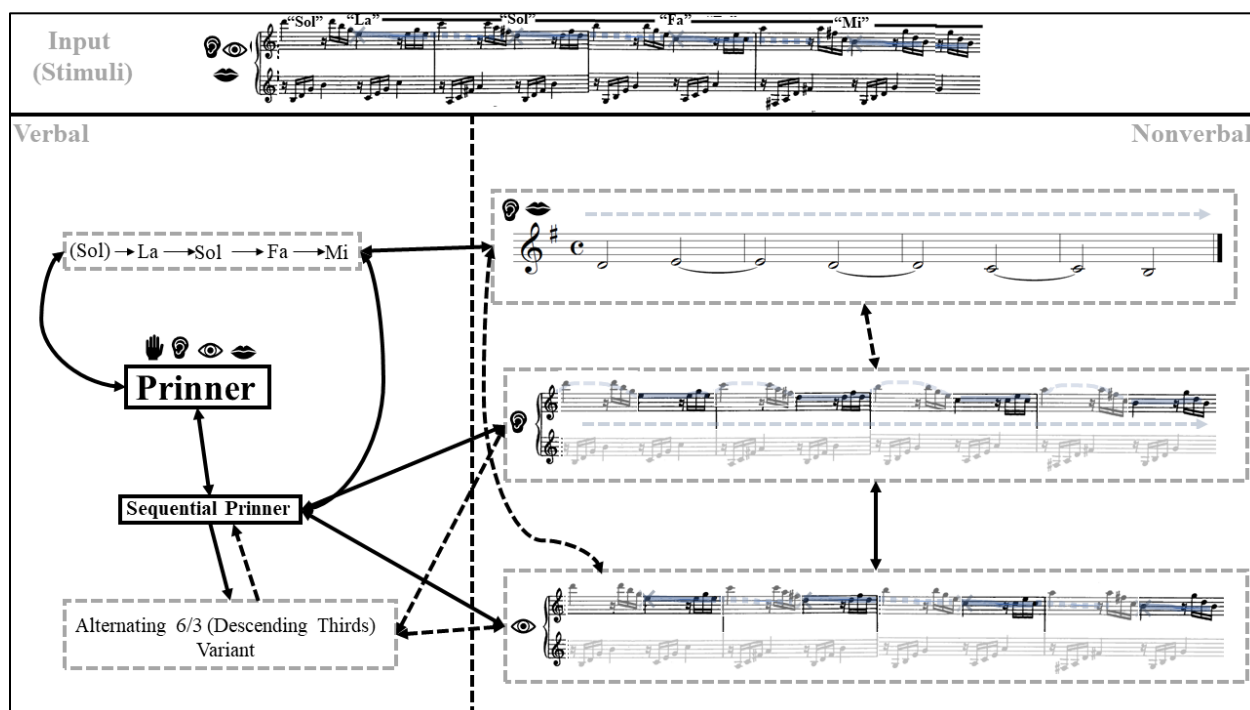


Figure 4.38. Recoding Prinner Simulator Through Addition of Soprano Voice Trace

<sup>73</sup> This prompts a kind of pop-out effect. Initially the analyst may not have explicitly recognized the Prinner, but with the addition of the vocal interaction, more Prinner simulators become active through association, allowing for the categorization decision.



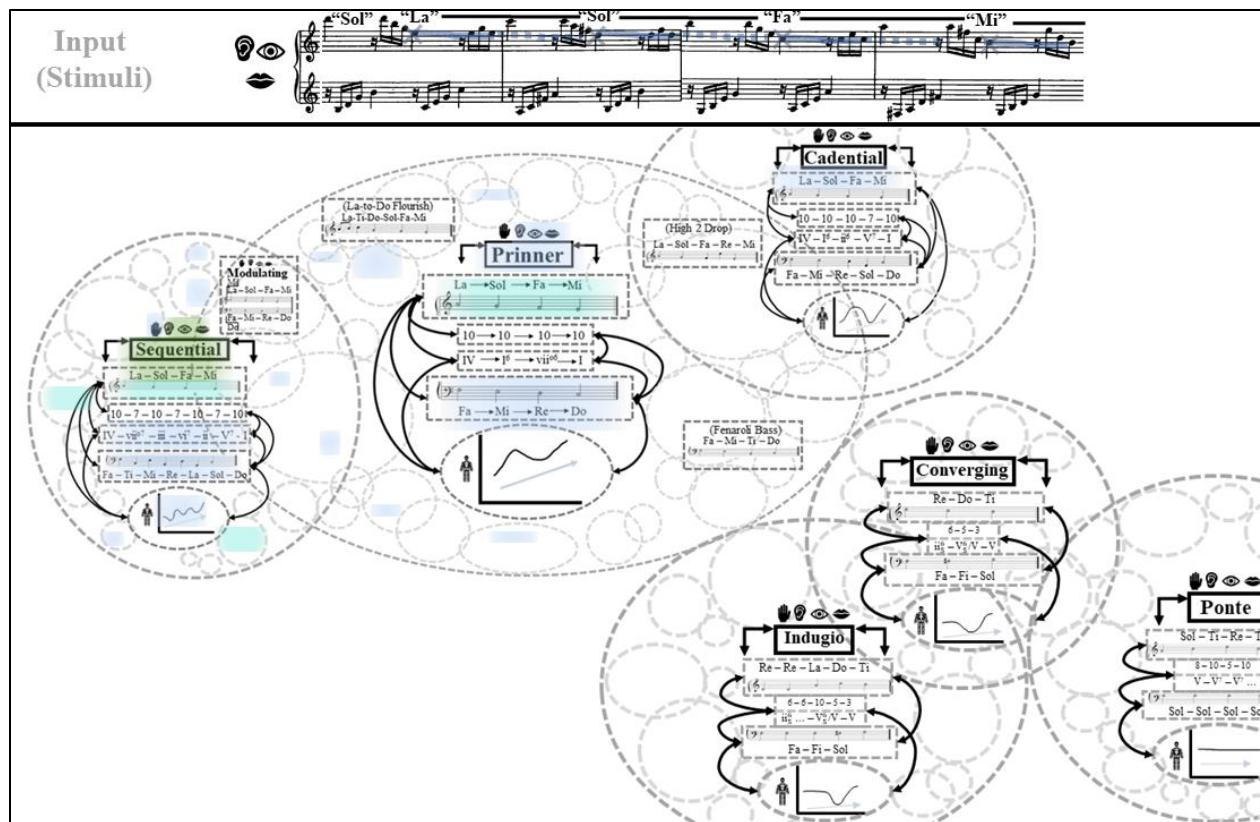


Figure 4.39. Activation of Prinner Traces Through Solmization After Recoding

In the final analytical session, the theorist listens to the piece, actively recalling the concepts encoded during previous interactions. Such actions may include verbalizations of previously encoded instances (e.g., Prinner—Indugio, “La-Sol-Fa-Mi”) and practicing directing attention to relevant features of such concepts over time in the bass, soprano, or counterpoint. Other associated interactions such as “Prinner-as-continuation” or “recapitulation in the subdominant,” may also be recalled. Over time, this analytical process helps to bind the probabilities of different simulators together. As more pieces are analyzed and committed to memory, conceptual ‘regions’ of schema subtypes may begin to take on the appearance of form-functional variants (see Figure 4.40). For example, sequential Prinner sub-type exemplar, typically four bars in length, may be associated with formally loose regions, such as transitions

and S spaces, whereas riposte-like Prinners, typically two measures in length, may become more highly associated with formally tight-knit areas, like P spaces.

To repeat: Formal knowledge in the current framework is represented in the storage of full pieces, as well as accompanying interactions, such as analytical interactions that chunk these exemplars and associate them with verbal labels for form, as well as re-coding of schemata associated in various locations. There are no separate abstract representations for form (top-down): such knowledge is captured by sequential probabilities between simulator pools. The analytical process is a means of ensuring activation of simulators over time, providing a multitude of interactions beyond those confined to the auditory domain and listening.

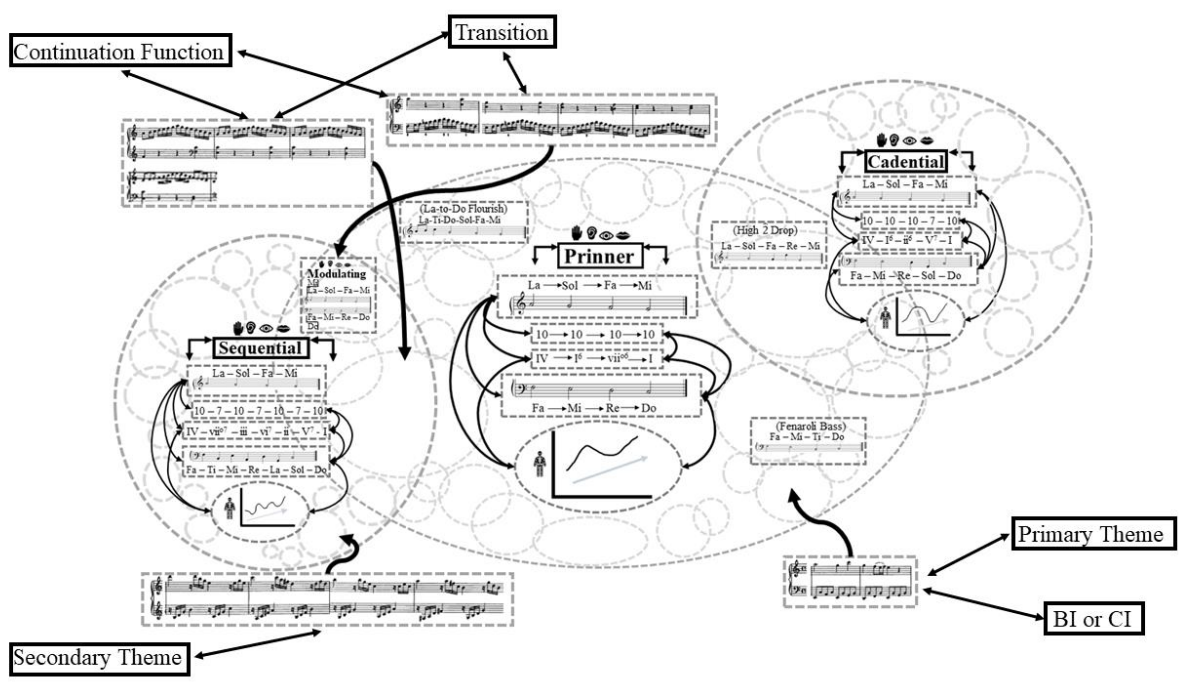


Figure 4.40. Prinner Subtype Exemplars and Associations with Formal Sections and Functions

## Summary and Conclusions

In this chapter, I have argued that Galant schema learning is a type of expert memory acquisition. Traditionally, Galant category knowledge was highly distributed across different expert domains (solfeggio, partimenti, counterpoint). Each of these domains functioned as an independent type of memory expertise, but also operated within a larger pedagogical context as necessary steps towards the expertise required for a career as a Galant composer. As students in Parisian and Neapolitan conservatories progressed in their training, each activity—solfeggio, partimenti, counterpoint—became increasingly focused on retrieval and recall rather than on encoding. Early on in training, students would learn simplified models—or prototypes—that served important roles in defining category interactions and category boundaries. They would then iteratively progress through more complex exemplars, working to greatly elaborate memory episodes stored in LTM.

Contrastingly, modern music-theoretic expertise in Galant schemata is constrained largely to skill in categorization in listening and score analysis contexts. I have suggested that the loop in music theory is an iterative practice of expertise acquisition, in which analysis is the target domain through which Galant simulators are encoded and brought online during simulation. I have demonstrated that the modern music theorist—using many of the same simulators as those trained in conservatories, such as scale degrees and counterpoint—works to realign simulators in memory to more probabilistically represent Galant schemata categories. This entails the recoding and association of scale degree lines and harmonies in memory through studying isolated exemplars, ensuring that simulators are encoded into LTM. As in traditional practices, prototypical models serve as a basis for category interaction, and help to carve out the category boundaries of a given schemata. Following this, music-theoretic expertise involves the

application of this knowledge in analysis, the process of bringing online previously learned simulators from LTM in WM in the right order and at the right time. Through the analytical process, theorists continue to encode schemata features and associations, including their relations to other concept simulators in memory, such as formal function. As expertise is acquired, the number of exemplars for schemata increases, as well as their level of elaboration in memory. As a result, the probabilities between particular exemplars—the position of the chunk relative to others, and their connection to other simulators through a continuation function—become more refined. This process provides so-called ‘top down’ sensitivities to schemata and their form-functional usages. As expertise increases, the fluency with which simulators come online during listening and analysis increases, resulting in automation of processing and categorization. In the following chapter, I detail the effects of this type of expertise in the context of how schemata are interpreted perceptually, arguing that expertise in fact leads to a rigidity of interpretation in schemata categorization due to less-probabilistic simulator pools’ decreased activations.

## Chapter 5

### Expertise in Action: LTWM as Control Over Simulation

In this chapter, I examine one aspect of music theory expertise in detail: the formation and manipulation of schema interpretation. I argue that forming and changing such an interpretation is a type of memory expertise which represents control over simulation in WM—essentially, the selection and instantiation of different property and relation simulators over time from LTM into WM. In the first part of the chapter, I discuss issues in interpretation formation, particularly the contribution of top-down and bottom-up factors to ambiguous figure perception. I then present a case study examining perceptual ambiguity of the Modulating Prinner Schema in Mozart piano sonatas. I discuss Mozart's use of the Modulating Prinner and Step-Descent Fauxbourdon Romanesca schemata within sonata form transitions, which I argue presents a case of perceptual ambiguity between these two schemata. I conclude by discussing the results from a qualitative survey on ambiguous figure perception in the transition of Mozart's Piano Sonata no. 2, iii. The results demonstrate that the transition is perceptually bistable, but that the extent to which an interpretation can be formed or changed depends on a participant's level of expertise with Galant schemata categories.

#### Issues in Top-Down Effects in Interpretation and Perception: Hazy Boundaries Between Perception and Cognition

One debated issue in cognitive sciences is *cognitive penetrability*—the idea that higher-level cognitive processes can *directly* affect lower-level perceptual processing. Such higher-level processes include beliefs, desires, emotions, intentions, and linguistic representations (Firestone

and Scholl 2016). Those in favor of the cognitive penetrability hypothesis argue that higher-level states directly affect what we *see* at the perceptual level, entailing a radical, non-modular blurring of perception and cognition. Research examining cognitive penetrability has primarily focused on the visual domain, phenomena such as ambiguous figure perception proposed as evidence for top-down effects on perception (see Long and Toppino 2004 for a review). While there seems to be a general consensus that ambiguous figure interpretation involves a mixture of top-down and bottom-up processing (e.g., Intaite *et al.* 2013; Leptourgos *et al.* 2020), recent research has strongly challenged the cognitive penetrability hypothesis, claiming that there exists little to no evidence for top-down effects on perception (see Firestone and Scholl 2014; 2015a,b,c; 2016). Instead, a counter-claim is made that effects found in previous research do not represent top-down effects on perception, but rather demonstrate confined effects of cognition, namely the formation of a judgment in higher-level processing as opposed to a direct modulation of perceptual processes (Firestone and Scholl 2016).<sup>74</sup> Effects on interpretation are driven primarily by higher-level cognitive processes involving attention and memory (*ibid.* 11, 15), which Firestone and Scholl argue are self-contained modular cognitive processes.

Regardless of one's position on cognitive penetrability, the scholarship in this area supports the notion that interpretation of a percept is heavily dependent on effects of attention and memory (see Toppino 2003; Meng and Tong 2004; Pearson and Brascamp 2008). Changing interpretations of bistable visual stimuli are mediated by shifting visual attention and fixation to different parts of the stimulus (see Taddei-Ferretti *et al.* 2008) and also affected by memory

---

<sup>74</sup> In this view, the formation of an interpretation depends primarily on higher-level cognitive processes which do not affect lower-level perceptual ones. Thus, the way in which the eye and lower-level visual system takes in and processes information is unaffected by cognition; the visual information remains the same, but what changes are higher-level cognitive processes. Conversely, other authors argue that this definition of perception is far too narrow (Cañal-Bruland, van der Kamp and Gray 2016).

priming (Hortlitz and O’Leary 1993). Imagery also plays an important role in ambiguous figure perception; participants with higher imagery ability demonstrate greater ease in generating, maintaining, and alternating interpretations (Pearson, Clifford and Tong 2008). Our primary concern here is with the effects of attention and memory in percept interpretation; I suggest that forming and changing an interpretation are inherently parts of music theory memory expertise, which is used in both imagery and perception. Forming an interpretation in this context is therefore understood to operate through top-down control and selection of representations present in simulator pools: the selective instantiation of different property and relation simulators from LTM into WM. The extent to which this process is under *active* control of the central executive or is governed by more automated ‘bottom-up’ perceptual priming is contingent on the type of activity (i.e., imagery versus perception).

### Control over Interpretation in Perception: Musical Ambiguous Figures and the Case of the Modulating Prinner

Controlling an interpretation in perception, as compared to imagery, presents a multitude of challenges. Unlike imagery, where the object for interpretation can be held and manipulated in WM, the new sensory information constant in auditory perception requires much more automation to form an interpretation, let alone to change it. Forming or modulating an interpretation during perception therefore relies much more on attention: shifting attention can be one means of selecting a different simulator base for interpretation from LTM. This does not rule out an important role for imagery during perceptual interpretation, however, as I will demonstrate later, with theorists using subvocalization (musical and/or verbal) to bring different simulator pools online. Using such techniques along with focusing attentional resources via

simulation using music theoretic concepts, theorists learn to toggle between interpretations and evaluate them during listening. This chapter now turns to investigating such phenomena, reporting the results from a qualitative survey on musical ambiguous figure perception. In the first section, I discuss the study's musical materials and contexts: modulating Primmers and Step-Descent Romanescas in sonata form transitions. I then outline the study's guiding questions, hypotheses, method, and results, followed by a discussion that contextualizes my findings within the framework developed in this dissertation.

### **Modulating Primmer and Sonata Form: An Overview**

The modulating Primmer is a subtype of the general schema, which is heard as ending in the key of the dominant and was commonly used to facilitate a modulation (Gjerdingen 2007, 52). It therefore is an ideal schema to use in formal locations where modulation is expected, such as transitional spaces of sonata forms. The sonata form transition (TR) is a space located between the primary (P) and secondary (S) theme areas in a two-part exposition. It culminates in a punctuation: a medial caesura, more often than not one that is modulatory, before the entrance of S in a new key. In major mode sonatas, the first-level default for modulatory transitions is a V: HC MC, and in minor mode sonatas it is either a III: HC or v: HC MC (Hepokoski and Darcy 2006, 25-26). Should the transition be non-modulatory, the second-level default for medial caesuras is a I: HC MC (*ibid.*, 26). The rarest third- and fourth- level defaults for medial caesuras are the V: PAC in modulatory transitions and I: PAC MC in non-modulatory transitions, respectively (*ibid.*, 27, 29).

Hepokoski and Darcy (2006) discuss the typical structure of the TR space as one that presents many analytical challenges due to the ambiguous nature of the P-TR boundaries,



particularly in the case of fused or merged transitions (Hepokoski and Darcy 2006, 95).<sup>75</sup> In general, however, transitional rhetoric is fairly distinct from the rhetoric of the P space, including normative expectations for energy gain up to the arrival of the medial caesura. Thus, while the opening of the TR space may be inherently ambiguous, transitional rhetoric is more easily identifiable as one gets closer to the medial caesura. Typical features of transitional rhetoric include *Fortspinnung*, sequential activity, *forte* dynamics, and drive toward a structural dominant. The drive toward the medial caesura has several recognized stages. The first is typically a ‘push’ toward arrival of a structural dominant, which is often prolonged through a dominant pedal (Hepokoski and Darcy 2006, 30-31). This is followed by a continuation of energy gain through increased dynamics, rhythmic activity, and other means (ibid, 33). Subsequently, the arrival of the medial caesura often occurs with several *forte* ‘hammer blows’ to mark the accumulation of energy gain toward its terminal peak (p. 34), followed by a grand pause, sometimes including *caesura* fills. This is followed by the launching of S, the secondary theme area, which is exemplified by a sudden change in texture and, often, *piano* dynamics in the second expositional key, which together confirm the prior event as the medial caesura (p. 36). The TR space in the recapitulation invites re-composition. As the S reappears in the tonic key, a modulatory transition is not required. However, composers often play with modulatory expectations, including ‘superfluous’ or ‘unnecessary’ re-composition in the P-TR spaces. Such reinterpretation may involve tonal shifts, often in the subdominant direction, before ‘correcting’ back to the home key (ibid, 236). Such procedures, both in the expositional and recapitulatory TR spaces, provide numerous interpretational challenges, including ‘perceptually ambiguous’ areas where multiple interpretational options exist, in both formal and tonal interpretation.

---

<sup>75</sup> Hepokoski and Darcy note that composers likely understood the TR space as one that was appended to the P-space as continuation modules (p. 93).

The rhetorical structure of the TR space is one that encourages particular pairings of schemata over time. In modulatory transitions, this often includes the pairing of the modulating Prinner, to initiate the tonal shift towards the dominant (in major-mode sonatas), with a converging cadence and Ponte to punctuate and extend the dominant arrival before the medial caesura. Such common pairings in the TR space are reflected in the transitional probabilities between the Prinner, modulating Prinner, and other schemata typically used in transitions as Gjerdingen (2007, Chapter 27) discusses (see Figure 5.1). The Prinner and Modulating Prinner schemata are most likely to be followed by cadences, half-cadences, Ponte, Cudworth, and converging cadence schemata, demonstrating the flexibility of Prinner schema in terms of tonal goals: it may shift toward closure either in the key of the tonic or in the dominant. Modulating, dominant trajectory Prinner pairings (e.g., Modulating Prinner to Converging) are particularly likely in transitional spaces, which has been proposed (Byros 2011) as a pairing that facilitates 18<sup>th</sup>-century interpretational responses, such as the perception of tonal modulation *prior* to receiving any new key information. In such instances, the modulating Prinner-converging cadence pair is contained within a large-scale Indugio schema, initiating perception of modulation, and shifting expectations towards a dominant arrival in the key of the dominant before the modulation is marked by a change of key (see Figure 5.2). Thus, while the introduction of key-confirming pitch information occurs in the final two stages of the modulating Prinner, the 18<sup>th</sup>-century interpretation is to perceive the change of key at the opening of the modulating Prinner, snapping the perception of tonal orientation to a new one when perception of the schema first occurs.

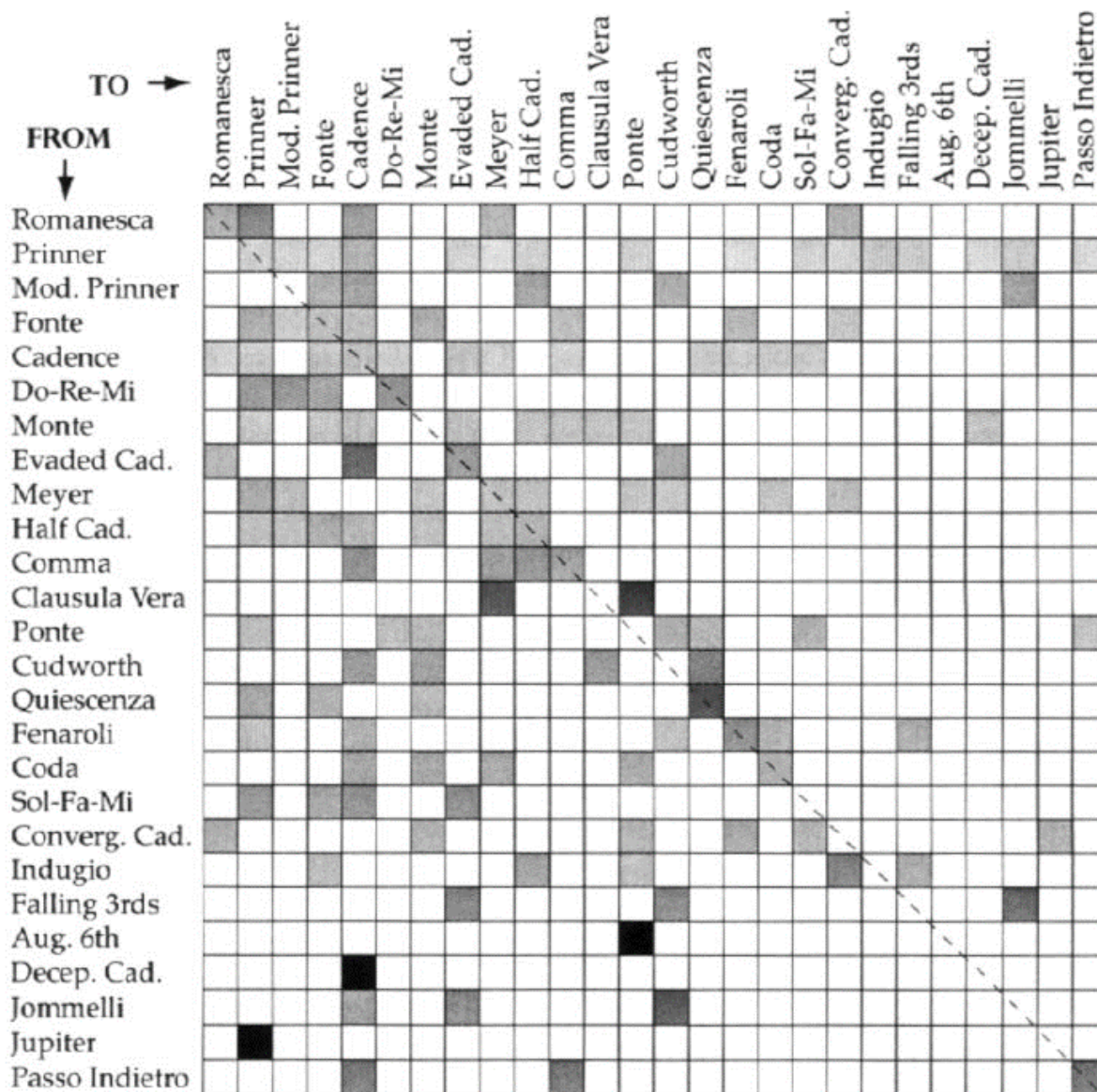


Figure 5.1. Figure 27.1 from Gjerdingen (2007, 372)

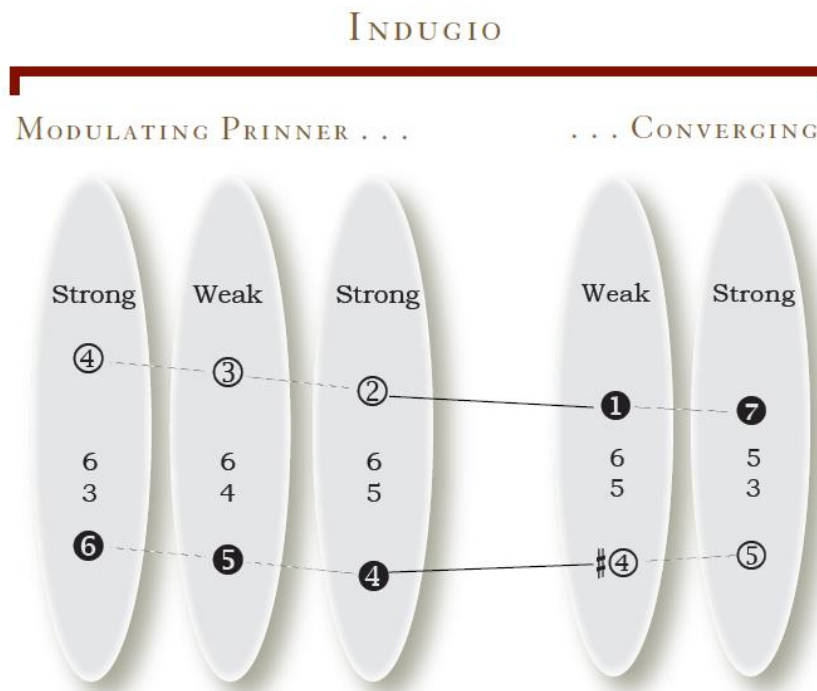


Figure 5.2. Embedded Prinner and Converging in an Indugio Schema from Byros (2011)

### A Mozart Case Study: Two Transitional Schema Options

The transitional space in a sonata form offers an abundance of interpretational affordances, particularly around the medial caesura goal. As each sonata essentially contains two versions of the transition—one in the exposition which may or may not modulate, and another, often re-composed, version in the recapitulation that usually does not modulate—different mental ‘templates’ exist for transitions in these contexts, both as a generalized ‘transition-concept’ (i.e., particular simulators activating over time), and on a piece-by-piece basis. Just as there are multiple options for the tonal goal of the transition, there are also, in the exposition, multiple potential ‘pathways’ through tonal space which drive toward the medial caesura. As composers often play with these expectations, the sonata form transition is one in which there are multiple ‘schema-templates’ available to the analyst, affording some flexibility in interpretation

in these spaces. In this section, I will discuss two different schema options for modulating and nonmodulating transitions—the Modulating Prinner and the Step-Descent Fauxbourdon Romanesca (hereafter Step-Descent Romanesca). I will begin by discussing the role of the Modulating Prinner in modulating transitions, followed by the Step-Descent Romanesca in tonic-confirming (nonmodulating) contexts.

### Modulating Prinner

The modulating Prinner, as previously discussed, is one of the primary schemata seen when modulating to the dominant, and therefore often occurs in modulating transitions whose goal is the key of the dominant (most typically V: HC MC). The modulating Prinner, in addition to its tonal pliability with regard to scale degree identity, has several other frequently-occurring features as discussed by Gjerdingen (2007); among these are sequential or *Fortspinnung*-like construction and an octave leap in the upper voice, which functions as a signal to change scale degree identity (where  $\hat{3}$  becomes  $\hat{6}$  in the new key, Gjerdingen 2007, 355). Here I will discuss two versions of such a modulating Prinner in Mozart piano sonatas K. 280, iii and K. 309, i.

*Piano Sonata no. 2, K. 280, iii.* The third movement of K. 280 is a *Presto*, Type 3 sonata, featuring an independent, modulating transition ending with a V: HC medial caesura. The boundary between the P and TR space is slightly obscured: the P space ends with a I: PAC in measure 16, and the TR begins in the following measure, taking the form of a dissolving P-codetta type over a tonic pedal point (see Figure 5.3). Rhetorically, the TR space is clearly initiated by measure 25, which aligns with the presence of the modulating Prinner, complete with *forte* dynamics, octave-leaps in the upper voice, and sequential, fragmentary construction typified by a continuation phrase. It is at measure 25 that the modulating Prinner marks the availability of the new key of the dominant (C major), which is reinforced in measure 27 with the occurrence of the new leading tone, and fully confirmed with the presence of the dominant arrival in C major in measure 32. As discussed in Byros (2011), the ordering of Galant schemata seen here is typical of Mozart: a modulating Prinner (measures 25-28) leads to a converging cadence (measures 31-32), which is prolonged until the medial caesura proper (measure 37). Together, both schemata form a larger-scale Indugio, which prolongs  $\hat{4}$  (measures 25-29), passing through the converging cadence to reach the dominant of the key. In this movement, the medial caesura is marked by a grand pause, which then continues to the S-theme in the key of the dominant (C major), starting as is typical on *piano* dynamic. The recapitulatory transition (measures 132-148) receives only a small adjustment to re-orient the listener and accommodate the lack of modulation in preparation for the S-theme. Here, the original modulating Prinner occurs in measure 132 (C major), which is then followed by a secondary modulating Prinner in the home key, which re-orient the listener back to F major (see Figure 5.3). The crux is at measure 140, which continues as the expositional TR: a large scale Indugio contains the second modulating Prinner (measures 136-139) and converging cadence (measures 142-148).

## Exposition

## TR (Dissolving, Post Cadential)

13 P  
 Cadence I: PAC Presentation BI  
 Tonic Pedal  
 F:

19 BI Continuation Indugio  
 Tonic Pedal Modulating Prinner  
 C:

29 Converging V: HC MC  
 Dominant Arrival  
 p f p f p f

Figure 5.3. Expository Transition of K. 280, ii, 13-37

**Recap**

P Cadence I: PAC TR BI

120

F: Modulating Prinner

126 Tonic Pedal Indugio C: Modulating Prinner Converging

136 F:

143 I: HC MC. Dominant Arrival

The musical score consists of four systems of piano music. The first system (measures 120-125) is marked 'P' and includes a 'Cadence' and 'I: PAC TR BI'. The second system (measures 126-135) features a 'Tonic Pedal' in the bass line and is annotated with 'Modulating Prinner' and 'Indugio'. The third system (measures 136-142) continues the 'Modulating Prinner' and ends with a 'Converging' section. The fourth system (measures 143-148) is marked 'I: HC MC.' and includes a 'Dominant Arrival'.

Figure 5.4. Recapitulatory Transition of K. 280, iii, 120-148



*Piano Sonata no. 10, K. 309, i.* Like K. 280, the first movement of K. 309 features an independent, modulating transition, facilitated by a modulating Prinner which leads to a V: HC MC (see Figure 5.5). In this movement, however, the boundary between the P and TR spaces is more clearly articulated, featuring an elided I: PAC in measure 21 marking the end of the P and beginning of TR. This boundary is also marked by typical TR rhetoric, including *forte* dynamics and sequential construction. And much as was seen in K. 280, this marks the beginning of the modulating Prinner which facilitates the modulation to the key of the dominant. This modulating Prinner also contains a similar octave leap construction, with a two-bar sequential stage presentation. The parallel tenth voice structure is emphasized at the beginning of each stage; however, the bassline is modified to reflect the mixing of the Prinner and Fenaroli schemata, which further emphasizes the tonic arrival in the new key in measure 27. This arrival is slightly undermined by the repeated iteration of  $\hat{4}$  (from the previous stage) above the final stage, only reaching resolution in measure 29. This is followed by a two-bar modified Pulcinella schema (measures 29-30), and the V: HC MC in measure 32. Mozart includes *caesura* fill in measures 33-34, at the beginning of the S in measure 35 in the key of the dominant, marked with *piano* dynamics.

**Exposition**

P ----- Continuation → Cadential -----

18

C:

I: PAC TR ----- Modulating Prinner -----

(elided)

21

G:

Pulcinella

V: HC MC

27

Figure 5.5. Expositional Transition from K. 309, i, 18-32

The recapitulatory TR is quite similar to that found in the exposition, with a slight modification to adjust for the lack of modulation (see Figure 5.6). The schematic layout is like the exposition, with the same Modulating Prinner and Fenaroli bassline in measures 116-124. The final stage of the Prinner includes a pedal point rather than tonic-dominant alternations and includes the melodic portion of the Pulcinella schema from the exposition, essentially merging these two areas into one. Within the context of the modulating Prinner version heard in the exposition and given the continued presence of F-sharp leading tones, the pedal is not yet reinterpreted as a dominant arrival, even though rhetorically the presence of a pedal point in this section of the TR space is highly indicative of dominant arrival. Instead, the crux occurs at measure 125, which acts as a perceptual snap back to a dominant interpretation to align with the

structure of the expositional TR. The I: HC MC is maintained here in measures 125-126, followed by caesura fill and the tonic key entrance of S in measure 129.

**Recap** **TR**

**P** Continuation → Cadential I: PAC (elided) Modulating Prinner

113 *f p f p* *fp cresc.*

C: G: Fenaroli (Bass)

Modulating Prinner (cont'd)

117 *fp cresc. fp cresc. f*

123 Pedal Point I: HC MC

C:

Figure 5.6. Recapitulatory Transition of K. 309, i, 113-126

### Step-Descent Fauxbourdon Romanesca

The Romanesca schema is typically used as an opening gambit and often occurs as a basic idea in thematic material. There are several variants of the Romanesca. The earliest form is the leaping, descending thirds sequential variant (Gjerdingen 2007, 29), and the second a scalar step-descent variant which descends through the entire octave from tonic to tonic, alternating 5/3 and 6/3 harmonies (ibid., 32). Both of these variants' sequential structures afford functional

usage in continuation phrases. The last variant of the Romanesca, a Galant hybrid, combines the leaping and stepwise variants, and forms the version commonly used as an opening gambit (ibid., 33). There is some structural similarity between the step-descent Romanesca and the Prinner, including the stepwise bassline and alternating 5/3 and 6/3 harmonies. The Prinner's distinguishing feature, however, is the parallel-tenth counterpoint in the outer voices; the Romanesca's upper voice motion, by way of contrast, typically alternates between tonic and dominant scale degrees. This can be more closely approximated in the step-descent Romanesca by the addition of fauxbourdon voice leading and harmonic inversion, similar to voicings used in the Rule of the Octave (Gjerdingen 2007, 468). This structural similarity is evident in the use of the Step-Descent Fauxbourdon Romanesca as a key-confirming schema in both non-modulating and modulating transitions. Compared to the modulating Prinner, the Step-Descent Romanesca is used to reinforce an ongoing key, rather than initiate a modulation. Below I present two versions of such a schema in the TR space, to demonstrate its typical usage in such spaces.

*Piano Sonata no. 15, K. 576, i.* The first movement of Mozart's Piano Sonata K. 576 features a non-modulating, independently thematized transition. Here, a Step-Descent Romanesca is used in the transition in a similar manner to the modulating Prinner, both in terms of ordering (Romanesca-Converging-Ponte) and realization, on the way toward the medial caesura (Figure 5.7). The transition begins in measure 16, and much like the first movement of K. 309, features an elided PAC following the end of the P. The first eight measures of the transition firmly iterate the home key of D major with two repetitions of the Aprile (measures 16-19) and Fenaroli (measures 20-23) schemata. This is followed by the Step-Descent Romanesca (measures 24-25), descending through nearly the entire octave in the bassline (from  $\hat{3}$  to  $\hat{4}$ ), and completely through the octave in the soprano ( $\hat{1}$  to  $\hat{1}$ ). Contextually, this Romanesca has many similarities to the Modulating Prinner, including the placement and progression of the schema to a converging cadence and Ponte, driving toward the I: HC MC. However, it also has several rhythmic and motivic similarities to the Modulating Prinner previously discussed, such as the octave displacement in the soprano voice, and rhythmic figuration similar to that used in K. 280.

## Exposition

TR I: PAC (elided) Aprile

16

20

24

D:

Step-Descent Romanesca

Converging

I: HC MC

Ponte


Dominant Arrival

Figure 5.7. Expositional Transition of K. 576, i, 16-27


The recapitulatory TR is recomposed to resemble a dissolving P-type, which articulates several key areas, including the subdominant, before returning to the home key for a I: HC MC (see Figure 5.8). The TR begins in measure 106, at the point in the exposition where a restatement of P was provided (measure 9). Here, the P-material shifts to a subdominant inflection for the contrasting idea in measures 110-112, which is followed by fragmentation and imitative material in the keys of the minor dominant (a) and submediant (b). The crux occurs at measure 118 where the original Step-Descent Romanesca is provided, which functions to suddenly snap back to the home key of D major and is followed by a Converging Cadence to Ponte I: HC MC in measures 119-121.

**Recap** **TR (Dissolving)**

CI I: HC BI CI I: PAC


101 **P** 

D: "BI" "CI"

108 

Subdominant Inflection  
G:


Fragmentation and Imitation

113 

a: b:

Romanesca Converging I: HC MC

Fauxbourdon Variant Ponte

118 

D: Dominant Arrival

Figure 5.8. Recapitulatory Transition of K. 576, i, 101-121

*Piano Sonata no. 14, K. 310, iii.* The third movement of K. 310 presents an interesting case. The final movement is a Type 4 sonata, or sonata rondo, which uses a trimodular block (TMB) strategy in the exposition, and a medial caesura declined in the recapitulation. Mozart uses a Step-Descent Romanesca as tonic confirmation within the TMB, as well as a hybrid modulating Prinner at the second medial caesura in the exposition (see Figure 5.9).

**Exposition** **TR (b)**

**P (a)** <sup>19</sup> i: PAC

a: <sup>29</sup> III: HC MC **S (a): TMB** BI C: BI

Continuation <sup>37</sup> Evaded PAC

Step-Descent Romanesca Cadential Evaded

<sup>45</sup> Step-Descent Romanesca Cadential PAC

<sup>55</sup> Evaded (Modulatory) Prinner? v: HC MC

e: Indugio

Figure 5.9. Expositional Trimodular Block (TMB) in K. 310, iii, 19-63



Here, the modulating transition (or b rotation) begins after a i:PAC, which quickly modulates to the major mode mediant, and weakly articulates a III: HC MC in measure 28. The S zone, or third rotation (a), begins directly afterwards in the minor mediant, and is structured as a sentence, in which the Step-Descent Romanesca functions as the continuation phrase, in the process confirming the key of C major. This schema is repeated twice, each time leading to an evaded cadence, which is followed by modulatory material leading to a second medial caesura in the key of the minor dominant. Interestingly, this second articulation of the MC is structured as an Indugio in which the upper voice (which had previously been the descending line, starting on tonic, in the Step-Descent Romanesca) is reinterpreted in the new key as a minor mode Prinner soprano line (le-sol-fa-me) in measures 60-61.

This procedure is like other usages by Mozart, in which the modulating Prinner and Converging Cadences are embedded into an Indugio (see K. 280 above). This compressed, fused hybrid is also typical of Mozart, where the Converging Cadence (bass line) is combined with the melody of the modulating Prinner (Gjerdingen 2007, 353). This recasting of the primary voice from the non-modulating Step-Descent Romanesca to a modulating Prinner variant highlights the plasticity and combinatorial potential of the different parts of these schemata. The recapitulatory version of the transition includes a medial caesura declined, without the expositional TMB (see Figure 5.10). Here, the modulating hybrid Indugio (Modulating Prinner + Converging) is excluded, but the tonic confirming Step-Descent Romanesca is maintained.

**Recapitulation**

**P (a)**<sub>190</sub> **TR (b)**  
*i: PAC*  
 a: *f* *p*

**MC declined**  
<sub>198</sub> **S (a): TR → FS BI**  
*f* *p*

**BI** **Step-Descent Romanesca**  
<sub>206</sub> *f*

**Cadential** **i: PAC ESC** **C** **Step-Descent Romanesca**  
<sub>214</sub>

Figure 5.10. Recapitulatory Transition of K. 310, iii, 190-221

### An Experimental Verification: Effects of Attention, Memory, and Expertise on Bistable Perception in Mozart's Piano Sonata no. 2, K. 280, iii

I have demonstrated that both the modulating Prinner and Step-Descent Romanesca occur within TR spaces in several of Mozart's piano sonatas. These schemata have several features in common, which suggests potential for perceptual ambiguity:

- Sequential organization and presentation of schemata stages, which features fragmentary-like construction typical of medial (continuation) function.

- Melodic and rhythmic similarities, viz. a rhythmically offset melodic voice that arpeggiates against a bass that is metrically stable.
- Parallel tenths between two of the three or four voices present.
- Similar spatial-temporal locations (i.e., in the TR space leading up to the medial caesura), and temporal orderings with other schemata (e.g., \_\_\_\_\_—Converging—Ponte).

However, important differences help distinguish one schema from the other. For example, the distinguishing Modulating Prinner features are:

- Equal-length schema stages that range from one to two measures long. This type of construction is ideal for providing the listener with time to perform a tonal reorientation as each stage is presented for enough time such that scale degrees and harmonies can be reinterpreted in the new key.
- Stages that are harmonized with a mixture of 5/3 and 6/3 sonorities which alternate.
- Temporal-spatial location near the beginning of the transition as it often initiates transitional rhetoric.

Contrastingly, the distinguishing features of the Step-Descent Romanesca are:

- Twice the number of schema stages as the Prinner (eight in total), descending the entire scale from tonic to tonic.
- Shorter and rhythmically unequal stages (e.g., long-short-long-short). This feature aligns with the tonic-confirming function of the Step-Descent Fauxbourdon Romanesca, with which there is no need for the composer to offer time to the listener to tonally reorient.
- Only 6/3 harmonies, not alternating with 5/3 harmonies.
- Occurs near the middle or end of the transition, and therefore does not initiate transitional rhetoric.

Given the overlapping features of these two schemata, there is a case to be made for perceptual ambiguity on the basis of ‘bottom-up’ features alone. However, an interpretation may also vary with expertise. Byros (2009a,b; 2012a,b) has argued that expertise with Galant schemata categories provides a means to access eighteenth-century modes of hearing—particularly as it relates to tonality perception. Byros’ examination of reception history of Beethoven’s *Eroica* reveals three primary strains of tonal interpretation: one which perceives a modulation to G-minor, a “Cloud” strain which proposes a form of tonal ambiguity, and lastly one which reads the whole passage within the key of E-flat major. Byros demonstrates that the tendency to perceive key change in these opening bars operates on a historical axis, indicating that the historical situation or context is a contributing factor in the variation of key perception. The G-minor hearing therefore appears to be historically and stylistically appropriate, an interpretation which Byros proposes arises from the perception of the *le-sol-fi-sol* schema.

Expertise with Galant schemata therefore appears to provide listeners with the ability to rapidly reorient tonally and perceive ‘unconfirmed’ modulations, which from a modern and more monotonal perspective are likely perceived as temporary tonicizations that merely function to prolong an ongoing key, rather than alter it. This suggests that modern listeners—lacking in familiarity with Galant schemata—are more likely to perceive tonal reorientation in modulatory passages as occurring when new key information is introduced (e.g., accidentals) in the signal (‘bottom-up’). Contrastingly, those with schemata expertise are more likely to associate particular schemata—understood as co-occurrences of scale degrees and specific harmonies—as belonging to a particular key, even when there have yet to be any changes in accidentals. Therefore, there are two possible interpretations of the transition from the Presto of K. 280: an historically informed one, where an early modulation is perceived as occurring with the onset of

a modulating Prinner (see Figure 5.11a,b), and a more modern one where the modulation is not perceived until the introduction of new key information, specifically,  $\sharp\hat{4}$  in the new key (see Figure 5.12a,b). However, there exists a grey area in this second interpretation in light of the overlapping features between the Prinner and Step-Descent Romanesca schemata discussed above. The question, therefore, is: is it possible, with at least some expertise in Galant schemata, to hear both Modulating Prinner and Step-Descent Romanesca interpretations in this transition?

(a). Expositional Modulating Prinner

**Exposition** **TR (Dissolving, Post Cadential)**

The musical score is divided into three systems, each with specific annotations:

- System 1 (Measures 13-18):**
  - Measures 13-15: **Cadence**
  - Measure 16: **I: PAC** (Perfect Authentic Cadence)
  - Measures 17-18: **Presentation** and **BI** (Basic Interval)
  - Measures 17-18: **Tonic Pedal**
- System 2 (Measures 19-28):**
  - Measures 19-21: **BI**
  - Measures 19-21: **Tonic Pedal**
  - Measures 22-28: **Modulating Prinner**
  - Measures 22-28: **Continuation**
  - Measures 22-28: **Indugio**
  - Measure 22: **C: ii<sup>6</sup>**
  - Measure 23: **i<sup>6</sup>**
  - Measure 24: **vii<sup>6</sup>**
  - Measure 25: **I**
- System 3 (Measures 29-34):**
  - Measures 29-31: **Converging**
  - Measures 29-31: **IV(?)**
  - Measures 32-34: **Dominant Arrival**
  - Measures 32-34: **V: HC MC** (Half-Cadence / Mediant Cadence)

## (b). Recapitulatory Modulating Prinner

Recap

P Cadence I: PAC TR BI

120

F: Modulating Prinner

126 Tonic Pedal Indugio C: ii<sup>6</sup> I<sup>6</sup> vii<sup>6</sup> I

Modulating Prinner Converging

136 F: ii<sup>6</sup> I<sup>6</sup> vii<sup>6</sup> I

143 I: HC MC. Dominant Arrival

Figure 5.11. Modulating Prinner Interpretations in the Exposition (a) and Recapitulation (b) of K. 280, iii

## (a). Expositional Step-Descent Romanesca

**Exposition** **TR (Dissolving, Post Cadential)**

Cadence I: PAC Presentation BI

P Tonic Pedal

F: Continuation

BI Step-Descent Romanesca (half-length)

Tonic Pedal  $vi^6$   $V^6$   $vii^6 / V$   $V$

Converging V: HC MC

I(?) C: Dominant Arrival

## (b). Recapitulatory Step-Descent Romanesca

**Recap**

P Cadence I: PAC TR BI

F: Step-Descent Romanesca

Tonic Pedal  $vi^6$   $V^6$   $vii^6 / V$   $V$

Converging

$ii^6$   $I^6$   $vii^6$   $I$

I: HC MC

Dominant Arrival

Figure 5.12. Step-Descent Romanesca Interpretations for the Exposition (a) and Recapitulation (b) of K. 280, iii, 120-148

Using qualitative methods, I will demonstrate that perceptual ambiguity or bistability in K. 280 is plausible due to 1) overlapping features between Prinner and Step-Descent Romanesca schemata within transitional (TR) spaces, and 2) memory organization and access (i.e., LTWM expertise). I argue that the degree to which these interpretations may be *equally* available will vary with memory expertise. Similarly, the degree to which interpretations appear plausible will also vary with attention: the theoretical concept used to form the assessment or interpretation will selectively alter which memory representations are accessed. This allows for flexibility in recognition and categorization. In this way, it may be plausible to form different scale degree interpretations of a given line through focused attention. Contrastingly, it may be more challenging to form and alternate between different schema interpretations, which requires diverted attention across multiple features simultaneously. The extent to which different interpretations may be available in either scale degrees or schemata may be unequal however, and, I argue, will likely depend on the level of expertise with the categories at hand. Therefore, for an expert, hearing a Step-Descent Romanesca in the *Presto* of K. 280 may seem entirely impossible. Such experts may, however, be able to access property simulators which are shared across schema categories, and so be able to perceive a later modulation—but only when attending to a single line (see Figure 5.13). In this way, for these experts the overall category identity may remain a Prinner, but they may be able to bring online a different set of bass line simulators to remain in the previous key (i.e., perceive a Dominant Prinner).



The figure shows a musical score for the Recapitulatory Transition of K. 280, iii, 120-148. The score is divided into four systems, each with specific annotations and brackets indicating musical structures and schemata interpretations.

- System 1 (Measures 120-125):** Labeled "Recap" and "P". A bracket labeled "Cadence" spans measures 120-125. Above the staff, "I: PAC TR BI" is written.
- System 2 (Measures 126-135):** Labeled "F:" and "Dominant Prinner". A bracket labeled "Tonic Pedal" spans measures 126-135. Below the staff, "Indugio" is written. Chord symbols  $vi^6$ ,  $V^6$ ,  $vii^6 / V$ , and  $V$  are present.
- System 3 (Measures 136-142):** Labeled "Prinner" and "Converging". Chord symbols  $ii^6$ ,  $I^6$ ,  $vii^6$ , and  $I$  are present.
- System 4 (Measures 143-148):** Labeled "I: HC MC" and "Dominant Arrival".

Figure 5.13. Monotonal Prinner Schemata Interpretation for the Recapitulatory Transition of K. 280, iii, 120-148

The phenomenon of interpretational flexibility—forming and assessing multiple interpretations of a musical excerpt or entire piece—is fundamental to the discipline of music theory (e.g., Cone 1977; Agawu 1994; Guck 2006). While perceptual bistability has been widely studied in the modality of vision (e.g., Brugger 1999), the investigation of the same phenomenon in the auditory modality is quite limited. Previous work has demonstrated that the effect does exist within the auditory modality, but these effects have been limited to those using artificially constructed stimuli to examine perceptual ambiguity in auditory stream segregation (so-called ‘bottom-up’ processes, e.g., Turgeon & Bregman 2001). Research has yet to be conducted either on bistable perception using ecologically valid stimuli, or on the ability to engage categorical knowledge in the same way that vision does (using more ‘top-down’ processes, e.g., to invoke

duck vs. rabbit). Previous research suggests that this type of bistability is plausible in the musical realm; interpretational flexibility has been demonstrated for tonal scale degrees in imagery for musically trained populations (Vuvan & Schmuckler 2011). The results from this study will therefore be the first to investigate perceptual bistability that engages conceptual knowledge in ecologically viable musical stimuli. In line with previous research (e.g., Hortlitz & O’Leary 1993; Firestone & Scholl 2016), the results will also seek to contribute to the understanding of bistability as arising from effects of attention (focused vs. diffuse) and memory (experience with music theory concepts).

### Guiding Questions and Hypotheses

There were three primary questions that guided the development of the survey. They were:

- 1) Are the excerpts from K. 280, in fact, perceptually bistable?
- 2) If the excerpts are found to be bistable, is the formation or flexibility of interpretation dependent on which conceptual knowledge is being used to form the interpretation? In other words, do differences in auditory attending strategies and memory change the availability of interpretations or the ease of changing between two interpretations?
- 3) Is either forming or changing an interpretation related to the amount of expertise with relevant music theory concepts (e.g., Galant schemata)?

The study thus has one null and three primary hypotheses for each excerpt:

**H<sub>0</sub>:** This excerpt is not amenable to multiple interpretations, using either scale degrees or Galant schemata.

**H1a:** This excerpt is amenable to multiple interpretations in terms of schemata; however, one interpretation (Prinner) may be easier to hear than the other (Romanesca).

**H1b:** Furthermore, the availability of and ease of change between interpretations may differ between the expositional and recapitulatory versions of the excerpt (with the Romanesca more available in the recapitulation).

**H2:** Participants should be able to more easily form and alternate between scale degree interpretations for a single voice (soprano, bass) than for Galant schemata (which are dependent on the presence of multiple, co-occurring features).

**H3:** The ability to form and alternate Galant schemata interpretations may be dependent on moderate familiarity, but not a high-level expertise with Galant schemata (Expertise categories = Novice, Intermediate, Expert). Expertise should therefore be related to increased rigidity of interpretation (Ease of Change on a scale from 1 to 7, low to high), particularly for Schema interpretations, as such categories are overlearned and more likely to be automatically active during listening.

The null hypothesis is that the excerpt is not bistable, meaning that only one interpretation may be available, or that neither interpretation may be available. Hypothesis 1(a) posits that, while the excerpts may be bistable, an early modulation or Prinner interpretation may be more easily available than a later modulation or Romanesca interpretation. Hypothesis 1(b) suggests however that hypothesis 1(a) may vary based on the formal location (exposition or recapitulation). I argue that a Romanesca interpretation may be more available in the recapitulation (as it is nonmodulatory and contains a full scale from *do* to *do*) than in the exposition. This can be attributed to ‘bottom-up’ factors of features in common across

categories—an effect of representational activation of auditory imagens in the nonverbal system. As the exposition has more features in common with existing traces for modulating Prinner schemata, more of these traces are likely to be primed and active during listening. However, because this trace has some similarity to the Step-Descent Romanesca previously encountered, the traces containing these interactions may also be primed, even though there may be too few of them to give rise to activation of that schema as a whole (see Figure 5.14).<sup>76</sup> Therefore, one would easily be able to form the interpretation “Prinner” in the exposition, including accessing any of the relevant property or relation simulators for this category (e.g., bass line scale degrees). One might also be able to hear some similarities, albeit minimal, with an interpretation of “Step-Descent Romanesca,” making the perception of Romanesca weak or unlikely, but not impossible. The availability of a Romanesca interpretation may shift for the version of the transition found in recapitulation because there are *slightly* more similarities to a Romanesca interpretation compared to the exposition. For example, the full descending major scale in the bass may help to activate more bassline traces associated with Romanescas than the version in the exposition (see Figure 5.15).<sup>77</sup>

---

<sup>76</sup> For example, previously encountered exemplar traces, like those pertaining to the transition of K. 576 may be primed but not fully activated during listening. In this case, not enough simulators for the Romanesca category (bass, soprano, harmony) are active for the category to be clearly, fully activated. Partial activation may still occur, reflecting the similarity between this instance and previously experienced instances.

<sup>77</sup> The full-scale interpretation is weakened by the fact that Mozart clearly creates two chunks using an octave displacement, making the reading of two consecutive Prinner quite a bit stronger. Similarly, the stronger reading of Prinner in the exposition may result in difficulties hearing any other interpretation in the recapitulation; the exposition sets the frame of reference for the second rotation.

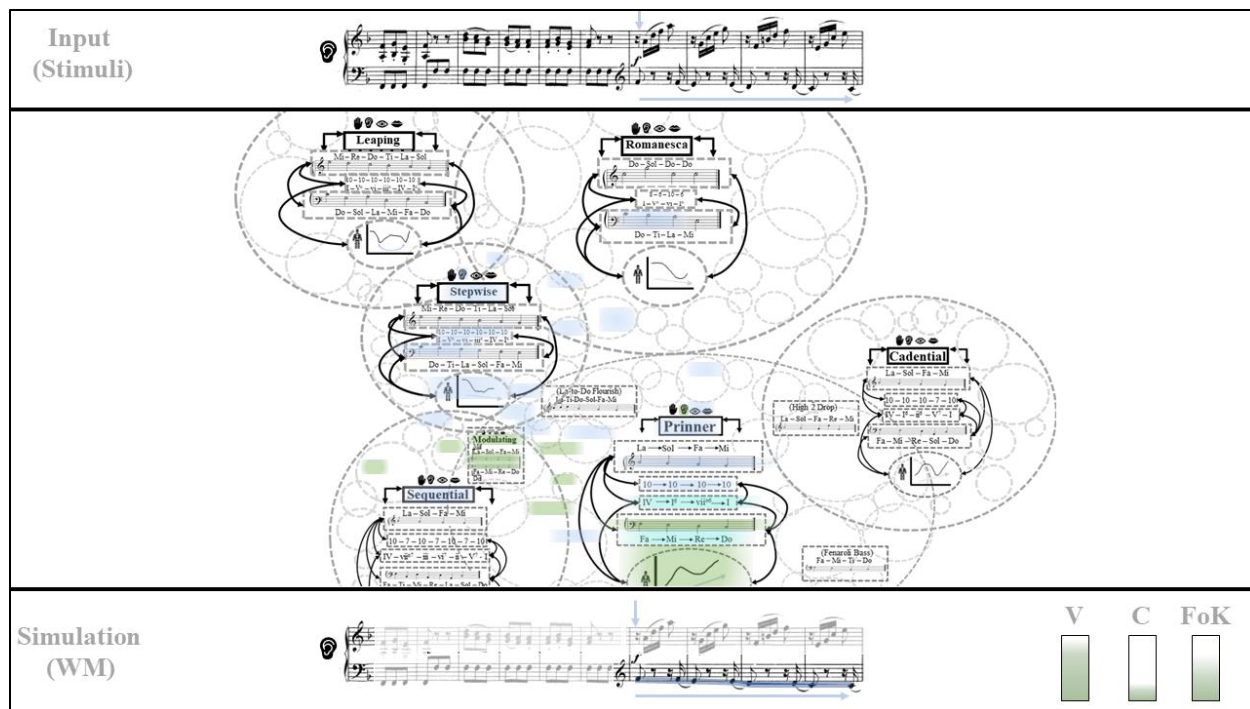


Figure 5.14. Direct Representational Activation of Prinner Traces and Partial Priming of Romanesca Traces in K. 280, iii, Exposition

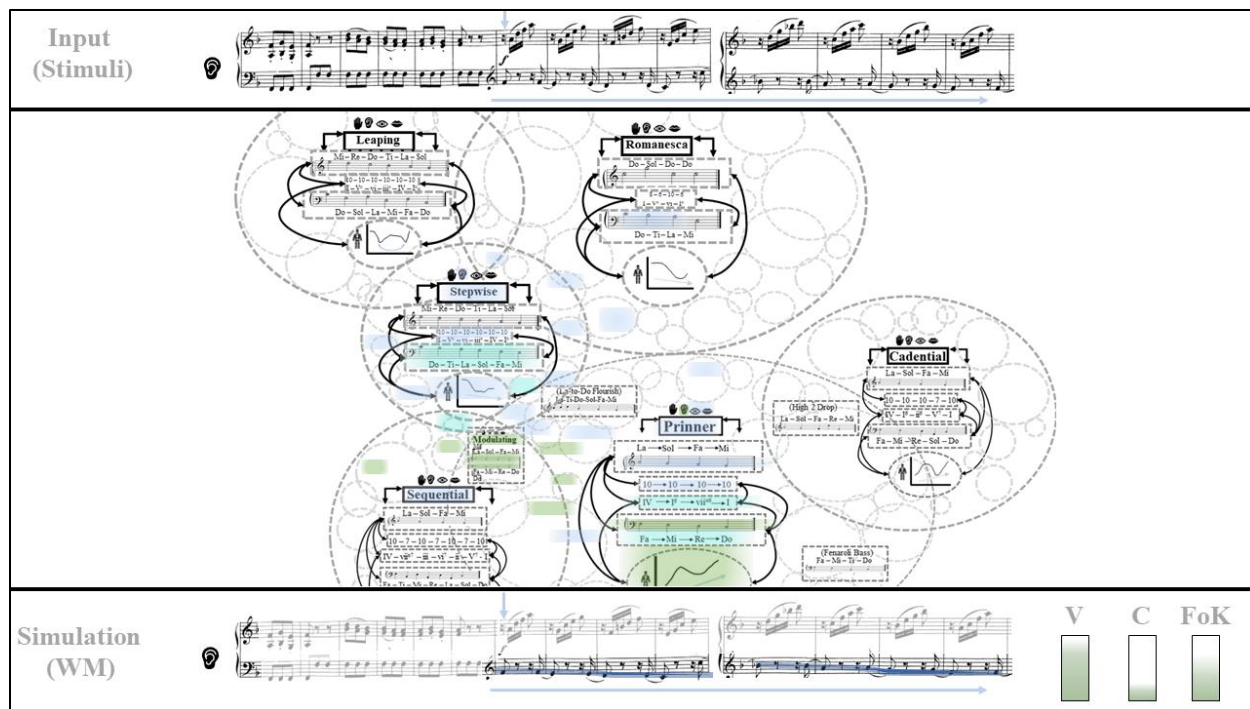
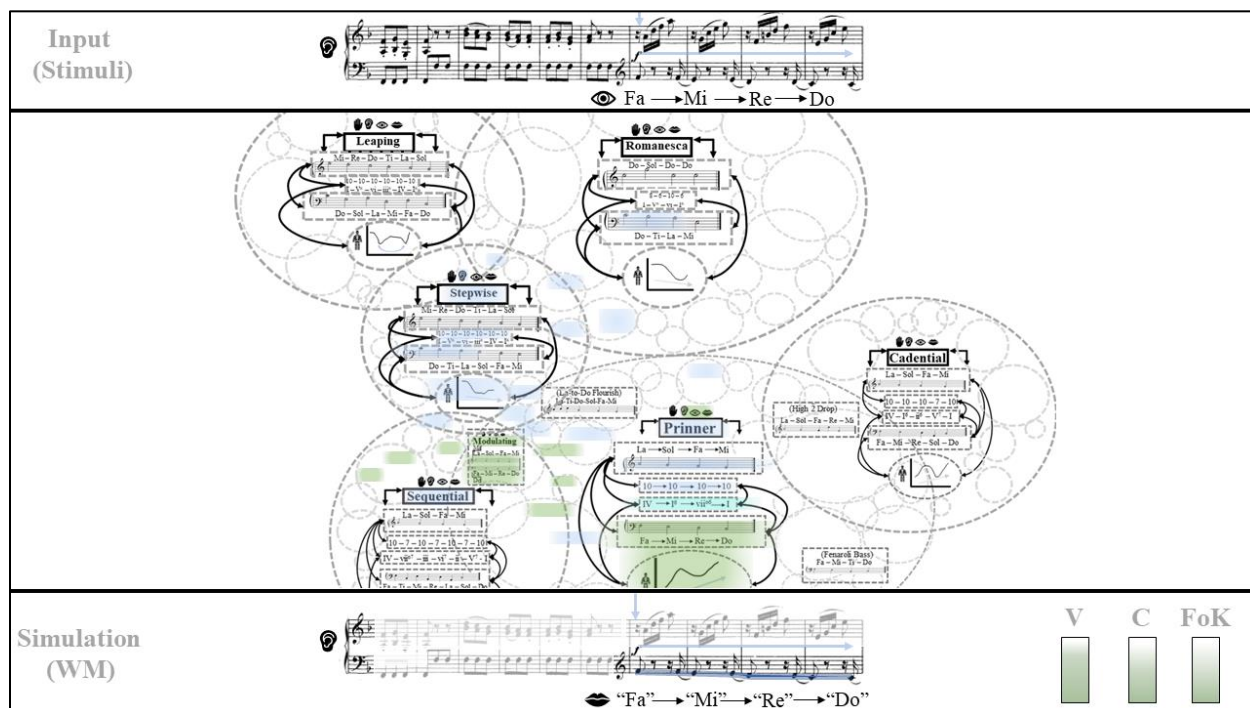


Figure 5.15. Direct Representation Activation of Prinner Traces and Increased Activation of Romanesca Traces in K. 280, iii, Recapitulation

The second hypothesis posits that ease of hearing and change ratings will vary with attending strategy, as a change in attention activates different memory representations. Single line attending (bass, soprano) will provide higher ease of hearing and ease of change ratings than diffuse attending (schema), as the single line attending activates a smaller pool of simulators than does diffuse attending. Essentially, the only traces that are ‘essential’ for a single-line simulator to be active are a smaller subset of traces that encompass previous interactions with single-line voices (e.g., sung interactions, auditory imagens, auditory-motor logogens). Other traces will certainly be active through associational and representational activation, but they do not need to be accessed and maintained in WM to make an evaluation about the goodness-of-fit for a single-line simulator (see Figure 5.16a). Hearing the line “Fa-Mi-Re-Do” is relatively simple, shown by the high vividness, control and feeling-of-knowing in WM in Figure 5.16a. Changing this

interpretation to one that is “Do-Ti-La-Sol” may be difficult, but not impossible. This requires suppression of the traces that were just active and held in WM (i.e., bass line simulators). The invocation of the verbalization “Do-Ti-La-Sol” helps to activate by referential activation, auditory imagens pertaining to that simulator (see Figure 5.16b). Forming this second tonic-based interpretation may prove more difficult as fewer traces are active pertaining to this interpretation. The difficulty in forming the interpretation is represented in Figure 5.16b by the lower perceived control and feeling-of-knowing assessments in WM.

(a). Direct Representational Activation and Referential Processing of Bass Line “Fa-Mi-Re-Do”



## (b). Alternation to Bass Line Interpretation “Do-Ti-La-Sol”

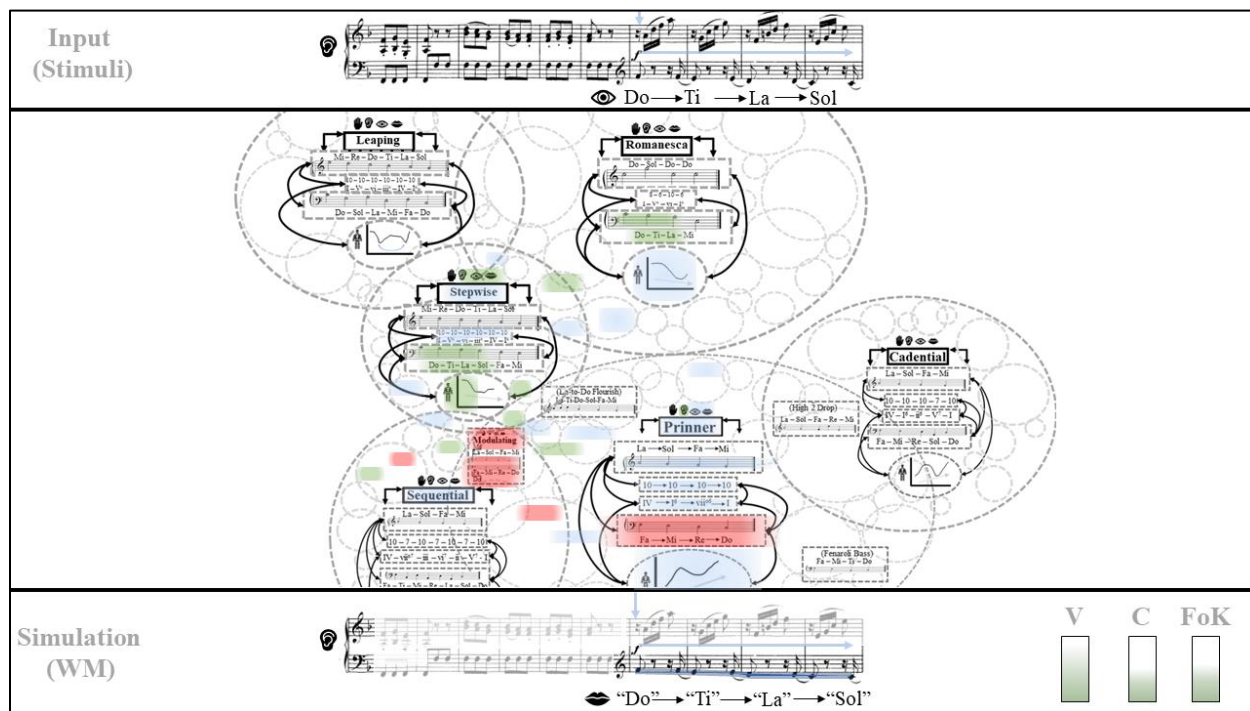


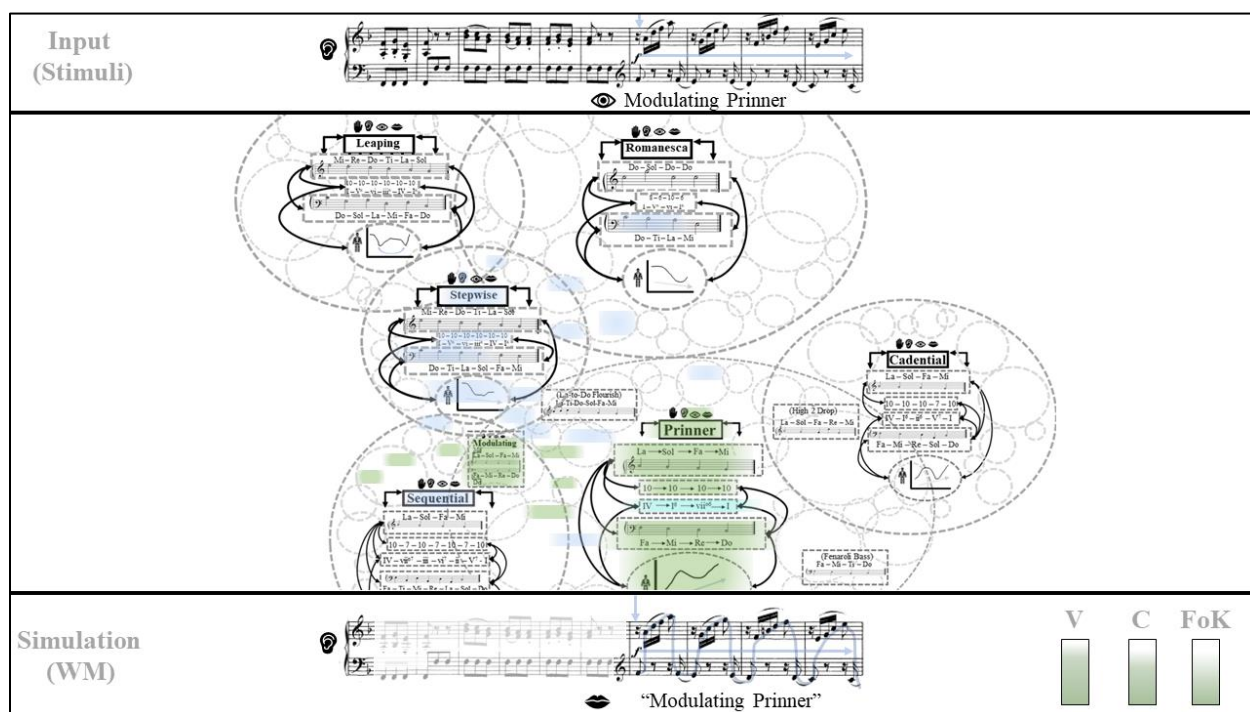
Figure 5.16. Formation of Modulating (a) and Change to Nonmodulating (b) Bass Line Interpretations in K. 280, iii, Exposition

Contrastingly, simulating and evaluating the entire category “Prinner” or “Romanesca” requires many more simulators to be active at any given time (e.g., bass line, harmony, soprano line) (see Figure 5.17a,b). While this may be relatively easy when the presented example activates previous traces for that category through representational activation, as is the case with a Prinner interpretation (see Figure 5.17a), this is much more challenging for a Romanesca interpretation. More WM control is required for a Step-Descent Romanesca interpretation because the listener will need to suppress the many Prinner traces that automatically activate through listening. At the same time, the listener will also need to bring online (through brute force associational or referential processing) other simulators for the Romanesca (e.g., soprano voice) that may not be active through representational activation alone (see Figure 5.17b). In



fact, the listener may indeed be unsuccessful at actively inhibiting Prinner simulator traces, contributing to difficulty in hearing a Romanesca interpretation.

(a). Direct Activation of Modulating Prinner



## (b). Suppression of Modulating Prinner Traces for Romanesca Interpretation

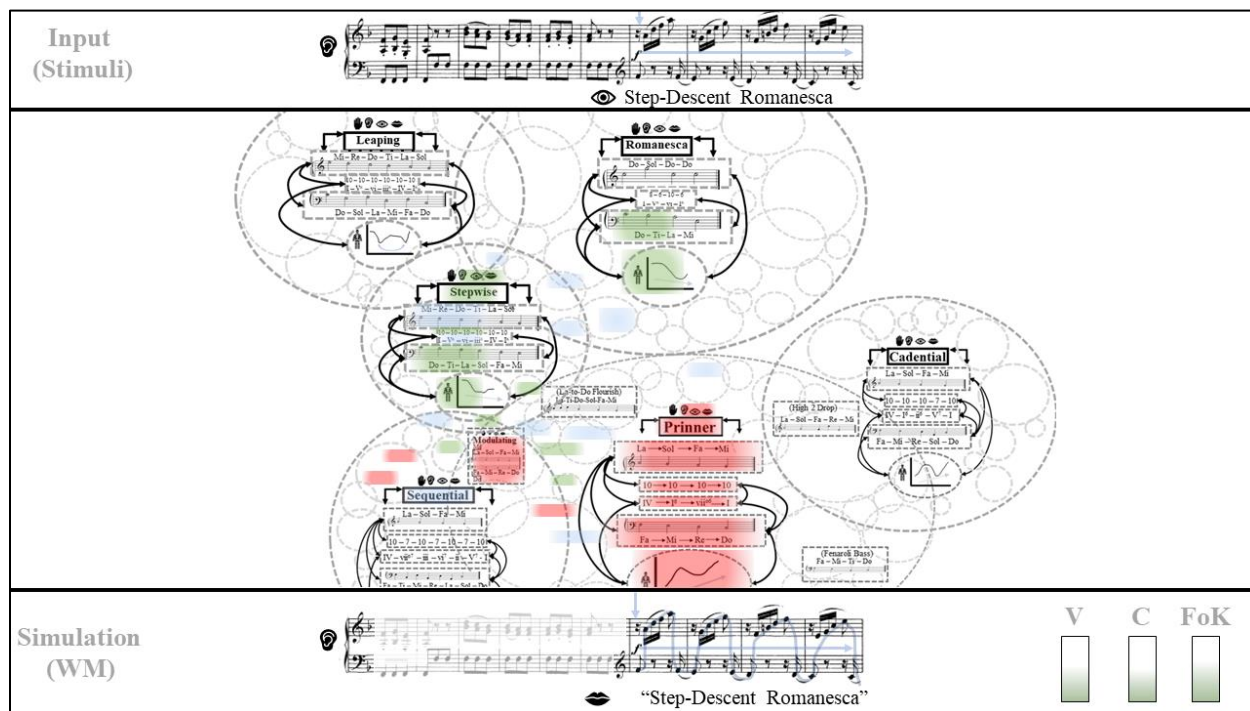


Figure 5.17. Formation of Modulating Prinner (a) and Step-Descent Romanesca (b) Schema Interpretations in K. 280, iii, Exposition

The final hypothesis states that the extent to which different interpretations can be formed and changed will vary with expertise: specifically, those with intermediate level expertise will be able to form and change between two schemata interpretations far more easily than will experts. A higher level of expertise should be associated with more elaborate simulators stored in memory. As a result, higher expertise should demonstrate a rigidity of interpretation, particularly for Galant schemata, as these categories are more differentiated in memory based both on their primary features (i.e., similarity of the stimuli, or ‘bottom-up’ factors), and spatial-temporal location in Sonata form (i.e., so-called ‘top-down’ factors pertaining to the probability of activation of simulator pools over time).

This is due to several factors pertaining to the structure of expert memory. Firstly, as experts' memory is more elaborated, the *distinguishing* features of Modulating Prinner and Step-Descent Romanescas in sonata form transitions are more prominent. Because of this, representational activation will be far narrower for experts, meaning that the traces that are directly activated by listening alone will be fewer and more specific. For example, Figure 5.18 shows the formation of a schema interpretation for a hypothetical expert. Here, the traces *directly* activated by auditory input are more narrowly within the modulating Prinner subcategory, and even fewer traces within the potential feature overlap with the Romanesca are primed. Experts' memory organization therefore may simply not facilitate activation of Romanesca traces, along with suppression of Prinner traces (see Figure 5.19). This effect may also stem from or be assisted by more 'top-down' features outside the stimuli. One example would be experts' experience with how schemata are employed in musical forms. Because the Modulating Prinner tends to occur with the onset of transitional rhetoric, memory traces pertaining to the Prinner may have a higher probability of activation when an expert recognizes—overtly or not—that a transition has begun. This would not be the case with the Romanesca, which more typically occurs *after* transitional rhetoric is underway. Listeners with intermediate, as opposed to expert, levels of schema experience should show a different pattern. Since the distinguishing features ('bottom up') and distributional sensitivities ('top-down') of schemata have not yet been fully acquired and honed, listeners with some, but not a great deal, of schema expertise will likely have an easier time in both forming and alternating between the two schema interpretations.

**Input (Stimuli)**

Modulating Prinner

**Simulation (WM)**

“Modulating Prinner”

V C FoK

Figure 5.18 illustrates the direct representational activation of Modulating Prinner traces in a hypothetical expert. The figure is divided into three horizontal sections. The top section, labeled 'Input (Stimuli)', shows a musical score for 'Modulating Prinner'. The middle section contains a complex network of interconnected nodes, each representing a different type of melodic trace: 'Leaping', 'Romanesca', 'Stepwise', 'Sequential', 'Prinner', and 'Cadential'. Each node includes a small musical score snippet, a list of notes (e.g., 'Mi- Re- Do- Ti- La- Sol'), and a waveform diagram. The bottom section, labeled 'Simulation (WM)', shows the same musical score as the input, but with a red lip icon below it, indicating activation. To the right of the simulation are three vertical bars labeled 'V', 'C', and 'FoK', representing different levels of activation or processing.

Figure 5.18. Direct Representational Activation of Modulating Prinner Traces of a Hypothetical Expert

**Input (Stimuli)**

Step-Descent Romanesca

**Simulation (WM)**

V C FoK

“Step-Descent Romanesca”

Figure 5.19. Suppression of Modulating Prinner Traces and Attempted Indirect Activation of Step-Descent Romanesca Traces of a Hypothetical Expert

## Method

Data was collected online in Qualtrics using qualitative survey methods. This study uses methods inspired by those used in ambiguous figure perception in the domain of vision (e.g., Brugger, 1999). Participants were presented with two short musical excerpts in auditory format—the transitions from the exposition and recapitulation of Mozart’s Piano Sonata no. 2, K. 280, iii—after which they were prompted to rate the ease in forming two different interpretations presented in visual format using scale degrees and/or Galant schemata labels. Following this, they were then asked to alternate between the two provided interpretations and rate the ease of change. Two types of attending strategies were investigated: 1) focused attention through scale-degree interpretations of either soprano or bass lines (i.e., a single auditory stream; Bregman, 1990), and 2) diffuse or shifting attention, prompted through assessing an interpretation of an

ordering of Galant schemata. Unlike focusing on a scale degree interpretation in a single voice, hearing a passage as a schema requires attending to multiple, co-occurring features at once. Because the latter requires more attentional resources (WM) and more active simulator traces (LTM), hearing and alternating schema interpretation is likely more challenging. The final portion of the survey collected general information regarding the participants' music theory training (years of training), their experience with scale degrees and Galant schemata, and their familiarity with the sonata movement in the study. Taken together, the information gathered in the first part of the study addresses the first and second research questions above (i.e., whether the excerpts are bistable, and if bistability changes with conceptual prompt and auditory attending strategy), and the biographical information regarding music theory training gathered at the end of the survey addresses the final research question (i.e., whether the results observed depend on previous training and experience). The study was reviewed and approved by Northwestern's Institutional Review Board, Study IRB# 00217214.

*Participants.* A total of nineteen persons ( $n = 19$ ) participated in the study. Of these, ten were professors of music theory, six were current graduate students in music theory, two were post-doctoral students in music theory, and one was a recent PhD graduate in music theory. Out of the 19 participants, a total of 3 reported having perfect pitch (all professors). Participants were recruited online through email (Music Theory and Cognition listserv in the Department of Music Studies at Northwestern, the Society for Music Theory Pedagogy Interest Group listserv, and by personal email contact) and online posts (Society for Music Theory Humanities Commons page).

*Stimuli.* The stimuli were two short excerpts from the expositional and recapitulatory transitions of Mozart's Piano Sonata no. 2, K. 280, iii. The audio clips were extracted from Kristian Bezuidenhout's *Mozart: Keyboard Music*, vols. 8 & 9 (2016) for fortepiano using Audacity. The expositional transition included the end of the main theme through the medial caesura (see Figure 5.20) totaling 16 seconds (from 0:08-0:24 in the recording), and the recapitulatory transition contained the same adjusted material (see Figure 5.21) totaling 19 seconds (from 2:09-2:28 in the recording). Two interpretations for each excerpt were created for bass and soprano lines, as well as orderings of Galant schemata (see Figure 5.22a,b, and c).

Figure 5.20. Expositional Transition of K. 280, iii

Figure 5.21. Recapitulatory Transition of K. 280, iii

## (a). Bass Line Interpretations

$$[\text{PAC} - \text{Tonic Pedal}] \rightarrow \hat{1} - \hat{7} - \hat{6} - \hat{5} - \hat{1} \left| \begin{array}{c} \% \\ \hat{4} - \#4 - \hat{5} \end{array} \right.$$

$$[\text{PAC} - \text{Tonic Pedal}] \rightarrow \hat{1} \left| \begin{array}{c} \% \\ \hat{4} - \hat{3} - \hat{2} - \hat{1} - \hat{4} - \#4 - \hat{5} \end{array} \right.$$

## (b). Soprano Line Interpretations

$$[\text{PAC} - \text{Tonic Pedal}] \rightarrow \hat{3} - \hat{2} - \hat{1} - \hat{7} - \hat{6} \left| \begin{array}{c} \% \\ \hat{2} - \hat{1} - 7 \end{array} \right.$$

$$[\text{PAC} - \text{Tonic Pedal}] \rightarrow \hat{3} \left| \begin{array}{c} \% \\ \hat{6} - \hat{5} - \hat{4} - \hat{3} - \hat{2} - \hat{1} - \hat{7} \end{array} \right.$$



## (c). Schemata Interpretations

PAC – Tonic Pedal – Step-Descent Romanesca – Converging

PAC – Tonic Pedal – Modulating Prinner – Converging

Figure 5.22. Visually Presented Interpretations for the Bass Line (a), Soprano Line (b) and Schemata (c) Interpretations of the Exposition

*Procedure.* The survey consisted of three parts: consent documentation and introduction, the experimental body of the survey (K. 280 ratings), and biographical information collection. At the beginning of the survey, participants read a letter of information and completed the consent documentation. They then completed an introductory training module, which introduced bistable perception through the original duck-rabbit figure drawn and presented in Jastrow (1899) (see Figure 5.23). They completed the task using this picture. Firstly, they were asked to form the interpretation RABBIT, and rate the ease of forming this interpretation on a Likert scale from 1 (cannot see at all) to 7 (can clearly see it). Following this, they were presented with a second interpretation, DUCK, and completed the same procedure. Participants were then prompted to view the image continuously and rate the ease of changing the interpretation from RABBIT to DUCK on a Likert scale from 1 (cannot alternate interpretations) to 7 (can easily alternate interpretations). Lastly, participants replicated the task in a practice session with a musical stimulus—the modulating transition from the exposition of Mozart’s Piano Sonata no. 7, K. 309, i. They were asked to rate the ease of hearing two bass line interpretations (see Figure 5.24) on a Likert scale from 1 (cannot hear at all) to 7 (can easily hear), as well as the ease of changing these two interpretations (1-7).

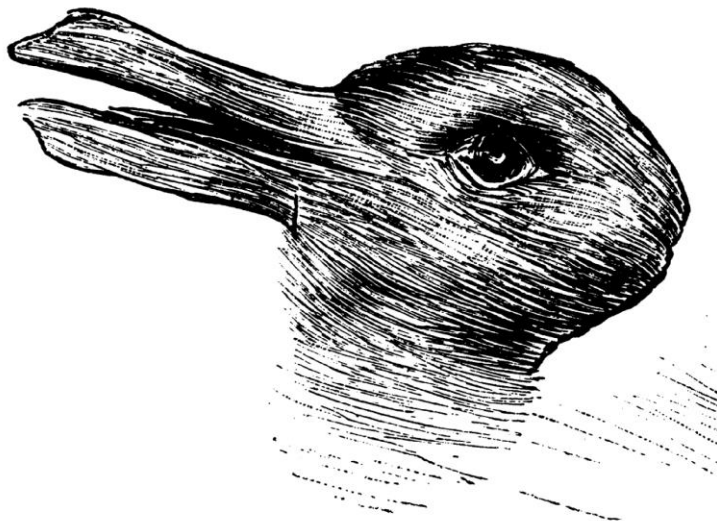


Figure 5.23. Duck-Rabbit Ambiguous Figure from Jastrow (1899)

$$\hat{1} \text{ \%} - \hat{7} \text{ \%} - \# \hat{4} \text{ \%} - \hat{5}$$

$$\hat{4} \text{ \%} - \hat{3} \text{ \%} - \hat{7} \text{ \%} - \hat{1}$$

Figure 5.24. Bass Line Interpretations for K. 309 used in the Survey Introduction

The experimental body of the survey gathered ease of hearing and ease of change ratings for the two excerpts from the Presto of K. 280. The survey was organized into two main blocks (exposition, recapitulation), which contained three blocks for each attentional condition (bass line, soprano line, schemata). Each attentional condition (bass, soprano, schemata) contained three trials each, one rating for each interpretation, and one rating for the ease of change between the two presented interpretations. This resulted in a total of 18 trials. The presentation order of the excerpts was randomized such that the exposition or recapitulation blocks could be presented in either order, the attentional condition blocks (bass, soprano, schema) could be presented in any order, and within each condition, the order of presentation of each interpretation trials (e.g.,

Prinner, Romanesca) was randomized. Before starting the first set of ratings in either the exposition or recapitulation, participants were prompted to listen to the excerpt without any interpretation provided. Once within a given block (e.g., Recapitulation) and condition (e.g., Schema), participants completed both ease of hearing and ease of change ratings before moving onto the next condition (e.g., bass line). Once all three conditions were completed (bass, soprano, schema), they moved onto the next main block (e.g., exposition). Participants were provided with an audio clip of the excerpt for each rating and could listen as many times as they liked. For the ease of change ratings, participants were encouraged to listen to the excerpt multiple times and to change interpretations on each successive listening.

The final part of the survey gathered various biographical information that was used to create expertise designations (see *Data cleaning and pre-processing* below). Participants were prompted to provide their education level, number of years of music theory and aural skills experience (both training and teaching), expertise in scale degree hearing and with Galant schemata, their familiarity with the excerpt (hearing and analysis), if they had perfect pitch, and what types of strategies they used for forming and changing interpretations in the task they had just completed.

*Data cleaning and pre-processing.* The raw data exported from Qualtrics was cleaned in python using pandas and numpy libraries. The data was aggregated into long format, which resulted in a total of 228 observations for ease of hearing ratings (DV1) and 114 observations for ease of change ratings (DV2). Data pre-processing included the labeling of block and trial orders (e.g., Sonata Order 1: exposition, recapitulation. Sonata Order 2: recapitulation, exposition), as well as the creation of schemata expertise designations from biographical data provided on schema familiarity and schema hearing. Schema familiarity involved a categorical response:

‘Completely unfamiliar’, ‘Somewhat familiar’, ‘Familiar’, ‘Very familiar’, and ‘I am an expert (analysis, playing, composition, etc.).’ Schemata hearing was measured on a Likert scale from 1 (Strongly disagree to 7 (Strongly agree) where participants rated the statement “I can hear Galant Schemata while listening.” Expertise groups for schemata therefore involved an aggregate of both responses. For the categorical response, those responding as experts were classified as experts, those familiar and very familiar were classified as intermediate, and those somewhat familiar and completely unfamiliar classified as novices. The ordinal schemata hearing response was used to verify these classifications. To remain in the intermediate level, participants required a schemata hearing value between 3-6, otherwise they were placed in the novice group. Those with the categorical response ‘somewhat familiar’ whose schema hearing level was 3 or higher were placed in the intermediate group. This resulted in three participants with ‘somewhat familiar’ responses to be placed in the intermediate group ( $M_{\text{Schema\_Hearing}} = 4.53$ ,  $sd = 0.45$ ) and two participants with ‘somewhat familiar’ responses to be placed in the novice group ( $M_{\text{Schema\_Hearing}} = 1.50$ ,  $sd = 0.51$ ). One participant was manually placed in the expert group whose categorical response was ‘very familiar’ and schema hearing value was 6.7 because this participant was specifically contacted by the author for their expertise in Galant schemata.

A summary of the schemata expertise group data is shown below (see Figure 5.25). Each group did not appear to differ greatly on their self-reported ability to hear scale degrees, indicated by the scale degree hearing ratings in the figure below. Average schema hearing levels scaled by group, with experts having the highest ratings ( $n = 5$ ,  $M_{\text{Schema}} = 6.38$ ,  $sd = 1.10$ ,  $M_{\text{Scale\_Degree}} = 6.64$ ,  $sd = 0.46$ ), novices having the lowest ( $n = 4$ ,  $M_{\text{Schema}} = 1.25$ ,  $sd = 0.43$ ,  $M_{\text{Scale\_Degree}} = 5.50$ ,  $sd = 1.51$ ), and intermediates in between ( $n = 9$ ,  $M_{\text{Schema}} = 4.85$ ,  $sd = 1.10$ ,  $M_{\text{Scale\_Degree}} = 6.01$ ,  $sd = 0.65$ ).

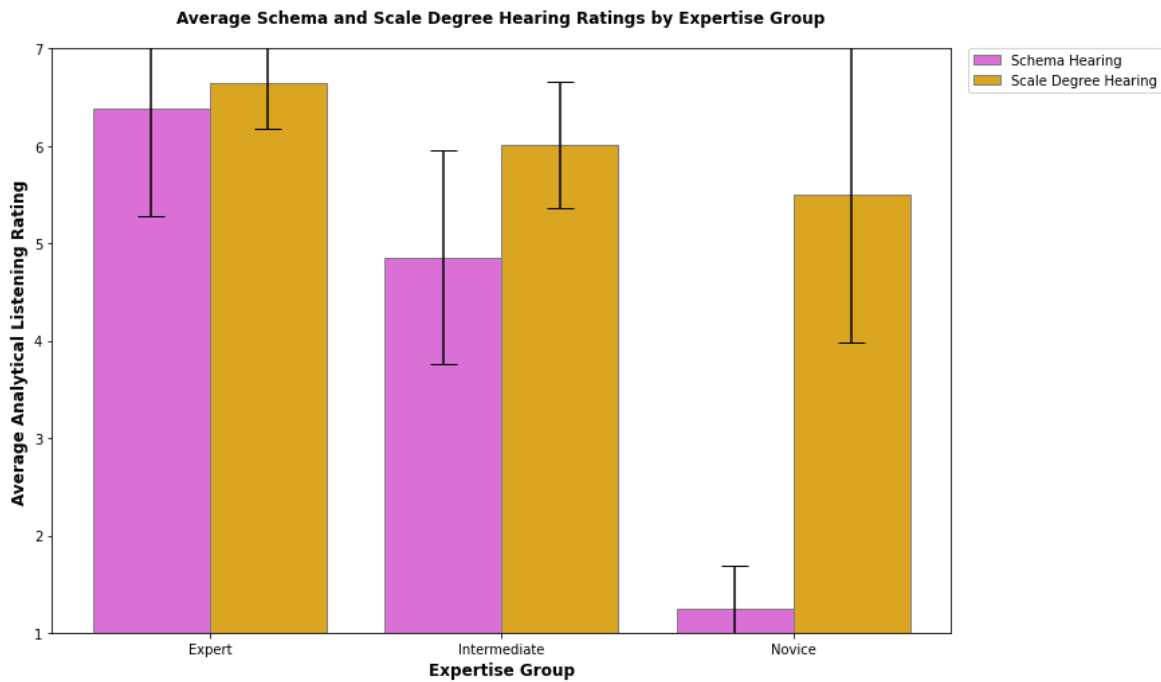


Figure 5.25. Average Schema Hearing and Scale Degree Hearing Ratings for Each Expertise Group

## Results

The survey design resulted in a four-way mixed ANOVA design, with  $2 \times 3 \times 2$  factorial design for the within-subject factors, and three levels of the between-subject factor. The within-subject factors included two levels of sonata section (exposition, recapitulation), three levels of attended feature (bass, soprano, schema) and two levels of modulation type (Prinner, Romanesca). The between-subject factor included the three levels of expertise group (expert, intermediate, novice). Upon exploratory analysis of the results, one participant in the expert group was dropped due to internal inconsistencies in their responses.

*Ease of hearing ratings (DVI).* A four-way mixed ANOVA was performed in R using the *rstatix* package to evaluate the effects of sonata section, attended feature, modulation type and expertise on the ease of hearing ratings. The dataset included sphericity violations for the within-subject factors as assessed by Shapiro-Wilk's test of normality for all conditions except for ratings in the exposition in the soprano condition for the Romanesca modulation type ( $p = 0.109$ ). As a result,  $p$ -values using Greenhouse-Geisser corrections (where applicable) were used. There was homogeneity of variances ( $p > 0.05$ ) as assessed by Levene's test of homogeneity of variances.

The analysis revealed main effects of expertise group,  $F(2, 15) = 6.81, p = 0.008$ , and attended feature,  $F(1.94, 29.10) = 11.37, p = 0.0002, \epsilon = 0.970$ . Post-hoc Tukey HSD tests revealed that Expert ( $M = 5.13, sd = 2.12$ ) and Novice ( $M = 3.56, sd = 2.17$ ) groups differed significantly,  $p < 0.001, 95\% \text{ C.I.} = [-2.44, -0.696]$ , and that Intermediate ( $M = 5.40, sd = 1.66$ ) and Novice ( $M = 3.56, sd = 2.17$ ) groups differed significantly,  $p < 0.001, 95\% \text{ C.I.} = [-2.63, -1.05]$  (see Figure 5.26). For attended feature, a post-hoc Tukey HSD test showed that the main effect of feature comes from a statistically significant difference between the Bass ( $M = 5.45, sd = 1.65$ ) and Soprano ( $M = 4.35, sd = 2.12$ ) features,  $p = 0.003, 95\% \text{ C.I.} = [-1.89, -0.313]$  (see Figure 5.27). The main effect for sonata section was nonsignificant,  $F(1, 15) = 0.009, p = 0.926$ , indicating that ratings for the exposition ( $M = 4.93, sd = 2.04$ ) were equal to the recapitulation ( $M = 4.90, sd = 2.05$ ). The main effect for modulation type was also nonsignificant,  $F(1, 15) = 4.41, p = 0.053$ , indicating that ratings for the Prinner ( $M = 5.21, sd = 1.99$ ) did not differ significantly from those for the Romanesca ( $M = 4.62, sd = 2.06$ ).

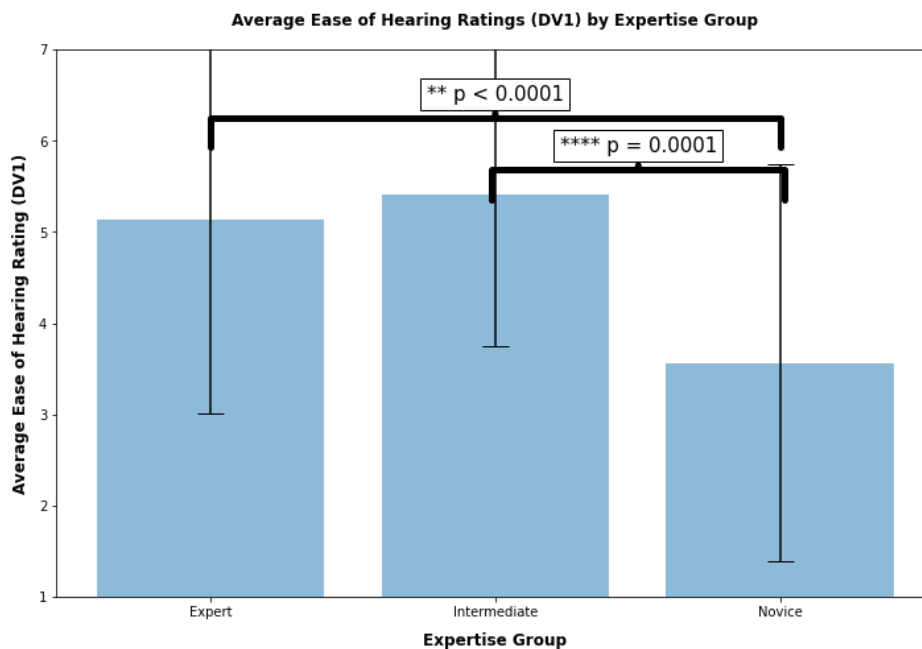


Figure 5.26. Average Ease of Hearing Ratings (DV1) by Expertise Group

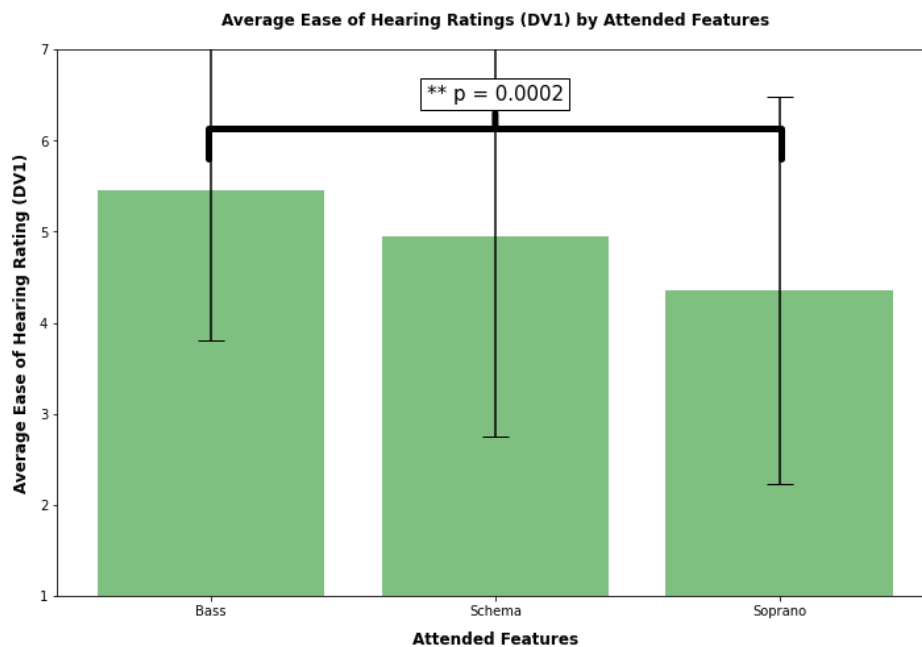


Figure 5.27. Average Ease of Hearing Ratings (DV1) by Attended Feature

The four-way mixed ANOVA analysis also showed four significant two-way interactions, but no significant three- or four-way interactions. There were two significant interactions that involved within-subject factors only, and two that involved interactions between the within- and between-subject factors (i.e., effects of expertise). For the within-subject factor interactions, the analysis revealed a significant interaction of sonata section by attended feature,  $F(1.64, 24.61) = 6.34, p = 0.009$ , which was corrected for sphericity violations ( $\epsilon = 0.820$ ). A post-hoc Tukey HSD test revealed that the Bass ( $M = 5.60, sd = 1.47$ ) and Soprano ( $M = 4.12, sd = 2.17$ ) differed significantly only for the recapitulation,  $p = 0.005$ , 95% C.I. = [-2.58, -0.364] (see Figure 5.28). There was also a significant interaction between attended feature and modulation type  $F(1.95, 29.10) = 4.75, p = 0.017$ , which was corrected for sphericity violation using the Greenhouse-Geisser correction ( $\epsilon = 0.970$ ). A post-hoc Tukey HSD test revealed that only for the Schema attending condition, the Prinner ( $M = 5.73, sd = 1.86$ ) was rated higher than the Romanesca ( $M = 4.16, sd = 2.23$ ),  $p = 0.001$ , 95% C.I. = [-2.53, -0.602] (see Figure 5.29). The two-way interaction between sonata section and modulation type was nonsignificant,  $F(1, 15) = 3.45, p = 0.083$ . Both Prinner and Romanesca were equally available in the exposition ( $M_{Prinner} = 5.33, sd = 1.93$  |  $M_{Romanesca} = 4.54, sd = 2.09$ ) and recapitulation ( $M_{Prinner} = 5.08, sd = 2.06$  |  $M_{Romanesca} = 4.71, sd = 2.04$ ).



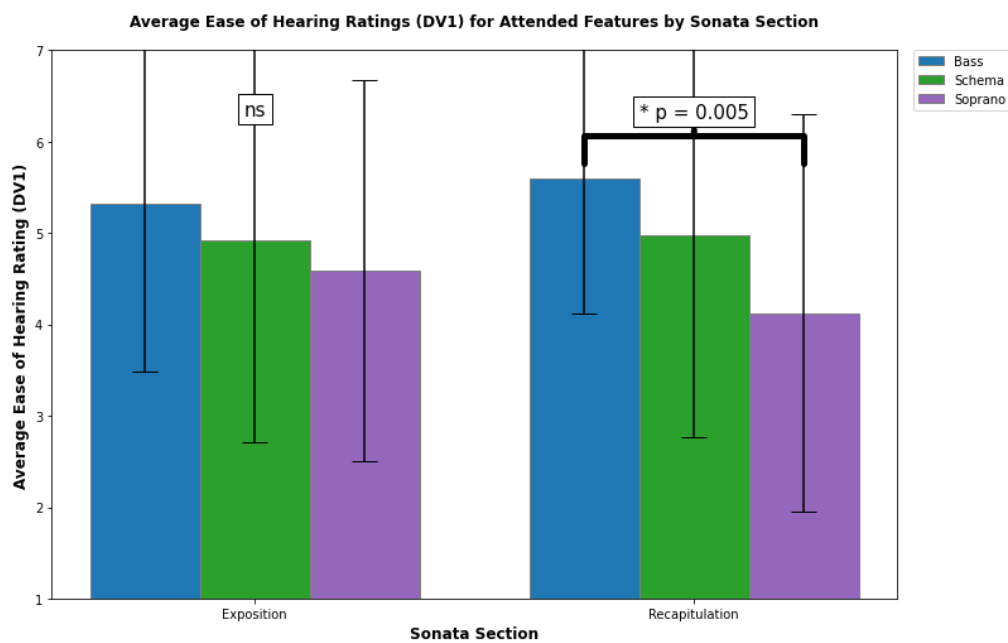


Figure 5.28. Average Ease of Hearing Ratings (DV1) for Attended Features by Sonata Section

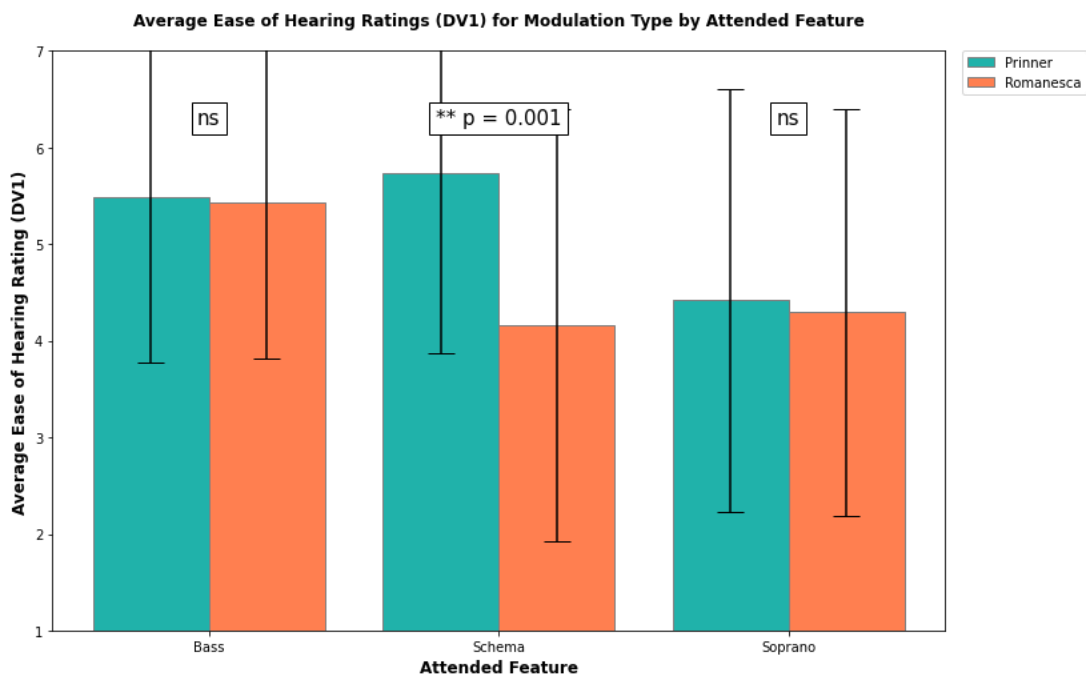


Figure 5.29. Average Ease of Hearing Ratings (DV1) for Modulation Type by Attended Feature

Lastly, the analysis revealed two significant interactions involving the between-subject factor, expertise group. Firstly, there was a significant two-way interaction between expertise group and attended feature,  $F(3.88, 29.10) = 3.12$ ,  $p = 0.031$ , which was corrected for sphericity violations ( $\epsilon = 0.970$ ). A post-hoc Tukey HSD test showed that for the Novice group only, the Bass ( $M = 4.83$ ,  $sd = 2.07$ ) and Soprano ( $M = 2.51$ ,  $sd = 1.62$ ) conditions differed significantly from one another,  $p = 0.005$ , 95% C.I. = [-4.03, -0.619] (see Figure 5.30). Secondly, there was a significant two-way interaction of expertise group and modulation type,  $F(2, 15) = 4.25$ ,  $p = 0.034$ . A post hoc Tukey HSD revealed that that for the Expert group only, the Prinner ( $M = 6.24$ ,  $sd = 1.49$ ) was rated higher than the Romanesca ( $M = 4.02$ ,  $sd = 2.08$ ),  $p < 0.0001$ , 95% C.I. = [-3.15, -1.28] (see Figure 5.31). The two-way interaction between expertise group and sonata section was nonsignificant,  $F(2, 15) = 0.803$ ,  $p = 0.467$ , showing that ratings for the exposition and recapitulation did not differ between groups: Expert ( $M_{\text{expo}} = 5.30$ ,  $sd = 2.08$  |  $M_{\text{recap}} = 4.96$ ,  $sd = 2.17$ ), Intermediate ( $M_{\text{expo}} = 5.42$ ,  $sd = 1.61$  |  $M_{\text{recap}} = 5.39$ ,  $sd = 1.72$ ), Novice ( $M_{\text{expo}} = 3.40$ ,  $sd = 2.19$  |  $M_{\text{recap}} = 3.72$ ,  $sd = 2.19$ ). All three- and four-way interactions were nonsignificant.

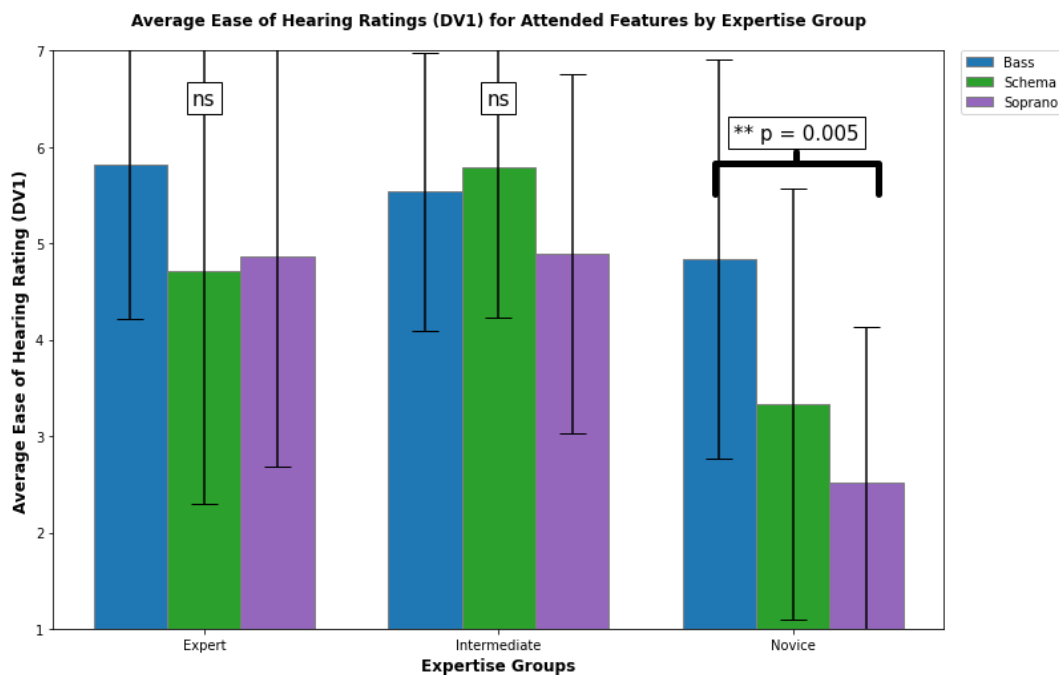


Figure 5.30. Average Ease of Hearing Ratings (DV1) for Attended Features by Expertise Group

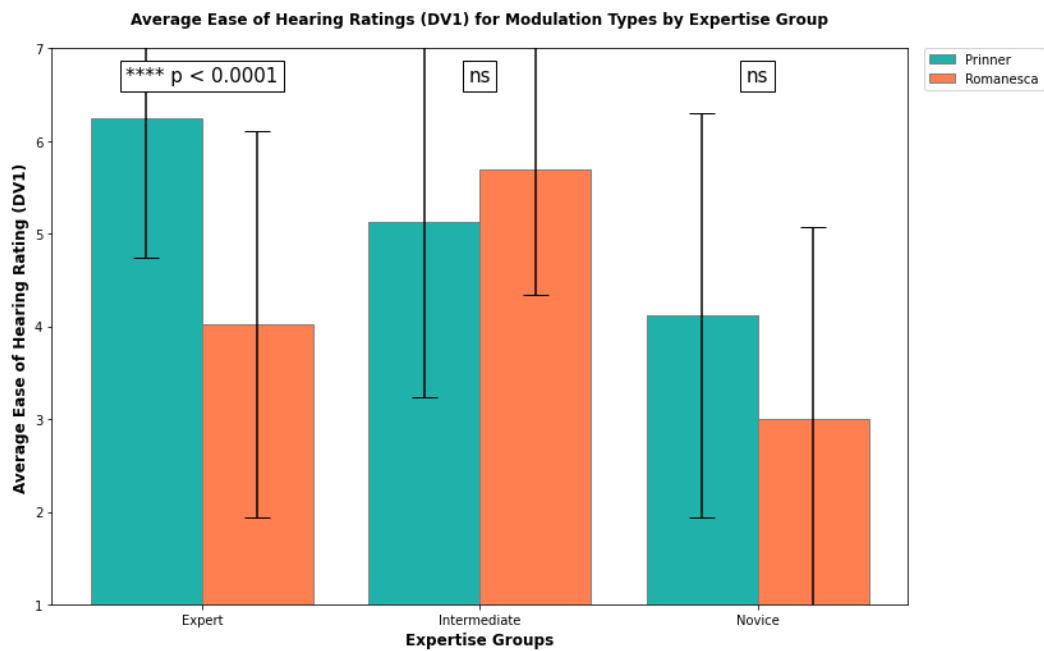


Figure 5.31. Average Ease of Hearing Ratings (DV1) for Modulation Type by Expertise Group

*Ease of change of interpretation ratings (DV2).* A second three-way mixed ANOVA was performed to examine the effects of expertise group (expert, intermediate, novice), sonata section (exposition, recapitulation), and attended feature (bass, soprano, schema) on ease of change ratings (DV2). The dataset included sphericity violations for the within-subject factors as assessed by Shapiro-Wilk's test of normality for all conditions except for bass and schema attending conditions in recapitulation ( $P > 0.05$ ). As a result, p-values using Greenhouse-Geisser corrections were used for the analysis. There was homogeneity of variances ( $p > 0.05$ ) as assessed by Levene's test of homogeneity of variances. The analysis revealed significant main effects of expertise,  $F(2, 15) = 4.54, p = 0.029$ , and feature,  $F(1.51, 22.64) = 4.81, p = 0.026$ . A post hoc Tukey HSD revealed that both the expert and intermediate groups differed from the novice group: Expert ( $M = 3.57, sd = 2.06$ ) and Novice ( $M = 1.88, sd = 1.50$ ),  $p = 0.00273$ , 95% C.I. = [-2.89, -0.511], and Intermediate ( $M = 4.49, sd = 1.82$ ) and Novice ( $M = 1.88, sd = 1.50$ ),  $p < 0.0001$ , 95% C.I. = [-3.68, -1.55] (see Figure 5.32). For attended feature, a similar pattern to DV1 was observed where the Bass ( $M = 4.11, sd = 1.99$ ) was rated higher than the Soprano ( $M = 3.09, sd = 2.06$ ), but a post hoc Tukey HSD revealed that this was nonsignificant ( $p = 0.113$ ). The main effect of sonata section was nonsignificant,  $F(1, 15) = 0.262, p = 0.616$ , ( $M_{\text{Exposition}} = 3.79, sd = 2.20$  |  $M_{\text{Recapitulation}} = 3.52, sd = 1.97$ ). All two-way interactions were nonsignificant.

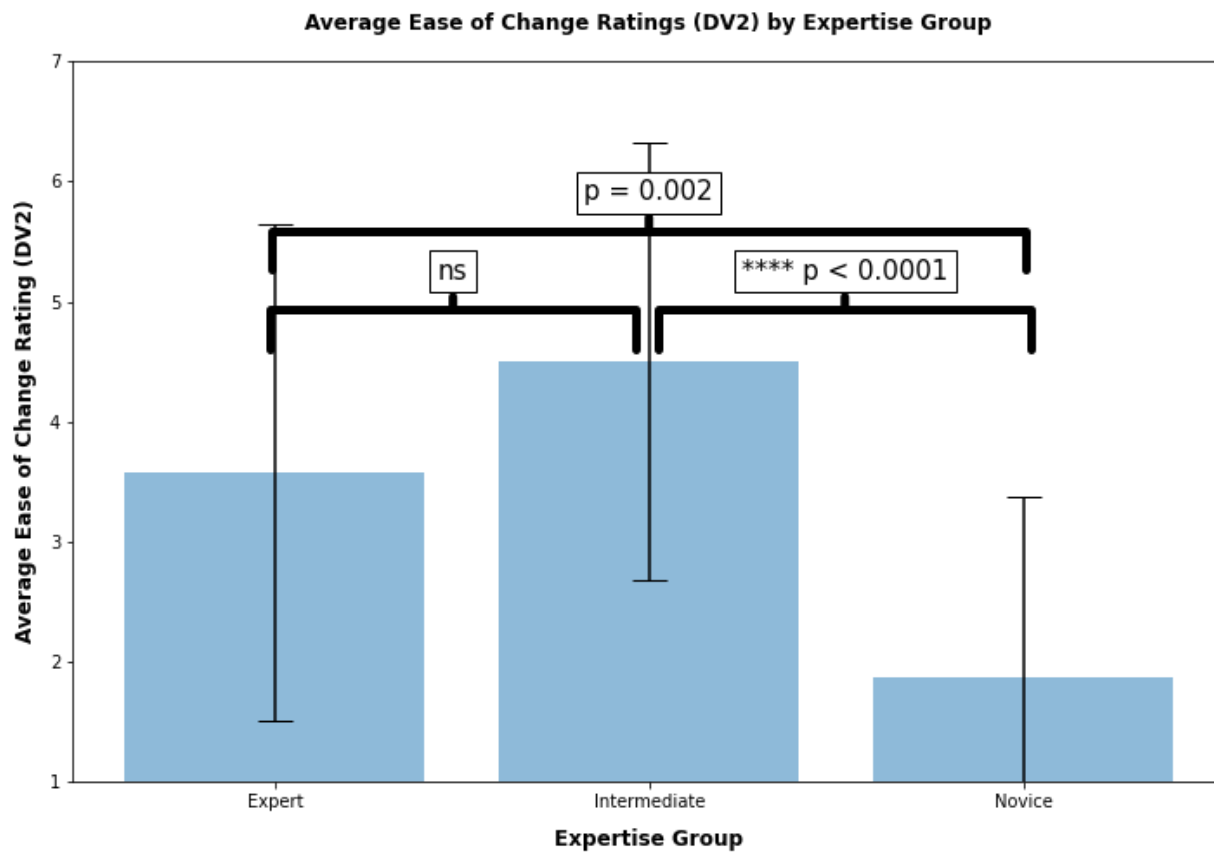


Figure 5.32. Average Ease of Change Ratings (DV2) by Expertise Group

*Excerpt familiarity and expertise.* The relationship between excerpt familiarity and expertise group was examined in order to ascertain if the observed effects of expertise group in the analyses above could be attributed to familiarity with the excerpt. It is plausible that higher familiarity with the piece could lead to rigidity in interpretation formation and ease of change. Two questions in the survey were designed to gather different types of familiarity ratings. First, one question addressed familiarity with the excerpt; participants selected one of three response types: “This was the first time I have heard this”, “I have heard this before”, “I am very familiar with this piece”. A second question gathered information on whether or not participants had analyzed the piece before. For this, participants selected one of three choices for having analyzed the piece: “No, not at all”, “Yes, somewhat”, or “Yes, extensively”. Table 5.1 shows the response breakdown by expertise group for both excerpt familiarity and excerpt analysis. For the expert group, all participants responded that they had heard the piece before; three indicated that they had not analyzed the piece and two indicated that they had somewhat analyzed the piece. For the intermediate group, four participants had never heard the piece before, four had heard it before, and one indicated that they were very familiar with the piece. For excerpt analysis, seven of the intermediate group indicated that they had never analyzed the piece, while the remaining two indicated that they had somewhat analyzed it. In the novice group, three participants indicated that they had never heard the piece before, and one indicated that they had heard it before. All participants in the novice group indicated that they had never analyzed the piece. No participants indicated that they had extensively analyzed the piece. From this data, familiarity categories were created (see Table 5.2). These categories were created using the following criteria:

- Low Familiarity: Indicated “This was the first time I have heard this” for excerpt familiarity and “No, not at all” for excerpt analysis.
- Mid Familiarity: Indicated “I have heard this before” for excerpt familiarity and “No, not at all” for excerpt analysis.
- High Familiarity: Indicated “I have heard this before” for excerpt familiarity and “Yes, somewhat” for excerpt analysis OR “I am very familiar with this piece” and for excerpt familiarity and “Yes, somewhat” for excerpt analysis.

Table 5.1. Excerpt Familiarity and Analysis Responses by Expertise Group

Question Type	Responses	Expert	Intermediate	Novice
Excerpt Familiarity	“This was the first time I have heard this”	0	4	3
	“I have heard this before”	5	4	1
	“I am very familiar with this piece”	0	1	0
Excerpt Analysis	“No, not at all”	3	7	4
	“Yes, somewhat”	2	2	0
	“Yes, extensively”	0	0	0

Table 5.2. Excerpt Familiarity Categories by Expertise Group

Excerpt Familiarity Categories	Expert	Intermediate	Novice
Low	0	4	3
Medium	3	3	1
High	2	2	0

An ordinal logistic regression was performed in order to see if familiarity category (low, medium, high) was able to significantly predict expertise group (novice, intermediate, expert). The predictive (AIC = 455.23) model was not a better fit than the null model (AIC = 451.82,  $X^2 = 0.59$ ,  $p = 0.7423$ ), Pseudo  $R^2 = 0.001$ (McFadden). This indicates that excerpt familiarity is not

predictive of expertise group. Therefore, the observed results of expertise group above can be attributed to generalized schemata expertise, and not to familiarity with the excerpt itself.

## Discussion

Here, I will contextualize the findings in light of the original hypotheses and theoretical framework developed in this dissertation. Firstly, we can safely reject the null hypothesis:

**H0:** This excerpt is not amenable to multiple interpretations, for either scale degree interpretations or Galant schemata.

Both excerpts were amenable to multiple interpretations for both scale degree lines and Galant schemata.

Regarding the first hypothesis:

**H1a:** This excerpt is amenable to multiple interpretations in terms of schemata; however, one interpretation (Prinner) may be easier to hear than the other (Romanesca).

**H1b:** Furthermore, the availability of and ease of change between interpretations may differ between the expositional and recapitulatory versions of the excerpt (with the Romanesca more available in the recapitulation).

Both hypotheses 1a and 1b were rejected as there was no main effect of modulation type, nor an interaction between modulation type and sonata section. This demonstrates that the excerpt demonstrated relatively equal perceptual bistability between interpretations (Prinner, Romanesca), and that this did not vary between the exposition and recapitulation.



**H2:** Participants should be able to more easily form and alternate between scale degree interpretations for a single voice (soprano, bass) than for Galant schemata (which are dependent on the presence of multiple, co-occurring features).

The second hypothesis was partially supported. The significant main effect of attended feature showed that, overall, bass lines and Galant schemata were equally perceptible, but that the soprano line was much more difficult to hear. The significant interaction of sonata section and attended feature showed that this effect was largely due to differences between the bass and soprano line ratings (DV1) in the recapitulation. I interpret this effect as arising from difficulty in forming interpretations in the exposition in general. However, without a main effect of sonata section, it is difficult to confirm this. The significant interaction of attended feature and modulation type also supports the second hypothesis: early and late modulation interpretations are equally available when attending to the bass voice; however, when switching to a multi-feature attending strategy needed to assess Galant schemata interpretations, Prinner and Romanesca become much less equally available.

**H3:** The ability to form and alternate Galant schemata interpretations may be dependent on moderate familiarity, but not a high-level expertise with Galant schemata (Expertise categories = Novice, Intermediate, Expert). Expertise should therefore be related to increased rigidity of interpretation (Ease of Change on a scale from 1 to 7, low to high), particularly for Schema interpretations, as such categories are overlearned and more likely to be automatically active during listening.

The final hypothesis regarding expertise was supported in multiple ways. The significant main effects of expertise for both DV1 and DV2 showed that the Intermediate group had the

highest ease of hearing and ease of change ratings, confirming that they were more easily able to hear both interpretations, and were much more able to alternate between interpretations. The low ease of change (DV2) ratings in both novice and expert groups are particularly interesting because these lower ratings can be attributed to different potential causes. For the novices, the low ease of change (DV2) ratings can be attributed to a *lack* of category representations for Galant schemata. Contrastingly, the expert groups' lower ease of change ratings can be attributed to narrow activation of highly elaborated simulators in memory, resulting in a rigidity of interpretation. For the intermediate group, there are enough Galant schemata representations to select from, but the sparser (less elaborated) nature of these traces means that alternating simulator bases is much easier.

The interaction of expertise and attended feature for DV1 indicates that for the novice group only, scale degree hearing for the soprano line (compared to the bass line) was very low, whereas for both intermediate and expert groups, this difference was nonsignificant. This result is particularly interesting given that the novice group reported relatively high levels of scale degree hearing ability (see Figure 5.25 above). I interpret this result in light of schemata expertise: as both intermediate and expert groups had higher familiarity with Galant schemata, I argue that they were likely able to automatically 'fill in' the misaligned soprano line information in imagery (LTM to WM), as soprano line simulators were more likely to automatically come online during listening (see Figure 5.33). Novices have not yet acquired the tight association between bass and soprano line simulators and thus they need to rely solely on the perceptual input—or representational activation alone—of existing scale degree lines, which results in a lower ability in interpretation for this voice.

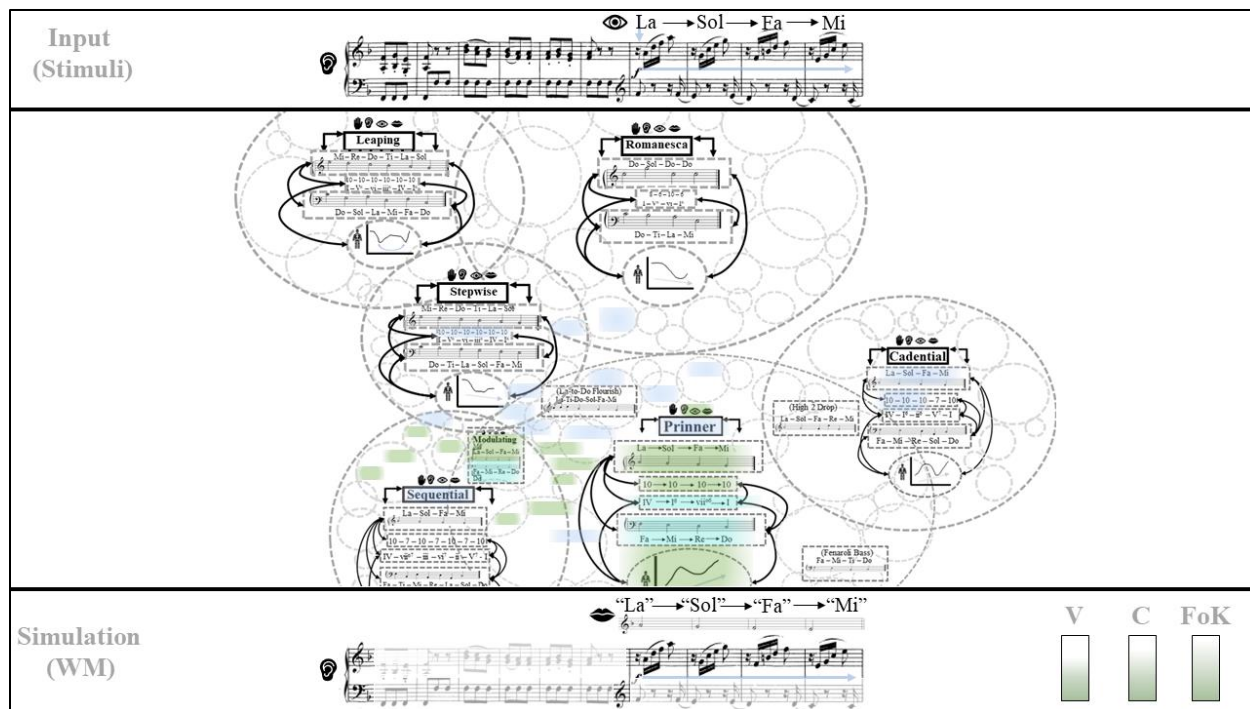


Figure 5.33. Soprano Voice Imagery During Simulation for Intermediate and Expert Groups

Lastly, the interaction of expertise group and modulation type for DV1 shows that for the expert group only, the Romanesca hearing is more difficult compared to a Prinner hearing. Contrastingly, and as predicted, Prinner and Romanesca categories are much more equally available for the intermediate group. As there is no significant three-way interaction between expertise group, feature, and modulation type, this indicates that the availability of both Prinner and Romanesca interpretations for the Expert group did not differ between attending types (see Figure 5.34). Therefore, even when only attending to the bass line, the experts' Galant schema knowledge (i.e., automatic activation of simulator pools through associational and referential processing) exerts an influence on the simulators that are available from LTM to WM. In this way, hearing a later modulation (or no modulation at all, i.e., Romanesca or dominant Prinner) is far less compelling than hearing an earlier modulation (i.e., modulating Prinner), regardless of whether the experts use a focused (i.e., bass line) or diffuse (i.e., schema) attending strategy.

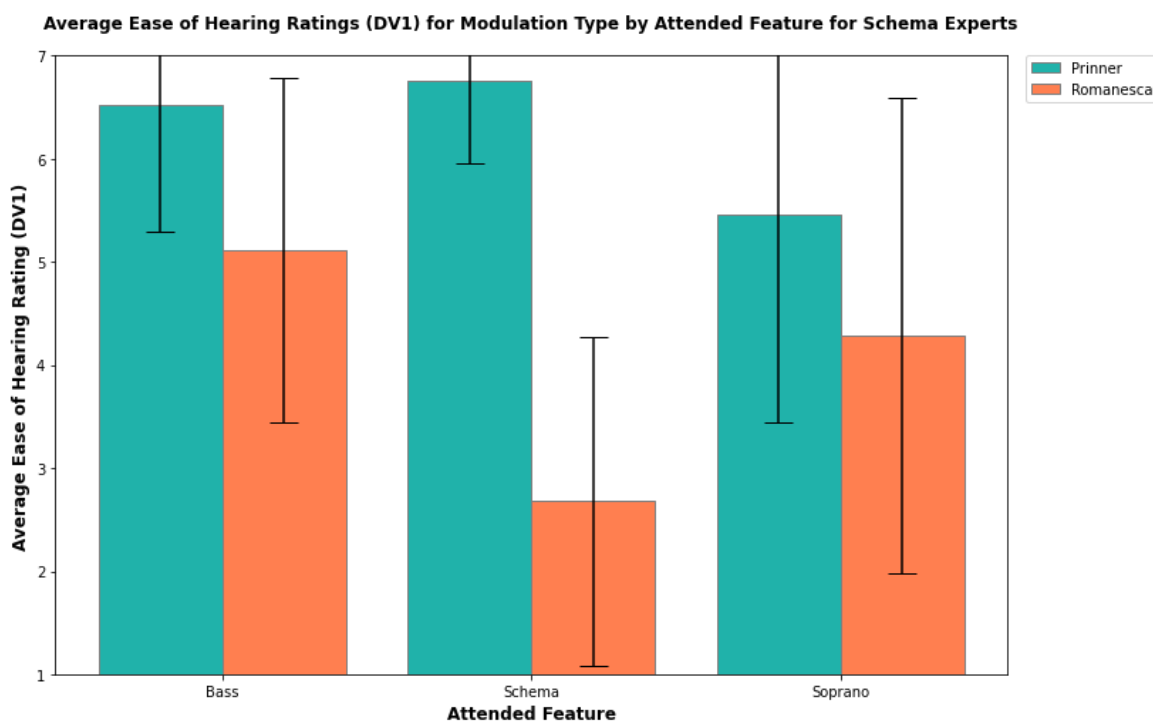


Figure 5.34. Average Ease of Hearing Ratings (DV1) for Modulation Type by Attended Feature for Schemata Experts

As predicted, it appears that this is due to both bottom-up and top-down factors. One of the experts' responses<sup>78</sup> for how they formed their interpretations suggests exactly this:

In addition, on a more reflective level, of course there are cues for the activation of the Modulating Prinner schema in a spot like this (e.g., surface activity that's speedy, harmonic modulation, connecting to Converging Cadence pattern at the end)—the top-down cue of "this is 3-over-1 becoming immediately 6-over-4 in the new key" is probably the strongest factor of all, which doesn't really leave much room for

<sup>78</sup> This participant was in the Mid-level excerpt familiarity group.

ambiguity. In terms of the formal grammar and even the surface affordances, there's simply no way that this is a (home-key) Prinner riposte.'

Therefore, it very much appears as if the 'surface cues' (bottom-up) facilitated direct, representational activation of traces pertaining to the modulating Prinner. Even when presented in a limited context, these cues were strong enough to elicit explicit recognition of the formal location of the excerpt through associational and referential processing. Along with the limited amount of distributional (order-over-time) information available in the short excerpt (top-down), this severely limited the categorization options for the schema here, resulting clearly in a reduction of ease of hearing ratings for Romanesca or late-modulation Prinner, and lower ease of change ratings overall for expert listeners.

Lastly, an examination of the interpretation formation strategies revealed important roles for imagery and subvocalization involving both musical and verbal content. Table 5.3 provides a breakdown of the strategies reported by participants for each expertise group (expert, intermediate, novice), and for all participants together (sum). Most categories were equally represented across groups. The most common strategies were imagining or subvocalizing scale degrees or solfège during listening (n = 19 responses). Some participants reported that they sang or spoke scale degrees or solfège aloud during listening (n = 6 responses) or sung or verbalized out loud without scale degrees or solfège while listening (n = 4 responses). Some participants reported speaking or singing between listenings (n = 4), and very few reported playing an instrument (n = 2). These responses suggest that imagery and verbalization play vital roles in simulation during interpretation formation.

Table 5.3. Interpretation Formation Types by Expertise Group

<b>Interpretation Formation Type</b>	<b>Expert</b>	<b>Intermediate</b>	<b>Novice</b>	<b>Sum</b>
I have no idea, I could just hear it!	1	1	2	4
I imagined scale degrees or solfège as I listened	3	6	4	13
I imagined or sung/spoke the interpretation between listenings	1	1	2	4
I sang or spoke (without scale degrees or solfège)	1	1	2	4
I subvocalized (silent rehearsal, spoken or sung) scale degrees or solfège while listening	2	4	2	8
I spoke or sang scale degrees or solfège while listening	1	4	1	6
I played an instrument	2	1	0	3
Other (please specify)	1	0	0	1

### Summary and Conclusions

In this chapter, I have presented a case for music theory LTWM as control over interpretation formation and modification in perception. Previous research into ambiguous figure perception has revealed important contributions of both top-down and bottom-up factors. The analytical case of Mozart's piano sonatas demonstrates a potential for perceptual ambiguity between Modulating Primmers and Step-Descent Romanescas in Sonata Form transitions due to overlapping features. The qualitative survey results provide support for this claim by demonstrating that the transition of K. 280, iii, is indeed perceptually bistable. However, this study also reveals effects of top-down factors on the ease of forming and changing an interpretation: in particular, the effects of expertise demonstrate that the extent to which interpretations are available and can be modified depends partially on the level of experience with Galant Schemata categories.

## Chapter 6

### A Project Recapitulation and Postscript

Throughout the course of this dissertation, I have created an account of the iterative loop between thinking and listening in music theory expertise, using Galant schemata as a case study. In the first chapter, I outlined the background, scope, and methodology for the project. I demonstrated that the thinking-listening dichotomy is positioned as one separating verbal (out-of-time) and imagery (experiential, in-time) processes. I outlined the conditions that scholars claim are required for synthesis between these two activities (e.g., the ordering of activities), as well as circumstances in which the connection between thinking and listening is disassociated rather than synthesized. One such account offered by Hansberry suggested that only complex music theories which use both phenomenal and theoretical concepts afford a blend between thinking and listening. I outlined current explanatory approaches in music cognition (e.g., Zbikowski 2002) for various aspects of music theoretic expertise, all of which emphasized the importance of category representation and online categorization for this expertise. Lastly, I highlighted four concrete effects of music theory expertise as discussed by scholars (enhanced perception, benefits of language, biased perception, intuition), which I will recontextualize in the section below using the framework developed throughout the dissertation. Overall, this chapter highlighted the importance of developing a framework capable of capturing imagery-verbalization interactions, multimodal representation, intuition, and flexibility in category representation and online category decisions.

In the second chapter, I synthesized Paivio's Dual-Coding theory and Barsalou's Dynamic Interpretation in Perceptual Symbols System in order to account for multimodal

category representation. The framework embraces a structurally unitary but functionally modular approach to mental representation. Therefore, the traditional modular memory types—semantic, declarative, episodic—arise from selective access to the same pool of representations, not from structurally distinct systems. The modular distinction between verbal and nonverbal systems in DCT is one that arises from specialization within subsections of overlapping brain regions (e.g., auditory cortex). As such, this framework also embraces an exemplar approach to category representation in which all encountered category exemplars are stored in episodic traces. Generalized category knowledge arises from co-activation of traces across many exemplars. This framework therefore accounts for both memory *organization* in category representation and memory *access* in online categorization tasks. Categories are organized in the mind as pools of simulators—interconnected imagens and logogens stored across the different modalities—each of which arises from particular interactions with the category at hand. Category representations—imagens and logogens, and the probabilistic connections in between them—will differ as a result of experience. Perceptual-cognitive processes such as recognition, identification, and categorization are operationalized as differences in DCT processing types (representational, associational, referential) over time, reflecting increased availability of information during categorization processes (or simulation). This framework also accounts for introspective judgements through several processes available in WM: imagery (vividness, control), metamemory (sensitivity to memory processing and functions, specifically feeling-of-knowing), and interoceptive representation in emotional construction.

In the third chapter, I used the framework developed in the second chapter to create an embodied account of Galant schema representation. Here I distinguished the differences in representation of these categories between encultured listeners and modern Galant experts as one



that stems from systematic differences in the interactions (and therefore the memory of) schema categories. I argued that encultured listeners possess more loose and holistic representations, in which Galant schemata categories are represented primarily by centralized auditory and interoceptive representations stemming from interactions involving listening activities.

Contrastingly, I argued that the modern experts' Galant schema representations are distributed and structured—i.e., varied across modality and system, and highly probabilistic—stemming from more varied interactions with these categories. I also argued that music theoretic concepts are used to explicitly store properties and relations for these categories, which function to direct attention to particular features during category interactions, ensuring that they are stored and connected with other category information in memory. I then demonstrated that this type of representation allows the expert more flexibility in simulation abilities (i.e., categorization and information manipulation).

In the fourth chapter, I extended and combined the framework with a memory expertise perspective (LTWM), proposing that the loop between thinking and listening discussed by theorists is a process for acquiring memory expertise. I discussed how Galant schemata acquisition, both in traditional conservatory training and in modern-day learning, represents a form of LTWM acquisition. To support this, I demonstrated that the three pillars of traditional conservatory training—*solfeggio*, *partimenti*, *counterpoint*—shift progressively from encoding to retrieval activities, placing increasingly more demand on efficient memory access. In the context of the modern theorist, I demonstrated that learning Galant schemata involves realigning property and relation simulators in memory to better reflect the cooccurrences and probabilities of schema categories. This is done through two primary activities. First, the studying of individual schemata and their prototypical presentations in isolation, which helps to facilitate

memory encoding and establish a baseline of systematic interaction with the categories. Second, studying schema in the context of analysis, which involves using both verbal and nonverbal interactions to ensure that simulators come online at the right time and in the right order (reflecting Galant schemata LTWM). To elucidate this, I created a hypothetical account of the steps and processes undertaken by a modern theorist attempting to learn the analysis of the first movement of Mozart's K. 545 outlined in *Music in the Galant Style*.

In the fifth chapter, I provided a concrete demonstration of music theoretic LTWM in action through an analytical case study and experimental verification of musical ambiguous figure perception using Galant schemata. In the analytical case study, I showed that there is substantial overlap between the modulating Prinner and step-descent Romanesca (fauxbourdon variant) in transitional spaces within Mozart piano sonatas. Along with some form-functional overlap, I showed that these schemata have substantial overlapping features, which I argued makes a case for potential perceptual ambiguity. I then provided support for this claim experimentally. I suggested, however, that schema expertise should affect ability to both form *and* modify interpretations during listening. I proposed that this is due to expert schema representation, in which property and relation simulators are more highly elaborated and more tightly structured (i.e., probabilistic), which effectively narrows representational activation of simulator pools in memory to those more like the exemplar being heard. Difficulties in changing interpretation stem from the high-probability associations of simulators, which makes effortful change of simulator pool (and suppression of those already active) much more difficult. These hypotheses were supported, as intermediate schema expertise showed greater ease in both forming and changing schema interpretations, whereas schema experts had both lower ease of hearing ratings for the Romanesca categories and lower ease of change ratings between

modulating Prinner and Romanesca schema. Taken together, the current framework allows for a conceptualization of category representation and category simulation powerful enough to explain differences between individuals and groups in an experimental context.

### Operationalizing Claims from Chapter 1

Given the work completed in chapters 2 through 5, it is now possible to return to and recontextualize claims made in chapter 1. To recapitulate, I outlined five primary claims that were addressed in the previous chapters:

1. The iterative growth of knowledge proposed by the ‘loop’ between thinking and listening stems from the co-operative independence between verbal and nonverbal systems.

I demonstrated this through the cognitive framework developed in chapter 2, and through the interactions of verbalization and imagery demonstrated in Galant schema category learning in chapters 3 and 4.

2. Music theory concepts are central to music theory expertise, including expertise in Galant Schemata. Music theory concepts function as cognitive tools which facilitate interactions with schema categories—in a sense, they function both as a means to direct attention, and as a kind of ‘container’ in memory for representational information.
3. Music theoretic concepts, including Galant schemata, are constructed out of different types of representations stemming from the various ways in which theorists interact with these categories: listening, score analysis, verbalization, singing, writing, piano playing, etc. The representational make up of concepts will differ based on these interactions, and they will therefore serve different functions in expertise.

These claims were primarily addressed in chapter 3 where I argued that music theoretic concepts are central in developing structured distributed representations for Galant schemata. They ensure that the feature of concern is attended to during perception, elaborated and encoded in memory. I showed that the varied ways in which music theorists interact with Galant schema categories create multimodal traces for these categories. In chapter 4, I showed that traces stemming from different types of activities—like solmization, singing, score study—are used to support different aspects of Galant schema categorization.

4. Schema representation is radically different between those explicitly trained in schemata categories (music theorists) and those whose interaction with the categories is from ‘natural’ exposure (i.e., statistical learning) alone. This is due to the distributed type of representations (verbal, nonverbal, multimodal) that experts have as a result of their more diverse interactions with schemata categories outside the realm of listening alone.

This claim was primarily addressed in chapters 3 and 4. In chapter 3, I showed that encultured listeners possess loose holistic simulators for Galant categories, which are more centrally represented in auditory and interoceptive modalities. Experts’ schema representations by contrast, are structured and distributed meaning that they are more distributed across system and mode, and that these representations are more probabilistically linked. In chapter 4, I demonstrated a modern theorists’ use of music theoretic concepts to re-align their simulator pools to more accurately reflect Galant schemata regularities.

5. The effect of expertise is ‘real’ and observable. Specifically, the acquisition of ‘eighteenth-century hearing’ and its impact on cognition should be observable in experimental contexts. The learning of Galant schemata therefore does concretely affect one’s hearing through memory priming, imagery, and online categorization processes.

This final claim was primarily addressed in chapter 5 where I showed that ease of hearing and ease of change ratings for Prinner and Romanesca schema differed on the basis of expertise.

Those with intermediate schema expertise were more easily able to hear Prinner and Romanesca interpretations, and more easily able to alternate between them. Schema experts, however, rated the Prinner schema as much easier to hear in these contexts and had lower ease of rating change ratings. This demonstrates that given experts more automatic and elaborated schema knowledge in which Prinner schema are more common in modulating transition contexts, they could not easily perceive a Romanesca, nor alternate between Prinner and Romanesca interpretations.

I will next seek to recontextualize the claims regarding the concrete effects of music theoretic expertise (claim five) within the context of Galant schema theory. I noted in the first chapter that scholars discuss four different concrete effects of music theory expertise: enhanced or modified perception, benefits of language, biased perception, and music theoretic intuition. I will now discuss each in turn.

*Enhanced or modified perception.* Enhanced or modified perception can be understood to stem from both the increased elaboration in simulator pools in LTWM, and in greater ease in accessing and manipulating these traces in WM during simulation. As more property and relation simulators are acquired, representational processing for these features becomes facilitated, as does associational processing for related or cooccurring properties and relations. With greater ease of representational activation, memory traces that store schema interactions come online faster and with less effort, affording a pop-out effect for these categories that can be described as a type of enhanced perception. Claims about the new type of perception captured by the terms ‘hearing-as’ and ‘audiation’ (which described a synthesis of verbal and nonverbal understanding) can be understood as increased facilitation in both associational and referential processing. As

the number of simulator pools increases in memory, and the probabilistic association between verbal and nonverbal representations increases, referential and associational processing becomes facilitated and more automatic, reflecting increased information availability during listening. Because logogens may become automatically primed or activated during listening, ‘perception’ becomes enhanced with conscious or explicit understanding.

*Language benefits the precision of thought.* Within my current framework, language functions to guide interactions with schema categories, and to make connections in memory within the verbal and nonverbal systems. Language is a means of ensuring that particular features and relations are encoded into LTM, and that they are retrievable. In this way, when verbalization is used consistently with particular nonverbal category features, simulator pools are created in which verbal and nonverbal representations are highly associated. Such directed attention towards particular types of interactions with categories facilitates property and relation simulator encoding as well as access to these traces in future categorization decisions. As an example, I showed that the encoding of solmization helps to encode melodic lines into memory, which can be recalled later during simulation. Recalling solmization can bootstrap categorization, particularly when a category decision is proving difficult to form, as cueing solfège allows the perceiver to activate relevant nonverbal schemata traces through referential processing. This allows attention to be more deliberately focused during perception and simulation, allowing for a precision of thought engendered by language.

*Biased perception and ‘inattentional deafness.’* Biased perception, particularly within the Galant schemata case study examined here, stems from the probabilistic connections between simulators and simulator pools. Essentially, as features and relations are encoded and associated in memory, they become facilitated during processing. Because attention has repeatedly been

allocated to particular properties and relations during category interactions, those interactions are highly elaborated in memory and come online more automatically. Other features, or in the case of ambiguous schema perception, other schema options, may *not* be available to experts as the overlearned nature of experts' categories are narrowly defined. As expertise increases, memory becomes more elaborated and high-probability, making it much harder to suppress active representations and actively work to bring online other simulator pools for a different schema choice.

*Intuition in music theory as automation and introspective availability.* As highlighted above, music theoretic intuition, understood as automation, can be understood as facilitated processing and increased availability of a more diverse set of simulator pools. One example of such facilitated or automated processing is the ability to recognize schema in degraded or unfamiliar conditions, much like the schema categorization from parallel lines as discussed in chapter 3. However, increased introspective availability in music theoretic intuition can also be understood not only as increased facilitation of verbal and nonverbal processing, but as increased sensitivity in metamemory and in the attachment of interoceptive representations to different types of category judgements, in a similar vein to emotional constructions. For example, a schema may be identified through increased facilitation of certain features, such as the melodic line La-Sol-Fa-Mi—but it may also be facilitated by the development of a sensitivity to *mental habits* or a pattern of representational priming. Therefore, sensitivity to simulator pools being primed but not yet activated in particular orders and particular configurations may indeed produce enough of an effect for perform an accurate category judgement; experts can make category judgements based on memory priming as they have sensitivity to patterns of activations of representations, even if these representations are not actively in use. Similarly, as this

sensitivity in metamemory increases, the types of interoceptive representations associated with these states increase and become more specific. This allows for increased reliance on introspective states and reflection during automated category decisions during listening, as theorists can rely on a wide range of introspectively available states to make a category judgement and need not solely rely on assessment of category features alone.

## Future Research

Here I consider two paths for future research. Firstly, an expansion of the current framework to examine other theories, particularly in examining concrete differences in the cognitive affordances offered by different analytical systems. Secondly, I discuss possible avenues for future experimental research.

The framework laid out here offers a potential means to hypothesize about the representational make-up of concepts in other music theoretical systems, and to be able to compare and contrast systems with greater depth. A music theoretic system can be understood to be a sub-specialization within music theory, essentially a different type of memory expertise (much like solfeggio, partimenti, etc.). Each system, for example Schenkerian analysis or set theory, has a different set of repertoires and a different set of concepts or cognitive tools, affording a varied set of interactions. Therefore, each system should foster a different set of affordances for the effects of training in these sub-disciplines. Combined with qualitative methods examining real analytical practices used by theorists, it would be possible to adapt the framework presented here to map out memory expertise in each of these subdomains. This could then be used to more accurately elucidate the practical, conceptual and epistemological differences between analytical systems. For example, much work has been done comparing the



score reductive approaches of Galant schema theory and Schenkerian analysis (Rabinovich 2013; Schwab-Felisch 2014). The current framework could be used to compare and contrast the similarities and differences between these two traditions with more precision. For example, while prototype mapping (i.e., locating a prototype using reduction) appears to be important for each tradition, the types of auditory imagens acquired for each system differ substantially. Where reduction may provide a similar set of visual-verbal referential connections (prototype differences notwithstanding), the types of auditory and interoceptive representations and their associations with visual and verbal units differs substantially. In a Galant approach, the units are much shorter, creating tight, highly associative packets where visual, verbal, haptic, and auditory representations for a given property or relation simulator (e.g., La-Sol-Fa-Mi) can be interchanged for one another because their interactions occur on a similar time scale. Contrastingly, auditory and visual-verbal interactions in a Schenkerian approach are highly divergent, where chunks are not the same size and therefore likely cannot be substituted for one another as they can be in Galant schema theory. This means that to *experience* a Schenkerian reduction is not necessarily solely an auditory phenomenon, but likely an act of reconstruction of many different experiential factors. One sees Schenkerian-inspired analysts employing a range of metaphors and analogies (e.g., Larson's forces) to connect an analysis to 'listening' with its many different embodied responses. Whereas Galant categories can be acquired through statistical learning, suggesting that coactivation of verbal, visual and auditory representations is likely quite automatic, Schenkerian reductions do not afford this coactivation. Therefore, more effect (and therefore more expertise, time-on-task) is required to bring a Schenkerian reduction 'to life.'

There are also several avenues for future experimental research. First, one avenue would be to expand and examine the conceptual peg hypothesis in a musical context. This could be done in several ways. Firstly, much like research examining the concreteness of vernacular language, it could be possible to gather several subjective measurements, like concreteness or imagery vividness. Given the framework here, music theoretic concepts classified as property simulators (e.g., scale degrees) should be rated as most concrete, while relation simulators (e.g., recapitulation) should be rated as less concrete. A second avenue following this survey would be to test the effects of concreteness on musical memory in an experimental context. Pairs of music theoretic words could be paired together, with those pairs including concrete concepts better remembered than those with more abstract words. A similar experiment could also pair concrete and abstract concepts with musical excerpts, and memory for those excerpts could be tested. Memory for those excerpts paired with concrete music theoretic concepts would likely be facilitated more than those paired with abstract concepts. Finally, a series of experiments examining supervised learning could be implemented to examine the effects of directed attention on Galant schema category learning. The framework developed in this dissertation argues that memory for Galant schema is facilitated by explicit encoding with music theoretic concepts. Directing attention to primary concrete features like scale degree lines (outer voices, one at a time), should help those unfamiliar with Galant schema learn these categories, and reduce confusion between perceptually overlapping categories (e.g., Prinner / Romanesca).

## Works Cited

- Abbot-Smith, Kirsten, and Michael Tomasello. 2006. "Exemplar-Learning and Schematization in a Usage-Based Account of Syntactic Acquisition." *The Linguistic Review* 23 (3): 275–90.  
<https://doi.org/10.1515/TLR.2006.011>.
- Agawu, Kofi. 1994. "Ambiguity in Tonal Music: A Preliminary Study." In *Theory, Analysis, and Meaning in Music*, edited by Anthony Pople, 86–107. Cambridge University Press.
- Aldwell, Edward, and Carl Schachter. 2002. *Harmony and Voice Leading*. 3rd ed. Australia ; United States: Cengage Learning.
- Altarriba, Jeanette, and Lisa M. Bauer. 2004. "The Distinctiveness of Emotion Concepts: A Comparison between Emotion, Abstract, and Concrete Words." *The American Journal of Psychology* 117 (3): 389–410. <https://doi.org/10.2307/4149007>.
- Altarriba, Jeanette, Lisa M. Bauer, and Claudia Benvenuto. 1999. "Concreteness, Context Availability, and Imageability Ratings and Word Associations for Abstract, Concrete, and Emotion Words." *Behavior Research Methods, Instruments, & Computers* 31 (4): 578–602.  
<https://doi.org/10.3758/BF03200738>.
- Anderson, Richard, C., and David Pearson P. 1984. "A Schema-Theoretic View of Basic Processes in Reading Comprehension." In *Handbook of Reading Research*, edited by David Pearson P., 255–91. New York: Longman.: University of Illinois at Urbana-Champaign.
- Arndt, Matthew. 2011. "Schenker and Schoenberg on the Will of the Tone." *Journal of Music Theory* 55 (1): 89–146.

- Ashby, F. G., S. Queller, and P. M. Berretty. 1999. "On the Dominance of Unidimensional Rules in Unsupervised Categorization." *Perception & Psychophysics* 61 (6): 1178–99.  
<https://doi.org/10.3758/bf03207622>.
- Ashby, F. Gregory, and W. Todd Maddox. 2005. "Human Category Learning." *Annual Review of Psychology* 56 (1): 149–78. <https://doi.org/10.1146/annurev.psych.56.091103.070217>.
- Ashby, F. Gregory, and Vivian Valentin V. 2017. "Multiple Systems of Perceptual Category Learning: Theory and Cognitive Tests." In *Handbook of Categorization in Cognitive Science*, edited by Henri Cohen and Claire Lefebvre, Second edition, 157–88. Amsterdam: Elsevier.
- Ashley, Richard. 2020. "On Prototypes and the Prototypical: An Investigation of Music-Theoretic Concepts." In *Society for Music Theory 34th Annual Meeting*. Virtual Conference.
- Baddeley, Alan. 2000. "The Episodic Buffer: A New Component of Working Memory?" *Trends in Cognitive Sciences* 4 (11): 417–23. [https://doi.org/10.1016/S1364-6613\(00\)01538-2](https://doi.org/10.1016/S1364-6613(00)01538-2).
- . 2003. "Working Memory: Looking Back and Looking Forward." *Nature Reviews Neuroscience* 4 (10): 829–39. <https://doi.org/10.1038/nrn1201>.
- Baddeley, Alan D. 2007. *Working Memory, Thought, and Action*. Oxford Psychology Series 45. Oxford; New York: Oxford University Press.
- Baddeley, Alan D., and Graham Hitch. 1974. "Working Memory." In *Psychology of Learning and Motivation*, edited by Gordon H. Bower, 8:47–89. Academic Press.  
[https://doi.org/10.1016/S0079-7421\(08\)60452-1](https://doi.org/10.1016/S0079-7421(08)60452-1).
- Baragwanath, Nicholas. 2020. *The Solfeggio Tradition: A Forgotten Art of Melody in the Long Eighteenth Century*. New York: Oxford University Press.
- Barrett, Lisa Feldman. 2014. "The Conceptual Act Theory: A Précis." *Emotion Review* 6 (4): 292–97.  
<https://doi.org/10.1177/1754073914534479>.

- . 2017a. *How Emotions Are Made: The Secret Life of the Brain*. Boston: Houghton Mifflin Harcourt.
- . 2017b. “The Theory of Constructed Emotion: An Active Inference Account of Interoception and Categorization.” *Social Cognitive and Affective Neuroscience* 12 (1): 1–23.  
<https://doi.org/10.1093/scan/nsw154>.
- Barrett, Lisa Feldman, Christine D. Wilson-Mendenhall, and Lawrence W. Barsalou. 2015. “The Conceptual Act Theory: A Road Map.” In *The Psychological Construction of Emotion*, edited by Lisa Feldman Barrett and James Russell, 83–110. New York, NY: The Guilford Press.
- Barsalou, Lawrence W. 1983. “Ad Hoc Categories.” *Memory & Cognition* 11 (3): 211–27.  
<https://doi.org/10.3758/BF03196968>.
- . 1990. “On the Indistinguishability of Exemplar Memory and Abstraction in Category Representation.” In *Advances in Social Cognition: Content and Process Specificity in the Effects of Prior Experiences*, edited by Thomas Srull K and Robert Wyer, Jr S, III:61–88. Hillsdale, N.J: Lawrence Erlbaum Associates.
- . 1999. “Perceptual Symbol Systems.” *Behavioral and Brain Sciences* 22 (4): 577–660.  
<https://doi.org/10.1017/S0140525X99002149>.
- . 2002. “Being There Conceptually: Simulating Categories in Preparation for Situated Action.” In *Representation, Memory, and Development: Essays in Honor of Jean Mandler*, edited by Nancy L. Stein, Patricia J. Bauer, Mitchell Rabinowitz, and George Mandler, 1–16. Psychology Press.
- . 2003a. “Abstraction in Perceptual Symbol Systems.” *Building Object Categories in Developmental Time* 358: 1177–87.

- . 2003b. “Situated Simulation in the Human Conceptual System.” *Language and Cognitive Processes* 18 (5–6): 513–62. <https://doi.org/10.1080/01690960344000026>.
- . 2005a. “Abstraction as Dynamic Interpretation in Perceptual Symbol Systems.” In *Building Object Categories in Developmental Time*, edited by Lisa Gershkoff-Stowe and David H. Rakison, 389–431. Mahwah, N.J: Psychology Press.
- . 2005b. “Continuity of the Conceptual System across Species.” *Trends in Cognitive Sciences* 9 (7): 309–11. <https://doi.org/10.1016/j.tics.2005.05.003>.
- . 2009. “Simulation, Situated Conceptualization, and Prediction.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 364 (1521): 1281. <https://doi.org/10.1098/rstb.2008.0319>.
- Barsalou, Lawrence W., Ava Santos, W. Kyle Simmons, and Christine D. Wilson. 2007. “Language and Simulation in Conceptual Processing.” In *Symbols, Embodiment, and Meaning*, edited by Manuel de Vega, Arthur M. Glenberg, and Arthur Graesser, 245–83. Oxford: Oxford University Press.
- Barsalou, Lawrence W., and Katja Wiemer-Hastings. 2005. “Situating Abstract Concepts.” In *Grounding Cognition*, edited by Diane Pecher and Rolf A. Zwaan, 1<sup>st</sup> ed., 129–63. Cambridge University Press. <https://doi.org/10.1017/CBO9780511499968.007>.
- Batiste, Édouard. 1865. *Solfèges du Conservatoire*. Paris: Heugel.
- Bazin, Francois. 1857. *Cours D’Harmonie Theorique Et Pratique*. Edited by Léon Escudier. 2nd ed. Kessinger Publishing, LLC.
- Best, Catherine A. 2020. “The Effect of Labels on Visual Attention: An Eye Tracking Study.” *Food Quality and Preference* 84: 7. <https://doi.org/10.1016/j.foodqual.2020.103948>.

- Best, Catherine A., Hyungwook Yim, and Vladimir M. Sloutsky. 2013. "The Cost of Selective Attention in Category Learning: Developmental Differences between Adults and Infants." *Journal of Experimental Child Psychology* 116 (2): 105–19. <https://doi.org/10.1016/j.jecp.2013.05.002>.
- Betz, Nicole, Katie Hoemann, and Lisa Feldman Barrett. 2019. "Words Are a Context for Mental Inference." *Emotion* 19 (8): 1463–77. <https://doi.org/10.1037/emo0000510>.
- Bezuidenhout, Kristian. 2016. *Mozart: Keyboard Music*. CD. Vol. 8 & 9. Harmonia Mundi.
- Blair, Mark R., Marcus R. Watson, R. Calen Walshe, and Phillip Maj. 2009. "Extremely Selective Attention: Eye-Tracking Studies of the Dynamic Allocation of Attention to Stimulus Features in Categorization." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 35 (5): 1196–1206. <https://doi.org/10.1037/a0016272>.
- Bourne, Janet Eileen. 2015. "A Theory of Analogy for Musical Sense-Making and Categorization: Understanding Musical Jabberwocky." Ph.D., Illinois: Northwestern University. <http://search.proquest.com/docview/1720843383/abstract/6B006307C1844A8DPQ/1>.
- Bowers, Jeffrey S., and Keely W. Jones. 2008. "Detecting Objects Is Easier than Categorizing Them." *The Quarterly Journal of Experimental Psychology* 61 (4): 552–57. <https://doi.org/10.1080/17470210701798290>.
- Bregman, Albert S. 1990. *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, Massachusetts: A Bradford Book.
- Bröker, Franziska, Bradley C. Love, and Peter Dayan. 2021. "When Unsupervised Training Benefits Category Learning." *Cognition*. <https://doi.org/10.1016/j.cognition.2021.104984>.
- Brown, Roger, and David McNeill. 1966. "The 'Tip of the Tongue' Phenomenon." *Journal of Verbal Learning and Verbal Behavior* 5 (4): 325–37. [https://doi.org/10.1016/S0022-5371\(66\)80040-3](https://doi.org/10.1016/S0022-5371(66)80040-3).

- Brugger, Peter. 1999. "One Hundred Years of an Ambiguous Figure: Happy Birthday, Duck/Rabbit!" *Perceptual and Motor Skills* 89 (3): 973–77. <https://doi.org/10.2466/pms.1999.89.3.973>.
- Byros, Vasileios. 2011. "Quo Vadis Corpus?" In *Society for Music Theory 34th Annual Meeting*. Minneapolis, MN.
- Byros, Vasili. 2009a. "Foundations of Tonality as Situated Cognition, 1730-1830: An Enquiry into the Culture and Cognition of Eighteenth -Century Tonality with Beethoven's 'Eroica' Symphony as a Case Study." ProQuest Dissertations Publishing. <http://search.proquest.com/docview/305040356/?pq-origsite=primo>.
- . 2009b. "Towards an 'Archaeology' of Hearing: Schemata and Eighteenth-Century Consciousness." *Musica Humana* 1: 72.
- . 2012a. "Meyer's Anvil: Revisiting the Schema Concept." *Music Analysis* 31 (3): 273–346. <https://doi.org/10.1111/j.1468-2249.2012.00344.x>.
- . 2012b. "Unearthing the Past: Theory and Archaeology in Robert Gjerdingen's 'Music in the Galant Style.'" *Music Analysis* 31 (1): 112–24.
- . 2015a. "'Haupttrühepunkte Des Geistes" Punctuation Schemas and the Late-Eighteenth-Century Sonata." In *What Is a Cadence?: Theoretical and Analytical Perspectives on Cadences in the Classical Repertoire*, edited by Markus Neuwirth and Pieter Bergé, 215–51. Leuven, Belgium: Leuven University Press.
- . 2015b. "Prelude on a Partimento: Invention in the Compositional Pedagogy of the German States in the Time of J. S. Bach." *Music Theory Online* 21 (3). <https://www.mtosmt.org/issues/mto.15.21.3/mto.15.21.3.byros.html>.
- Cañal-Bruland, Rouwen, John van der Kamp, and Rob Gray. 2016. "Acting Is Perceiving!" *The Behavioral and Brain Sciences* 39: E223. <https://doi.org/10.1017/S0140525X15002460>.



- Caplin, William E. 1998. *Classical Form: A Theory of Formal Functions for the Instrumental Music of Haydn, Mozart, and Beethoven*. Oxford: Oxford University Press, U.S.A.
- Caplin, William E. 2015. "Harmony and Cadence in Gjerdingen's 'Prinner.'" In *What Is a Cadence?: Theoretical and Analytical Perspectives on Cadences in the Classical Repertoire*, edited by Markus Neuwirth and Pieter Bergé, 17–58. Leuven, Belgium: Leuven University Press.
- Chandrasekaran, Bharath, Seth R. Koslov, and W. T. Maddox. 2014. "Toward a Dual-Learning Systems Model of Speech Category Learning." *Frontiers in Psychology* 5. <https://doi.org/10.3389/fpsyg.2014.00825>.
- Chandrasekaran, Bharath, Han-Gyol Yi, and W. Todd Maddox. 2014. "Dual-Learning Systems during Speech Category Learning." *Psychonomic Bulletin & Review* 21 (2): 488–95. <https://doi.org/10.3758/s13423-013-0501-5>.
- Chandrasekaran, Bharath, Han-Gyol Yi, Kirsten E. Smayda, and W. Todd Maddox. 2016. "Effect of Explicit Dimensional Instruction on Speech Category Learning." *Attention, Perception, & Psychophysics* 78 (2): 566–82. <https://doi.org/10.3758/s13414-015-0999-x>.
- Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.
- . 2002. *Syntactic Structures*. Berlin, New York: Mouton de Gruyter.
- Choron, Alexandre-Etienne. 1804. *Principes d'accompagnement des écoles d'Italie, extraits des meilleurs auteurs: Leo, Durante, Fenaroli, Sala, Azopardi, Sabbatini, le père Martini, et autres*. Paris: Imbault.
- Chua, Elizabeth F., and Eliza Bliss-Moreau. 2016. "Knowing Your Heart and Your Mind: The Relationships between Metamemory and Interoception." *Consciousness and Cognition* 45: 146–58. <https://doi.org/10.1016/j.concog.2016.08.015>.

- Church, Barbara A., Eduardo III Mercado, Matthew G. Wisniewski, and Estella H. Liu. 2013. "Temporal Dynamics in Auditory Perceptual Learning: Impact of Sequencing and Incidental Learning." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 39 (1): 270–76. <https://doi.org/10.1037/a0028647>.
- Clendinning, Jane Piper, and Elizabeth West Marvin. 2016. *The Musician's Guide to Theory and Analysis*. 3rd ed. United Kingdom: W. W. Norton & Company.
- Clercq, Trevor de, and David Temperley. 2011. "A Corpus Analysis of Rock Harmony." *Popular Music* 30 (1): 47–70. <https://doi.org/10.1017/S026114301000067X>.
- Colley, Ian D, Peter E Keller, and Andrea R Halpern. 2018. "Working Memory and Auditory Imagery Predict Sensorimotor Synchronisation with Expressively Timed Music." *Quarterly Journal of Experimental Psychology* 71 (8): 1781–96. <https://doi.org/10.1080/17470218.2017.1366531>.
- Cone, Edward T. 1977. "Three Ways of Reading a Detective Story or a Brahms Intermezzo." *The Georgia Review* 31 (3): 554–74.
- Connell, Louise, Dermot Lynott, and Briony Banks. 2018. "Interoception: The Forgotten Modality in Perceptual Grounding of Abstract and Concrete Concepts." *Philosophical Transactions of the Royal Society B: Biological Sciences* 373 (1752): 20170143. <https://doi.org/10.1098/rstb.2017.0143>.
- Cox, Arnie. 2011. "Embodying Music: Principles of the Mimetic Hypothesis." *Music Theory Online* 17 (2). <http://www.mtosmt.org/issues/mto.11.17.2/mto.11.17.2.cox.html>.
- . 2016. *Music & Embodied Cognition*. Bloomington, IN: Indiana University Press.
- Craig, A. D. 2015. *How Do You Feel?: An Interoceptive Moment with Your Neurobiological Self*. Princeton: Princeton University Press.

- De Souza, Jonathan, Adam Roy, and Andrew Goldman. 2020. "Classical Rondos and Sonatas as Stylistic Categories." *Music Perception* 37 (5): 373–91.  
<https://doi.org/10.1525/mp.2020.37.5.373>.
- DeBellis, Mark. 2005. "Conceptual And Nonconceptual Modes Of Music Perception." *Postgraduate Journal of Aesthetics* 2 (2): 17.
- Divjak, Dagmar, and Antti Arppe. 2013. "Extracting Prototypes from Exemplars What Can Corpus Data Tell Us about Concept Representation?" *Cognitive Linguistics* 24 (2): 221–74.  
<https://doi.org/10.1515/cog-2013-0008>.
- Dubiel, Joseph. 2017. "Music Analysis and Kinds of Hearing-As." *Music Theory and Analysis (MTA)* 4 (2): 233–42. <https://doi.org/10.11116/MTA.4.2.4>.
- Dubiel, Joseph, Marion A. Guck, and Bryan Parkhurst. 2017. "Hearing As Hearing-As." *Music Theory and Analysis (MTA)* 4 (2): 229–32. <https://doi.org/10.11116/MTA.4.2.3>.
- Dubois, Théodore. 1891. *Traité d'Harmonie Théorique et Pratique*. Paris: Heugel.
- Dunlosky, John, and Robert A. Bjork, eds. 2013. *Handbook of Metamemory and Memory*. New York: Psychology Press. <https://doi.org/10.4324/9780203805503>.
- Durand, Émile. 1892. *Traité d'accompagnement au piano*. Paris: Leduc.
- Durante, Francesco, and Robert O. Gjerdingen. n.d. "Partimenti Diminuti." *Monuments of Partimenti*. Accessed April 15, 2022. <https://partimenti.org/partimenti/collections/durante/index.html>.
- Ell, Shawn, F. Ashby, and Steven Hutchinson. 2012. "Unsupervised Category Learning with Integral-Dimension Stimuli." *Quarterly Journal of Experimental Psychology (2006)* 65: 1537–62.  
<https://doi.org/10.1080/17470218.2012.658821>.

- Erickson, Lucy C., and Erik D. Thiessen. 2015. "Statistical Learning of Language: Theory, Validity, and Predictions of a Statistical Learning Account of Language Acquisition." *Developmental Review* 37: 66–108. <https://doi.org/10.1016/j.dr.2015.05.002>.
- Ericsson, K. Anders. 2018. "Superior Working Memory in Experts." In *The Cambridge Handbook of Expertise and Expert Performance*, edited by K. Anders Ericsson, Robert Hoffman, Aaron Kozbelt, and A. Mark Williams, 2nd ed., 696–714. Cambridge, United Kingdom ; New York, NY, USA: Cambridge University Press.
- Ericsson, K. Anders, and Walter Kintsch. 1995. "Long-Term Working Memory." *Psychological Review* 102 (2): 211–45. <https://doi.org/10.1037/0033-295X.102.2.211>.
- Fenaroli, Fedele. 1780. *Partimenti Ossia Basso Numerato, Opera Completa Di Fedele Fenaroli*. Paris.  
<https://urresearch.rochester.edu/institutionalPublicationPublicView.action?institutionalItemId=28260>.
- Ferguson, Brock, and Sandra Waxman. 2017. "Linking Language and Categorization in Infancy." *Journal of Child Language* 44 (3): 527–52. <https://doi.org/10.1017/S0305000916000568>.
- Fiacconi, Chris M., Jane E. Kouptsova, and Stefan Köhler. 2017. "A Role for Visceral Feedback and Interoception in Feelings-of-Knowing." *Consciousness and Cognition* 53: 70–80.  
<https://doi.org/10.1016/j.concog.2017.06.001>.
- Firestone, Chaz, and Brian J. Scholl. 2014. "'Top-Down' Effects Where None Should Be Found: The El Greco Fallacy in Perception Research." *Psychological Science* 25 (1): 38–46.  
<https://doi.org/10.1177/0956797613485092>.

- . 2015a. “Can You Experience ‘Top-down’ Effects on Perception?: The Case of Race Categories and Perceived Lightness.” *Psychonomic Bulletin & Review* 22 (3): 694–700. <https://doi.org/10.3758/s13423-014-0711-5>.
- . 2015b. “Enhanced Visual Awareness for Morality and Pajamas? Perception vs. Memory in ‘Top-down’ Effects.” *Cognition* 136: 409–16. <https://doi.org/10.1016/j.cognition.2014.10.014>.
- . 2015c. “When Do Ratings Implicate Perception versus Judgment? The ‘Overgeneralization Test’ for Top-down Effects.” *Visual Cognition* 23 (9–10): 1217–26. <https://doi.org/10.1080/13506285.2016.1160171>.
- . 2016. “Cognition Does Not Affect Perception: Evaluating the Evidence for ‘Top-down’ Effects.” *Behavioral and Brain Sciences* 39. <https://doi.org/10.1017/S0140525X15000965>.
- Floridou, Georgia A., Victoria J. Williamson, Lauren Stewart, and Daniel Müllensiefen. 2015. “The Involuntary Musical Imagery Scale (IMIS).” *Psychomusicology: Music, Mind, and Brain* 25 (1): 28–36. <https://doi.org/10.1037/pmu0000067>.
- Forde, Emer M. E., and Glyn W. Humphreys. 1999. “Category Specific Recognition Impairments: A Review of Important Case Studies and Influential Theories.” *Aphasiology* 13 (3): 169–93. <https://doi.org/10.1080/026870399402172>.
- Forder, Lewis, and Gary Lupyan. 2019. “Hearing Words Changes Color Perception: Facilitation of Color Discrimination by Verbal and Visual Cues.” *Journal of Experimental Psychology: General* 148 (7): 1105–23. <https://doi.org/10.1037/xge0000560>.
- Forkmann, Thomas, Anne Scherer, Judith Meessen, Matthias Michal, Hartmut Schächinger, Claus Vögele, and André Schulz. 2016. “Making Sense of What You Sense: Disentangling Interoceptive Awareness, Sensibility and Accuracy.” *International Journal of Psychophysiology* 109: 71–80. <https://doi.org/10.1016/j.ijpsycho.2016.09.019>.

- Foster, Jonathan, K., and Marko Jelacic. 1999. "Chapter 1: Memory Structures, Procedures, and Processes." In *Memory: Systems, Process, or Function?*, edited by Jonathan Foster K. and Marko Jelacic, 1–10. Oxford: Oxford University Press.
- Fotiadis, Fotis A., and Athanassios Protopapas. 2014. "The Effect of Newly Trained Verbal and Nonverbal Labels for the Cues in Probabilistic Category Learning." *Memory & Cognition* 42 (1): 112–25. <https://doi.org/10.3758/s13421-013-0350-5>.
- Frost, Ram, Blair C. Armstrong, Noam Siegelman, and Morten H. Christiansen. 2015. "Domain Generality versus Modality Specificity: The Paradox of Statistical Learning." *Trends in Cognitive Sciences* 19 (3): 117–25. <https://doi.org/10.1016/j.tics.2014.12.010>.
- Gahl, Susanne, and Alan C. L. Yu. 2006. "Introduction to the Special Issue on Exemplar-Based Models in Linguistics" 23 (3): 213–16. <https://doi.org/10.1515/TLR.2006.007>.
- Garfinkel, Sarah N., Adam B. Barrett, Ludovico Minati, Raymond J. Dolan, Anil K. Seth, and Hugo D. Critchley. 2016. "What the Heart Forgets: Cardiac Timing Influences Memory for Words and Is Modulated by Metacognition and Interoceptive Sensitivity." *Psychophysiology* 50 (6): 505–12. <https://doi.org/10.1111/psyp.12039>.
- Garfinkel, Sarah N., Anil K. Seth, Adam B. Barrett, Keisuke Suzuki, and Hugo D. Critchley. 2015. "Knowing Your Own Heart: Distinguishing Interoceptive Accuracy from Interoceptive Awareness." *Biological Psychology* 104: 65–74. <https://doi.org/10.1016/j.biopsycho.2014.11.004>.
- Garner, Wendell R. 1974. *The Processing of Information and Structure*. The Experimental Psychology Series. Potomac, Md.: Lawrence Erlbaum.
- Gedalgé, André. 1901. *Traité de la fugue*. Paris: Enoch.

- Gelding, Rebecca W., William Forde Thompson, and Blake W. Johnson. 2015. "The Pitch Imagery Arrow Task: Effects of Musical Training, Vividness, and Mental Control." *PLoS ONE* 10 (3): e0121809. <https://doi.org/10.1371/journal.pone.0121809>.
- Gentner, Dedre. 2016. "Language as Cognitive Tool Kit: How Language Supports Relational Thought." *American Psychologist* 71 (8): 650–57. <https://doi.org/10.1037/amp0000082>.
- Gentner, Dedre, and Jennifer Asmuth. 2019. "Metaphoric Extension, Relational Categories, and Abstraction." *Language, Cognition and Neuroscience* 34 (10): 1298–1307. <https://doi.org/10.1080/23273798.2017.1410560>.
- Gjerdingen, Robert, and Janet Bourne. 2015. "Schema Theory as a Construction Grammar." *Music Theory Online* 21 (2). [https://mtosmt.org/issues/mto.15.21.2/mto.15.21.2.gjerdingen\\_bourne.html](https://mtosmt.org/issues/mto.15.21.2/mto.15.21.2.gjerdingen_bourne.html).
- Gjerdingen, Robert O. 1988. *A Classic Turn of Phrase: Music and the Psychology of Convention*. 1st ed. Philadelphia: University of Pennsylvania Press.
- . 1996. "Courtly Behaviors." *Music Perception: An Interdisciplinary Journal* 13 (3): 365–82. <https://doi.org/10.2307/40286175>.
- . 2007. *Music in the Galant Style*. New York: Oxford University Press.
- . 2020. *Child Composers in the Old Conservatories: How Orphans Became Elite Musicians*. New York: Oxford University Press.
- Goldstone, Robert L., Alan Kersten, and Paulo F. Carvalho. 2012. "Concepts and Categorization." In *Handbook of Psychology, Second Edition*. American Cancer Society. <https://doi.org/10.1002/9781118133880.hop204022>.

- Goldstone, Robert L., and Mark Steyvers. 2001. "The Sensitization and Differentiation of Dimensions during Category Learning." *Journal of Experimental Psychology: General* 130 (1): 116–39. <https://doi.org/10.1037/0096-3445.130.1.116>.
- Gordon, Edwin E. 2004. *The Aural/Visual Experience of Music Literacy: Reading & Writing Music Notation*. Chicago, IL: GIA Publications, Inc.
- . 2012. *Learning Sequences in Music: Skill, Content, and Patterns*. Chicago, IL: GIA Publications, Inc.
- Greenspon, Emma B., Peter Q. Pfordresher, and Andrea R. Halpern. 2017. "Pitch Imitation Ability in Mental Transformations of Melodies." *Music Perception: An Interdisciplinary Journal* 34 (5): 585–604. <https://doi.org/10.1525/mp.2017.34.5.585>.
- Guck, Marion A. 2006. "Analysis as Interpretation: Interaction, Intentionality, Invention." *Music Theory Spectrum; Oxford* 28 (2): 191,193-209,321.
- . 2017. "Perceptions, Impressions: When Is Hearing 'Hearing-As'?" *Music Theory and Analysis (MTA)* 4 (2): 243–54. <https://doi.org/10.11116/MTA.4.2.5>.
- Hall, Anne Carothers. 2004. *Studying Rhythm*. 3rd ed. Upper Saddle River, N.J: Pearson.
- Halpern, Andrea R. 2015. "Differences in Auditory Imagery Self-Report Predict Neural and Behavioral Outcomes." *Psychomusicology: Music, Mind, and Brain, Musical Imagery*, 25 (1): 37–47. <https://doi.org/10.1037/pmu0000081>.
- Halpern, Andrea R, and Katie Overy. 2019. "Voluntary Auditory Imagery and Music Pedagogy." In *The Oxford Handbook of Sound and Imagination*, edited by Mark Grimshaw-Aagaar, Mads Walther-Hansen, and Martin Knakkegaard, 2:390–407. New York, NY: Oxford University Press.



- Halpern, Andrea R., and Robert J. Zatorre. 1999. "When That Tune Runs Through Your Head: A PET Investigation of Auditory Imagery for Familiar Melodies." *Cerebral Cortex* 9 (7): 697–704. <https://doi.org/10.1093/cercor/9.7.697>.
- Hampton, J. A. 1995. "Testing the Prototype Theory of Concepts." *Journal of Memory and Language* 34 (5): 686–708. <https://doi.org/10.1006/jmla.1995.1031>.
- Hampton, James A. 1995. "Similarity-Based Categorization: The Development of Prototype Theory." *Psychologica Belgica* 35 (2–3): 103–25. <https://doi.org/10.5334/pb.881>.
- Hanninen, Dora A. 2001. "Orientations, Criteria, Segments: A General Theory of Segmentation for Music Analysis." *Journal of Music Theory* 45 (2): 345–433. <https://doi.org/10.2307/3653443>.
- Hansberry, Benjamin. 2017. "Phenomenon and Abstraction: Coordinating Concepts in Music Theory and Analysis." Ph.D., United States -- New York: Columbia University. <http://search.proquest.com/pqdtglobal/docview/1906696381/abstract/5C662099A3394438PQ/1>.
- Harrison, Daniel. 1994. *Harmonic Function in Chromatic Music: A Renewed Dualist Theory and an Account of Its Precedents*. 1st ed. Chicago: University of Chicago Press.
- Hepokoski, James, and Warren Darcy. 2011. *Elements of Sonata Theory: Norms, Types, and Deformations in the Late-Eighteenth-Century Sonata*. Oxford: Oxford University Press.
- Herholz, Sibylle C., Andrea R. Halpern, and Robert J. Zatorre. 2012. "Neuronal Correlates of Perception, Imagery, and Memory for Familiar Tunes." *Journal of Cognitive Neuroscience* 24 (6): 1382–97.
- Hessels, Roy S., Diederick C. Niehorster, Marcus Nyström, Richard Andersson, and Ignace T. C. Hooge. 2018. "Is the Eye-Movement Field Confused about Fixations and Saccades? A Survey among 124 Researchers." *Royal Society Open Science* 5 (8): 180502. <https://doi.org/10.1098/rsos.180502>.

- Hintzman, Douglas L. 1986. “‘Schema Abstraction’ in a Multiple-Trace Memory Model.” *Psychological Review* 93 (4): 411–28. <https://doi.org/10.1037/0033-295X.93.4.411>.
- Hintzman, Douglas L., and Genevieve Ludlam. 1980. “Differential Forgetting of Prototypes and Old Instances: Simulation by an Exemplar-Based Classification Model.” *Memory & Cognition* 8 (4): 378–82. <https://doi.org/10.3758/BF03198278>.
- Hoemann, Katie, Madeleine Devlin, and Lisa Feldman Barrett. 2020. “Comment: Emotions Are Abstract, Conceptual Categories That Are Learned by a Predicting Brain.” *Emotion Review* 12 (4): 253–55. <https://doi.org/10.1177/1754073919897296>.
- Holm-Hudson, Kevin. 2016. *Music Theory Remixed: A Blended Approach for the Practicing Musician*. 1st ed. New York: Oxford University Press.
- Horlitz, Krista L., and Ann O’Leary. 1993. “Satiation or Availability? Effects of Attention, Memory, and Imagery on the Perception of Ambiguous Figures.” *Perception & Psychophysics* 53 (6): 668–81. <https://doi.org/10.3758/BF03211743>.
- Hubbard, Timothy L. 2010. “Auditory Imagery: Empirical Findings.” *Psychological Bulletin* 136 (2): 302–29. <https://doi.org/10.1037/a0018436>.
- Humphreys, Michael S., Gerald Tehan, Oliver Baumann, and Shayne Loft. 2020. “Explaining Short-Term Memory Phenomena with an Integrated Episodic/Semantic Framework of Long-Term Memory.” *Cognitive Psychology* 123: 101346. <https://doi.org/10.1016/j.cogpsych.2020.101346>.
- Hunt, Earl, and Franca Agnoli. 1991. “The Whorfian Hypothesis: A Cognitive Psychology Perspective.” *Psychological Review* 98 (3): 377–89. <https://doi.org/10.1037/0033-295X.98.3.377>.
- Ijzerman, Job. 2018. *Harmony, Counterpoint, Partimento: A New Method Inspired by Old Masters*. New York: Oxford University Press.

- Intaitė, Monika, Valdas Noreika, Alvydas Šoliūnas, and Christine M. Falter. 2013. “Interaction of Bottom-up and Top-down Processes in the Perception of Ambiguous Figures.” *Vision Research* 89: 24–31. <https://doi.org/10.1016/j.visres.2013.06.011>.
- Iverson, Paul, Valerie Hazan, and Kerry Bannister. 2005. “Phonetic Training with Acoustic Cue Manipulations: A Comparison of Methods for Teaching English /r/-/l/ to Japanese Adults.” *The Journal of the Acoustical Society of America* 118 (5): 3267–78. <https://doi.org/10.1121/1.2062307>.
- Janata, Petr, Jeffrey L. Birk, John D. Van Horn, Marc Leman, Barbara Tillmann, and Jamshed J. Bharucha. 2002. “The Cortical Topography of Tonal Structures Underlying Western Music.” *Science* 298 (5601): 2167–70. <https://doi.org/10.1126/science.1076262>.
- Jastrow, Joseph. 1899. “The Mind’s Eye.” *Popular Science Monthly* 54: 299–312.
- Johnson, Mark. 2009. *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. Chicago: University of Chicago Press.
- Karpinski, Gary S. 2000. *Aural Skills Acquisition: The Development of Listening, Reading, and Performing Skills in College-Level Musicians*. 1st ed. New York, NY: Oxford University Press.
- . 2017. *Manual for Ear Training and Sight Singing*. 2nd ed. New York, NY: W. W. Norton & Company.
- Kim, ShinWoo, and Bob Rehder. 2011. “How Prior Knowledge Affects Selective Attention during Category Learning: An Eyetracking Study.” *Memory & Cognition* 39 (4): 649–65. <https://doi.org/10.3758/s13421-010-0050-3>.
- Kinghorn, Elizabeth, Rebekka Lagace-Cusiac, Ozgen Demirkaplan, Jessica A. Grahn, Jonathan De Souza, and Christine Carter. 2021. “The Effects of Interleaved and Blocked Practice on Musical Style Recognition.” In *ICMPC16-ESCOM11*. Virtual Conference.

- Kintsch, Walter. 1998. *Comprehension: A Paradigm for Cognition*. Cambridge; New York: Cambridge University Press.
- Kivunja, Charles. 2018. "Distinguishing between Theory, Theoretical Framework, and Conceptual Framework: A Systematic Review of Lessons from the Field." *International Journal of Higher Education* 7 (6): 44. <https://doi.org/10.5430/ijhe.v7n6p44>.
- Knowlton, Barbara J., and Larry R. Squire. 1993. "The Learning of Categories: Parallel Brain Systems for Item Memory and Category Knowledge." *Science* 262 (5140): 1747–49.
- Koreimann, Sabrina, Bartosz Gula, and Oliver Vitouch. 2014. "Inattentional Deafness in Music." *Psychological Research* 78 (3): 304–12. <https://doi.org/10.1007/s00426-014-0552-x>.
- Kostka, Stefan, Dorothy Payne, and Byron Almen. 2012. *Tonal Harmony*. 7th ed. New York: McGraw-Hill Education.
- Kousta, Stavroula-Thaleia, Gabriella Vigliocco, David P. Vinson, Mark Andrews, and Elena Del Campo. 2011. "The Representation of Abstract Words: Why Emotion Matters." *Journal of Experimental Psychology: General* 140 (1): 14–34. <https://doi.org/10.1037/a0021446>.
- Krogh, Lauren, Haley Vlach, and Scott Johnson. 2013. "Statistical Learning Across Development: Flexible Yet Constrained." *Frontiers in Psychology* 3: 598. <https://doi.org/10.3389/fpsyg.2012.00598>.
- Krumhansl, Carol L. 2001. *Cognitive Foundations of Musical Pitch*. Oxford: Oxford University Press.
- Krumhansl, Carol L., and Petri Toiviainen. 2001. "Tonal Cognition." *Annals of the New York Academy of Sciences* 930 (1): 77–91. <https://doi.org/10.1111/j.1749-6632.2001.tb05726.x>.
- Kruschke, John K. 1992. "ALCOVE: An Exemplar-Based Connectionist Model of Category Learning." *Psychological Review* 99 (1): 22–44. <https://doi.org/10.1037/0033-295X.99.1.22>.

- . 1993. “Human Category Learning: Implications for Backpropagation Models.” *Connection Science* 5 (1): 3–36. <https://doi.org/10.1080/09540099308915683>.
- Laitz, Steven G. 2016a. *The Complete Musician: An Integrated Approach to Theory, Analysis and Listening*. 4th ed. Oxford; New York: Oxford University Press.
- . 2016b. *The Complete Musician: An Integrated Approach to Theory, Analysis, and Listening*. 4th ed. Oxford; New York: Oxford University Press.
- Lakoff, George. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: The University of Chicago Press.
- Larson, Steve. 1993. “Scale-Degree Function: A Theory of Expressive Meaning and Its Application to Aural Skills Pedagogy.” *Journal of Music Theory Pedagogy* 7: 69–84.
- . 2012. *Musical Forces: Motion, Metaphor, and Meaning in Music*. Bloomington, IN: Indiana University Press.
- <http://ebookcentral.proquest.com/lib/northwestern/detail.action?docID=670304>.
- Lawrence, Douglas H. 1952. “The Transfer of a Discrimination along a Continuum.” *Journal of Comparative and Physiological Psychology* 45 (6): 511–16. <https://doi.org/10.1037/h0057135>.
- Leptourgos, Pantelis, Charles-Edouard Notredame, Marion Eck, Renaud Jardri, and Sophie Denève. 2020. “Circular Inference in Bistable Perception.” *Journal of Vision* 20 (4): 12. <https://doi.org/10.1167/jov.20.4.12>.
- Lerdahl, Fred. 2001. *Tonal Pitch Space*. Oxford: Oxford University Press.
- Lerdahl, Fred, and Ray S. Jackendoff. 1983. *A Generative Theory of Tonal Music*. MIT Press.
- Letailleur, Alain, Erica Bisesi, and Pierre Legrain. 2020. “Strategies Used by Musicians to Identify Notes’ Pitch: Cognitive Bricks and Mental Representations.” *Frontiers in Psychology* 11: 1480. <https://doi.org/10.3389/fpsyg.2020.01480>.

Levesque, and Bèche. 1779. *Solfèges d'Italie Avec La Basse Chiffrée*. 2nd ed. Paris.

<https://gallica.bnf.fr/ark:/12148/bpt6k11689561>.

Lewin, David. 1986. "Music Theory, Phenomenology, and Modes of Perception." *Music Perception: An Interdisciplinary Journal* 3 (4): 327–92. <https://doi.org/10.2307/40285344>.

———. 2007a. *Generalized Musical Intervals and Transformations*. New York: Oxford University Press.

———. 2007b. "Making and Using a PCset Network for Stockhausen's Klavierstücke III." In *Musical Form and Transformation: Four Analytic Essays*, 16–67. New York: Oxford University Press.

———. 2007c. *Musical Form and Transformation: Four Analytic Essays*. New York: Oxford University Press.

Lima, César F., Nadine Lavan, Samuel Evans, Zarinah Agnew, Andrea R. Halpern, Pradheep Shanmugalingam, Sophie Meekings, et al. 2015. "Feel the Noise: Relating Individual Differences in Auditory Imagery to the Structure and Function of Sensorimotor Systems." *Cerebral Cortex* 25 (11): 4638–50. <https://doi.org/10.1093/cercor/bhv134>.

Long, Gerald M., and Thomas C. Toppino. 2004. "Enduring Interest in Perceptual Ambiguity: Alternating Views of Reversible Figures." *Psychological Bulletin* 130 (5): 748–68.

<https://doi.org/10.1037/0033-2909.130.5.748>.

Love, Bradley C., Douglas L. Medin, and Todd M. Gureckis. 2004. "SUSTAIN: A Network Model of Category Learning." *Psychological Review* 111 (2): 309–32. <https://doi.org/10.1037/0033-295X.111.2.309>.

- Lupyan, Gary, David H. Rakison, and James L. McClelland. 2007. "Language Is Not Just for Talking: Redundant Labels Facilitate Learning of Novel Categories." *Psychological Science* (0956-7976) 18 (12): 1077–83. <https://doi.org/10.1111/j.1467-9280.2007.02028.x>.
- Macdonald, James S. P., and Nilli Lavie. 2011. "Visual Perceptual Load Induces Inattentional Deafness." *Attention, Perception, & Psychophysics* 73 (6): 1780–89. <https://doi.org/10.3758/s13414-011-0144-4>.
- Mack, Michael L., and Thomas J. Palmeri. 2010. "Decoupling Object Detection and Categorization." *Journal of Experimental Psychology: Human Perception and Performance* 36 (5): 1067–79. <https://doi.org/10.1037/a0020254>.
- Mackintosh, N. J., and Lydia Little. 1970. "An Analysis of Transfer along a Continuum." *Canadian Journal of Psychology/Revue Canadienne de Psychologie* 24 (5): 362–69. <https://doi.org/10.1037/h0082872>.
- Martinovic, Jasna, Galina V. Paramei, and W. Joseph MacInnes. 2020. "Russian Blues Reveal the Limits of Language Influencing Colour Discrimination." *Cognition* 201: 104281. <https://doi.org/10.1016/j.cognition.2020.104281>.
- Maxfield, Justin T., and Gregory J. Zelinsky. 2012. "Searching through the Hierarchy: How Level of Target Categorization Affects Visual Search." *Visual Cognition* 20 (10): 1153–63. <https://doi.org/10.1080/13506285.2012.735718>.
- McAvinue, Laura P., and Ian H. Robertson. 2007. "Measuring Visual Imagery Ability: A Review." *Imagination, Cognition and Personality* 26 (3): 191–211. <https://doi.org/10.2190/3515-8169-24J8-7157>.
- McCandliss, Bruce D., Julie A. Fiez, Athanassios Protopapas, Mary Conway, and James L. McClelland. 2002. "Success and Failure in Teaching the [r]-[l] Contrast to Japanese Adults:

Tests of a Hebbian Model of Plasticity and Stabilization in Spoken Language Perception.”

*Cognitive, Affective, & Behavioral Neuroscience* 2 (2): 89–108.

<https://doi.org/10.3758/CABN.2.2.89>.

McClelland, James L., Julie A. Fiez, and Bruce D. McCandliss. 2002. “Teaching the /r/-/l/

Discrimination to Japanese Adults: Behavioral and Neural Aspects.” *Physiology & Behavior* 77

(4): 657–62. [https://doi.org/10.1016/S0031-9384\(02\)00916-2](https://doi.org/10.1016/S0031-9384(02)00916-2).

McRae, Ken, and Michael Jones. 2013. “Chapter 14: Semantic Memory.” In *The Oxford Handbook of*

*Cognitive Psychology*, edited by Daniel Reisberg, 206–19. New York: Oxford University Press.

Medin, Douglas L., and Judy E. Florian. 1992. “Abstraction and Selective Coding In Exemplar-Based

Models of Categorization.” In *From Learning Processes to Cognitive Processes: Essays in*

*Honor of William K. Estes*, edited by Alice Healy F., Stephen M. Kosslyn, and Richard M.

Shiffrin, 2:207–34. Psychology Press.

Medin, Douglas L., and Marguerite M. Schaffer. 1978. “Context Theory of Classification Learning.”

*Psychological Review* 85 (3): 207–38. <https://doi.org/10.1037/0033-295X.85.3.207>.

Meng, Ming, and Frank Tong. 2004. “Can Attention Selectively Bias Bistable Perception?

Differences between Binocular Rivalry and Ambiguous Figures.” *Journal of Vision* 4 (7): 2–2.

<https://doi.org/10.1167/4.7.2>.

Meyer, Leonard B. 1956. *Emotion and Meaning in Music*. Chicago, IL: University of Chicago Press.

———. 1973. *Explaining Music*. Berkeley, CA: University of California Press.

———. 1994. *Music, the Arts, and Ideas: Patterns and Predictions in Twentieth-Century Culture*.

Chicago: University of Chicago Press.



- Minda, John Paul, and Sarah J. Miles. 2010. "The Influence of Verbal and Nonverbal Processing on Category Learning." In *Psychology of Learning and Motivation*, 52:117–62. Elsevier.  
[https://doi.org/10.1016/S0079-7421\(10\)52003-6](https://doi.org/10.1016/S0079-7421(10)52003-6).
- Minda, John Paul, and J. David Smith. 2001. "Prototypes in Category Learning: The Effects of Category Size, Category Structure, and Stimulus Complexity." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27 (3): 775–99. <https://doi.org/10.1037/0278-7393.27.3.775>.
- . 2002. "Comparing Prototype-Based and Exemplar-Based Accounts of Category Learning and Attentional Allocation." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28 (2): 275–92. <https://doi.org/10.1037/0278-7393.28.2.275>.
- Nelson, Thomas O., and Louis Narens. 1990. "Metamemory: A Theoretical Framework and New Findings." In *Psychology of Learning and Motivation*, 26:125–73. Elsevier.  
[https://doi.org/10.1016/S0079-7421\(08\)60053-5](https://doi.org/10.1016/S0079-7421(08)60053-5).
- Nosofsky, Robert M. 1986. "Attention, Similarity, and the Identification–Categorization Relationship." *Journal of Experimental Psychology: General* 115 (1): 39–57.  
<https://doi.org/10.1037/0096-3445.115.1.39>.
- . 1988. "Exemplar-Based Accounts of Relations between Classification, Recognition, and Typicality." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14 (4): 700–708. <https://doi.org/10.1037/0278-7393.14.4.700>.
- . 2000. "Exemplar Representation without Generalization? Comment on Smith and Minda's (2000) 'Thirty Categorization Results in Search of a Model.'" *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26 (6): 1735–43. <https://doi.org/10.1037/0278-7393.26.6.1735>.

- Nosofsky, Robert M., and Mark K. Johansen. 2000. "Exemplar-Based Accounts of 'Multiple-System' Phenomena in Perceptual Categorization." *Psychonomic Bulletin & Review* 7 (3): 375–402. <https://doi.org/10.1007/BF03543066>.
- Nosofsky, Robert M., John K. Kruschke, and Stephen C. McKinley. 1992. "Combining Exemplar-Based Category Representations and Connectionist Learning Rules." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18 (2): 211–33. <https://doi.org/10.1037/0278-7393.18.2.211>.
- Nosofsky, Robert M., and Safa R. Zaki. 2002. "Exemplar and Prototype Models Revisited: Response Strategies, Selective Attention, and Stimulus Generalization." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28 (5): 924–40. <https://doi.org/10.1037/0278-7393.28.5.924>.
- Paivio, A. 1978. *Imagery and Verbal Processes*. Psychology Press.
- Paivio, Allan. 1965. "Abstractness, Imagery, and Meaningfulness in Paired-Associate Learning." *Journal of Verbal Learning and Verbal Behavior* 4 (1): 32–38. [https://doi.org/10.1016/S0022-5371\(65\)80064-0](https://doi.org/10.1016/S0022-5371(65)80064-0).
- . 1986. *Mental Representations: A Dual Coding Approach*. Oxford Psychology Series, no. 9. New York; Oxford: Oxford University Press ; Clarendon Press.
- . 2007. *Mind and Its Evolution*. 1<sup>st</sup> ed. Mahwah, N.J: Routledge.
- . 2013. "Dual Coding Theory, Word Abstractness, and Emotion: A Critical Review of Kousta et al. (2011)." *Journal of Experimental Psychology: General* 142 (1): 282–87. <https://doi.org/10.1037/a0027004>.

- Paivio, Allan, James M. Clark, and Mustaq Khan. 1988. "Effects of Concreteness and Semantic Relatedness on Composite Imagery Ratings and Cued Recall." *Memory & Cognition* 16 (5): 422–30. <https://doi.org/10.3758/BF03214222>.
- Paivio, Allan, and Mary Walsh. 1993. "Psychological Processes in Metaphor Comprehension and Memory." In *Metaphor and Thought*, edited by Andrew Ortony, 2nd ed., 307–28. Cambridge; New York: Cambridge University Press.
- Pearson, Joel, Colin W. G. Clifford, and Frank Tong. 2008. "The Functional Impact of Mental Imagery on Conscious Perception." *Current Biology* 18 (13): 982–86. <https://doi.org/10.1016/j.cub.2008.05.048>.
- Pearson, Joel, and Stephen M. Kosslyn. 2015. "The Heterogeneity of Mental Representation: Ending the Imagery Debate." *Proceedings of the National Academy of Sciences of the United States of America* 112 (33): 10089–92.
- Peretz, Isabelle, Dominique Vuvan, Marie-Élaine Lagrois, and Jorge L. Armony. 2015. "Neural Overlap in Processing Music and Speech." *Philosophical Transactions of the Royal Society B: Biological Sciences* 370 (1664): 20140090. <https://doi.org/10.1098/rstb.2014.0090>.
- Pfordresher, Peter Q., and Andrea R. Halpern. 2013. "Auditory Imagery and the Poor-Pitch Singer." *Psychonomic Bulletin & Review* 20 (4): 747–53. <https://doi.org/10.3758/s13423-013-0401-8>.
- Piston, Walter. 1969. *Harmony*. 3rd ed. New York: W. W. Norton.
- Rabinovitch, Gilad. 2013. "'Schenker the Galant?' Tacit Knowledge, Contradiction, and Complementation in the Interaction between Gjerdingen's Theory of Galant Schemata and Schenkerian Analysis." Ph.D., New York: University of Rochester. <http://search.proquest.com/docview/1465052118/abstract/1992D124F8C145EBPQ/1>.

- . 2015. “Tracing Galant Threads: Gjerdingen’s Schemata and the Evolution of Musical Form, 1730-1780.”
- . 2018. “Gjerdingen’s Schemata Reexamined.” *Journal of Music Theory* 62 (1): 41–84. <https://doi.org/10.1215/00222909-4450636>.
- . 2019. “Implicit Counterpoint in Gjerdingen’s Schemata.” *Music Theory and Analysis (MTA)* 6 (1): 1–50. <https://doi.org/10.11116/MTA.6.1.1>.
- . 2020. “Hidden Polyphony, Linear Hierarchy, and Scale-Degree Associations in Galant Schemata.” *Indiana Theory Review* 36 (1): 114–66.
- Raveh, Dana, and Nilli Lavie. 2015. “Load-Induced Inattentional Deafness.” *Attention, Perception, & Psychophysics* 77 (2): 483–92. <https://doi.org/10.3758/s13414-014-0776-2>.
- Rehder, Bob, and Aaron B. Hoffman. 2005a. “Eyetracking and Selective Attention in Category Learning.” *Cognitive Psychology* 51 (1): 1–41. <https://doi.org/10.1016/j.cogpsych.2004.11.001>.
- . 2005b. “Thirty-Something Categorization Results Explained: Selective Attention, Eyetracking, and Models of Category Learning.” *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31 (5): 811–29. <https://doi.org/10.1037/0278-7393.31.5.811>.
- Rips, Lance J., Edward E. Smith, and Douglas L. Medin. 2012. “Concepts and Categories: Memory, Meaning, and Metaphysics.” In *The Oxford Handbook of Thinking and Reasoning*, 177–209. Oxford; New York: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199734689.013.0011>.
- Roark, Casey L., and Lori L. Holt. 2015. “Rule-Based and Information-Integration Categorization during an Incidental Learning Task.” *The Journal of the Acoustical Society of America* 137 (4): 2385–2385. <https://doi.org/10.1121/1.4920681>.

- . 2018. “Task and Distribution Sampling Affect Auditory Category Learning.” *Attention, Perception & Psychophysics* 80 (7): 1804–22. <https://doi.org/10.3758/s13414-018-1552-5>.
- Roark, Casey L., Kirsten Smayda, and Bharath Chandrasekaran. 2020. “Auditory and Visual Category Learning in Musicians and Non-Musicians.” PsyArXiv. <https://doi.org/10.31234/osf.io/q9gwx>.
- Rogers, Michael. 2004. *Teaching Approaches in Music Theory: An Overview of Pedagogical Philosophies*. 2nd ed. Carbondale: Southern Illinois University Press.
- Rosch, Eleanor. 1975. “Cognitive Representations of Semantic Categories.” *Journal of Experimental Psychology: General* 104 (3): 192–233. <https://doi.org/10.1037/0096-3445.104.3.192>.
- Rosch, Eleanor, and Carolyn B Mervis. 1975. “Family Resemblances: Studies in the Internal Structure of Categories.” *Cognitive Psychology* 7 (4): 573–605. [https://doi.org/10.1016/0010-0285\(75\)90024-9](https://doi.org/10.1016/0010-0285(75)90024-9).
- Ross, Brian H, and Valerie Makin S. 1999. “Prototype versus Exemplar Models.” In *The Nature of Cognition*, edited by Robert J Sternberg, 205–43. Cambridge, MA: MIT Press.
- Rumelhart, David E. 2017. “Schemata: The Building Blocks of Cognition.” In *Theoretical Issues in Reading Comprehension: Perspectives from Cognitive Psychology, Linguistics, Artificial Intelligence and Education*, edited by Rand J. Spiro, Bertram C. Bruce, and William F. Brewer. Routledge.
- Rumelhart, David E., and Andrew Ortony. 1977. “The Representation of Knowledge in Memory.” In *Schooling and the Acquisition of Knowledge*, edited by Richard Anderson C., Rand J. Spiro, and William Montague E., 99–135. Hillsdale, N.J.: Erlbaum.
- Sadoski, Mark, and Allan Paivio. 2012. *Imagery and Text: A Dual Coding Theory of Reading and Writing*. 2nd ed. New York: Routledge.

- Salamé, Pierre, and Alan Baddeley. 1989. "Effects of Background Music on Phonological Short-Term Memory." *The Quarterly Journal of Experimental Psychology Section A* 41 (1): 107–22.  
<https://doi.org/10.1080/14640748908402355>.
- Salzer, Felix. 1962. *Structural Hearing: Tonal Coherence in Music*. Vol. 1. 2 vols. New York: Dover Publications.
- Sanguinetti, Giorgio. 2012. *The Art of Partimento: History, Theory and Practice*. New York: Oxford University Press.
- Savard, Augustin. 1877. *Cours complet d'harmonie théorique et pratique*. Paris: E. & A. Girod.  
<https://gallica.bnf.fr/ark:/12148/bpt6k1172790s>.
- Schwab-Felisch, Oliver. 2014. "The Butterfly and the Artillery: Models of Listening in Schenker and Gjerdingen." *Music Theory and Analysis (MTA)* 1 (1–2): 107–20.  
<https://doi.org/10.11116/MTA.1.6>.
- Schwartz, Bennett L., and Janet Metcalfe. 1992. "Cue Familiarity but Not Target Retrievalability Enhances Feeling-of-Knowing Judgments." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18 (5): 1074–83. <https://doi.org/10.1037/0278-7393.18.5.1074>.
- . 2011. "Tip-of-the-Tongue (TOT) States: Retrieval, Behavior, and Experience." *Memory & Cognition* 39 (5): 737–49. <https://doi.org/10.3758/s13421-010-0066-8>.
- Sears, David, William E. Caplin, and Stephen McAdams. 2014. "Perceiving the Classical Cadence." *Music Perception: An Interdisciplinary Journal* 31 (5): 397–417.  
<https://doi.org/10.1525/mp.2014.31.5.397>.
- Sears, David R. W., Marcus T. Pearce, William E. Caplin, and Stephen McAdams. 2018. "Simulating Melodic and Harmonic Expectations for Tonal Cadences Using Probabilistic Models." *Journal of New Music Research* 47 (1): 29–52. <https://doi.org/10.1080/09298215.2017.1367010>.

- Sears, David R. W., Jacob Spitzer, William E. Caplin, and Stephen McAdams. 2020. "Expecting the End: Continuous Expectancy Ratings for Tonal Cadences." *Psychology of Music* 48 (3): 358–75. <https://doi.org/10.1177/0305735618803676>.
- Sears, David Robert William. 2016. "The Classical Cadence as a Closing Schema: Learning, Memory, & Perception." Doctoral Dissertation, Montreal, CA: McGill University.
- Simons, Daniel J, and Christopher F Chabris. 1999. "Gorillas in Our Midst: Sustained Inattentional Blindness for Dynamic Events." *Perception* 28 (9): 1059–74. <https://doi.org/10.1068/p281059>.
- Smayda, Kirsten E., Bharath Chandrasekaran, and W. Todd Maddox. 2015. "Enhanced Cognitive and Perceptual Processing: A Computational Basis for the Musician Advantage in Speech Learning." *Frontiers in Psychology* 6. <https://doi.org/10.3389/fpsyg.2015.00682>.
- Smith, David J., and John Paul Minda. 2000. "Thirty Categorization Results in Search of a Model." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26 (1): 3–27. <https://doi.org/10.1037/0278-7393.26.1.3>.
- Smith, Edward E., and Douglas L. Medin. 1981. *Categories and Concepts*. Cognitive Science Series 4. Cambridge, Mass: Harvard University Press.
- Symons, James. 2017. "A Cognitively Inspired Method for the Statistical Analysis of Eighteenth-Century Music, as Applied in Two Corpus Studies." Ph.D., Illinois: Northwestern University. <http://www.proquest.com/docview/1984374211/abstract/D18648606BE841FCPQ/1>.
- Taddei-Ferretti, C., J. Radilova, C. Musio, S. Santillo, E. Cibelli, A. Cotugno, and T. Radil. 2008. "The Effects of Pattern Shape, Subliminal Stimulation, and Voluntary Control on Multistable Visual Perception." *Brain Research, Brain and Vision*, 1225: 163–70. <https://doi.org/10.1016/j.brainres.2008.04.064>.

- Temperley, David. 1999. "The Question of Purpose in Music Theory: Description, Suggestion, and Explanation." *Current Musicology; New York, N. Y.*, 66–85.
- . 2001. *The Cognition of Basic Musical Structures*. Cambridge, Mass: MIT Press.
- . 2006. "Music in the Galant Style." *Journal of Music Theory* 50 (2): 277–90.  
<https://doi.org/10.1215/00222909-2008-018>.
- . 2009. "In Defense of Introspectionism: A Response to DeBellis." *Music Perception* 27 (2): 131–38. <https://doi.org/10.1525/mp.2009.27.2.131>.
- Teschner, Gustav Wilhelm. 1872. *Solfeggi Für Sopranstimme von Niccola Zingarelli*. Magdeburg: Heinrichshofen'sche Musikalien-Handlung. Plate H.M.2240, H.M. 2241.
- Thiessen, Erik D., and Lucy C. Erickson. 2013. "Beyond Word Segmentation: A Two-Process Account of Statistical Learning." *Current Directions in Psychological Science* 22 (3): 239–43.
- Thiessen, Erik D., Alexandra T. Kronstein, and Daniel G. Hufnagle. 2013. "The Extraction and Integration Framework: A Two-Process Account of Statistical Learning." *Psychological Bulletin* 139 (4): 792–814. <https://doi.org/10.1037/a0030801>.
- Thiessen, Erik D., and Philip I. Pavlik. 2013. "IMinerva: A Mathematical Model of Distributional Statistical Learning." *Cognitive Science* 37 (2): 310–43. <https://doi.org/10.1111/cogs.12011>.
- Tomasello, Michael. 2005. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. 1<sup>st</sup> ed. Cambridge, Mass.: Harvard Univ. Press.
- Toppino, Thomas C. 2003. "Reversible-Figure Perception: Mechanisms of Intentional Control." *Perception & Psychophysics* 65 (8): 1285–95. <https://doi.org/10.3758/BF03194852>.
- Tour, Peter van. 2015. *Counterpoint and Partimento: Methods of Teaching Composition in Late Eighteenth-Century Naples*. 3rd ed. Uppsala: Uppsala Universitet.



- Tulving, Endel. 1972. "Episodic and Semantic Memory." In *Organization and Memory*, edited by Endel Tulving and Wayne Donaldson, 381–403. New York: Academic Press.
- . 1991. "Concepts of Human Memory." In *Memory: Organization and Locus of Change*, edited by Larry R. Squire, Norman M. Weinberger, Gary Lynch, and James L. McGaugh, 3–32. Oxford University Press.
- . 1993. "What Is Episodic Memory?" *Current Directions in Psychological Science* 2 (3): 67–70.
- Turgeon, Martine, and Albert S. Bregman. 2001. "Ambiguous Musical Figures." *Annals of the New York Academy of Sciences* 930 (1): 375–81. <https://doi.org/10.1111/j.1749-6632.2001.tb05746.x>.
- Van Overschelde, James, P. 2013. "Metacognition: Knowing About Knowing." In *Handbook of Metamemory and Memory*, edited by John Dunlosky and Robert A. Bjork, 47–72. New York: Psychology Press. <https://doi.org/10.4324/9780203805503>.
- Vanpaemel, Wolf, and Gert Storms. 2008. "In Search of Abstraction: The Varying Abstraction Model of Categorization." *Psychonomic Bulletin & Review* 15 (4): 732–49. <https://doi.org/10.3758/PBR.15.4.732>.
- Vasuki, Pragati Rao Mandikal, Mridula Sharma, Katherine Demuth, and Joanne Arciuli. 2016. "Musicians' Edge: A Comparison of Auditory Processing, Cognitive Abilities and Statistical Learning." *Hearing Research* 342: 112–23. <https://doi.org/10.1016/j.heares.2016.10.008>.
- Vigliocco, Gabriella, Stavroula Kousta, David Vinson, Mark Andrews, and Elena Del Campo. 2013. "The Representation of Abstract Words: What Matters? Reply to Paivio's (2013) Comment on Kousta et al (2011)." *Journal of Experimental Psychology: General* 142 (1): 288–91. <https://doi.org/10.1037/a0028749>.

- Vuvan, Dominique T., and Mark A. Schmuckler. 2011. "Tonal Hierarchy Representations in Auditory Imagery." *Memory & Cognition* 39 (3): 477–90. <https://doi.org/10.3758/s13421-010-0032-5>.
- Wang, Jing, Julie A. Conder, David N. Blitzer, and Svetlana V. Shinkareva. 2010. "Neural Representation of Abstract and Concrete Concepts: A Meta-Analysis of Neuroimaging Studies." *Human Brain Mapping* 31 (10): 1459–68. <https://doi.org/10.1002/hbm.20950>.
- Warrington, Elizabeth, K., and Rosaleen McCarthy A. 1987. "Categories of Knowledge: Further Fractionations and an Attempted Integration." *Brain* 110 (5): 1273–96. <https://doi.org/10.1093/brain/110.5.1273>.
- Weiser, Margaret. 1990. "Rating Cadence Stability: The Effects of Chord Structure, Tonal Context and Musical Training." Thesis, McMaster University. <https://macsphere.mcmaster.ca/handle/11375/23585>.
- Wiemer-Hastings, Katja, and Xu Xu. 2005. "Content Differences for Abstract and Concrete Concepts." *Cognitive Science* 29 (5): 719–36. [https://doi.org/10.1207/s15516709cog0000\\_33](https://doi.org/10.1207/s15516709cog0000_33).
- Wiener, Seth, and Evan D. Bradley. 2020. "Harnessing the Musician Advantage: Short-Term Musical Training Affects Non-Native Cue Weighting of Linguistic Pitch." *Language Teaching Research*, 1362168820971791. <https://doi.org/10.1177/1362168820971791>.
- Wisniewski, Matthew G., Barbara A. Church, Eduardo Mercado, Milen L. Radell, and Alexandria C. Zakrzewski. 2019. "Easy-to-Hard Effects in Perceptual Learning Depend upon the Degree to Which Initial Trials Are 'Easy.'" *Psychonomic Bulletin & Review* 26 (6): 1889–95. <https://doi.org/10.3758/s13423-019-01627-4>.
- Wisniewski, Matthew G., Milen L. Radell, Barbara A. Church, and Eduardo Mercado. 2017. "Benefits of Fading in Perceptual Learning Are Driven by More than Dimensional Attention." *PLOS ONE* 12 (7): e0180959. <https://doi.org/10.1371/journal.pone.0180959>.

- Witzel, Christoph, and Karl R. Gegenfurtner. 2015. "Categorical Facilitation with Equally Discriminable Colors." *Journal of Vision* 15 (8): 22. <https://doi.org/10.1167/15.8.22>.
- . 2016. "Categorical Perception for Red and Brown." *Journal of Experimental Psychology: Human Perception and Performance* 42 (4): 540–70. <https://doi.org/10.1037/xhp0000154>.
- Zaki, Safa R., and Robert M. Nosofsky. 2004. "False Prototype Enhancement Effects in Dot Pattern Categorization." *Memory & Cognition* 32 (3): 390–98. <https://doi.org/10.3758/BF03195833>.
- Zaki, Safa R., Robert M. Nosofsky, Roger D. Stanton, and Andrew L. Cohen. 2003. "Prototype and Exemplar Accounts of Category Learning and Attentional Allocation: A Reassessment." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 29 (6): 1160–73. <https://doi.org/10.1037/0278-7393.29.6.1160>.
- Zatorre, Robert J., and Andrea R. Halpern. 1993. "Effect of Unilateral Temporal-Lobe Excision on Perception and Imagery of Songs." *Neuropsychologia* 31 (3): 221–32. [https://doi.org/10.1016/0028-3932\(93\)90086-F](https://doi.org/10.1016/0028-3932(93)90086-F).
- Zatorre, Robert J., Andrea R. Halpern, and Marc Bouffard. 2009. "Mental Reversal of Imagined Melodies: A Role for the Posterior Parietal Cortex." *Journal of Cognitive Neuroscience* 22 (4): 775–89. <https://doi.org/10.1162/jocn.2009.21239>.
- Zatorre, Robert J., Andrea R. Halpern, David W. Perry, Ernst Meyer, and Alan C. Evans. 1996. "Hearing in the Mind's Ear: A PET Investigation of Musical Imagery and Perception." *Journal of Cognitive Neuroscience* 8 (1): 29–46. <https://doi.org/10.1162/jocn.1996.8.1.29>.
- Zbikowski, Lawrence M. 2002. *Conceptualizing Music: Cognitive Structure, Theory, and Analysis*. New York, NY: Oxford University Press.

Zettersten, Martin, and Gary Lupyan. 2020. "Finding Categories through Words: More Nameable Features Improve Category Learning." *Cognition* 196: 104135.

<https://doi.org/10.1016/j.cognition.2019.104135>.