Tonal Emergence: An Agent-Based Model of Tonal Coordination

Matthew D. Setzler[1], Robert L. Goldstone

Cognitive Science Program, Indiana University. 107 S Indiana Ave, Bloomington IN, 47405, USA

Author Note

Corresponding author: Matthew Setzler, mattsetz@gmail.com

[1] Current Affiliation: Pacific Northwest National Labs. 1100 Dexter Ave N, Seattle, WA 98109, USA

Abstract

Humans have a remarkable capacity for coordination. Our ability to interact and act jointly in groups is crucial to our success as a species. Joint Action (JA) research has often concerned itself with simplistic behaviors in highly constrained laboratory tasks. But there has been a growing interest in understanding complex coordination in more open-ended contexts. In this regard, collective music improvisation has emerged as a fascinating model domain for studying basic JA mechanisms in an unconstrained and highly sophisticated setting. A number of empirical studies have begun to elucidate coordination mechanisms underlying joint musical improvisation, but these empirical findings have yet to be cached out in a working computational model. The present work fills this gap by presenting TonalEmergence, an idealized agent-based model of improvised musical coordination. TonalEmergence models the coordination of notes played by improvisers to generate harmony (i.e., tonality), by simulating agents that stochastically generate notes biased towards maximizing harmonic consonance given their partner's previous notes. The model replicates an interesting empirical result from a previous study of professional jazz pianists: feedback loops of mutual adaptation between interacting agents support the production of consonant harmony. The model is further explored to show how complex tonal dynamics, such as the production and dissolution of stable tonal centers, are supported by agents that are characterized by (i) a tendency to strive toward consonance, (ii) stochasticity, and (iii) a limited memory for previously played notes. TonalEmergence thus provides a grounded computational model to simulate and probe the coordination mechanisms underpinning one of the more remarkable feats of human cognition: collective music improvisation.

Tonal Emergence: An Agent-Based Model of Tonal Coordination

1 Introduction

Coordination is central to human life. Coordination comes in many forms, running the gamut from banal everyday tasks to highly sophisticated domains that require years of training and expertise (Hasson, Ghazanfar, Galantucci, Garrod, & Keysers, 2012; Sebanz, Bekkering, & Knoblich, 2006). Some of our most impressive cognitive abilities are carried out by interacting groups of people — performing music ensembles, teams of surgeons carrying out an operation, scientific collaborators discussing competing hypotheses. Joint action (JA) research has traditionally concerned itself with simple behaviors, such as synchronization, in laboratory studies of highly constrained tasks (Goebl & Palmer, 2009; Hennig, 2014; Konvalinka, Vuust, Roepstorff, & Frith, 2010; Noy, Dekel, & Alon, 2011). But some of the most interesting forms of human coordination are not scripted, but improvised, and entail not just synchrony, but also complementary coordination in support of abstract goals. In recognition of this, there has been growing interest in understanding more complex coordination that occurs in open-ended contexts, such as verbal communication and joint music performance (Hasson & Frith, 2016).

Collective music improvisation has emerged as a fascinating exemplar domain for studying JA in an unconstrained and highly sophisticated setting (Aucouturier & Canonne, 2017; Borgo, 2005; A. E. Walton, Richardson, Langland-Hassan, & Chemero, 2015). A growing body of empirical studies have begun to systematically elucidate the phenomenon of collective musical improvisation — how it is by implicitly guided by shared mental models (Canonne & Garnier, 2011), constrained by genre and performance context (A. E. Walton et al., 2018) and supported by feedback loops of mutual adaptation between interacting musicians (Setzler & Goldstone, 2020). While these experimental studies have examined professional jazz musicians, we would like to emphasize that collaborative improvisation is a vital component of many diverse musical traditions around the world and throughout history, including the raga in Indian classical music, maqam (ubiquitous in Middle Eastern music), European classical music (in the baroque era), and “progressive rock” in the 1970s (Berkowitz, 2010; Kassebaum, 1987; Malvinni, 2013; Solis & Nettl, 2009; Touma, 1971). Furthermore, music improvisation is used in educational and therapeutic settings to stimulate creativity and a foster sense of empowerment and group membership (Koutsoupidou & Hargreaves, 2009; Vougioukalou, Dow, Bradshaw, & Pallant, 2019). Thus, we take jazz improvisation as a model domain to study a much more widespread human behavior which is important in many cultures.

In “free” jazz, improvising ensembles possess the uncanny ability to collectively generate novel musical expressions in real-time, without any written score, prior songform or explicit advance planning. Remarkably, despite the lack of musical constraints, experienced improvising ensembles are capable of producing music that is coherently structured along a number of musical dimensions, namely: rhythm, melody and tonality (harmony). Of these dimensions, *tonal coordination* — the coordination of pitches to produce harmony — constitutes an especially interesting mode of coordination for JA research. Tonal coordination departs from synchronization and sensorimotor coupling, which have dominated joint music performance studies to date (Palmer & Zamm, 2017), because it involves the generation of complementary sets of notes that combine to produce harmony with rich, time-evolving structure.

Given the immense nuance and complexity of musical tonality, attested to by dozens of music theory textbooks and academic courses on the subject (Aldwell, Schachter, & Cadwallader, 2018; Christensen, 2006; Kostka, Payne, & Almén, 2017; Mathieu, 1997)), it is difficult to imagine that it can be generated collectively, in real-time, without an *a priori* song template or explicit advance planning. Yet this is precisely what happens in expert free improvisation. As an example, consider this video (`https://mattsetz.github.io/dissertation-media/`) of a freely improvised duet between two world-class pianists: Craig Taborn and Vijay Iyer. There are several things worth paying attention to as you watch the video. First, Taborn and Iyer are obviously capable of collectively producing coherent, compelling tonal structure. Second, throughout much of the improvisation, tonality is organized in terms of “tonal basins”

—- well defined tonal centers (e.g., C major or F melodic minor) that persist for sustained periods of time (Rush, 2016).[2] Third, tonality is dynamic. Taborn and Iyer sometimes transition between qualitatively different tonal centers; and periods of stable, coherent tonality are interspersed with less structured, transient passages of atonality or "quasi-tonality". Lastly, tonality emerges out of mutual adaptations between Taborn and Iyer. There is not one unambiguous leader; instead, "leadership" seems to be more or less evenly distributed between the two of them.

With respect to this last point, Setzler and Goldstone (2020) presented an empirical study of coordination in pairs of freely improvising jazz pianists, like Taborn and Iyer, which experimentally isolated the effects of "mutual coupling". At a high-level, the study sought to answer the question: how is the music produced by groups of improvising musicians influenced by the presence or absence of mutual adaptations amongst ensemble-members? Musicians were instructed to freely improvise (i.e., without an underlying tune or song structure) in one of two interaction conditions: a "coupled" condition, in which both pianists improvised simultaneously (as in the video with Taborn and Iyer), and a "one-way" condition, in which a single pianist improvised along with a recording produced by an individual in a previous coupled trial[3]. This latter condition occurs in the popular studio recording technique of overdubbing. The two interaction conditions were devised to experimentally isolate mutual coupling between musicians — something that was present in coupled trials where musicians could respond to one another in ongoing feedback loops, but not in one-way trials, where there was only one possible direction of influence (i.e., from the ghost partner to the live musician, but not the other way around).

In a follow-up listener study it was found that naive listeners (with no particular musical experience) preferred music clips produced in coupled trials, despite the fact that they could not guess which condition a clip was produced in above chance-level.

[2] We intentionally use the term "basins" to evoke the idea of a phase space in a dynamical system. This dynamical systems framing of improvising music ensembles is unpacked in the following subsection.

[3] throughout the paper we refer to the static recording in one-way trials as a "ghost partner"

Furthermore, it was found that the presence or absence of mutual coupling systematically constrained the tonal structure produced by co-improvising musicians. A music-theory informed measure of tonal consonance (Chew et al., 2014), which measured the degree to which notes in a given time window formed consonant (stable, pleasant) harmonies or dissonant (tense, clashing) harmony, was applied to music generated in each condition. This analysis revealed that musicians harmonize with previous notes of their partners, resulting in bidirectional tonal coordination in coupled trials (both musicians harmonized with one another's previous notes), and unidirectional coordination in overdubbed trials in (live musicians harmonized with previous notes from the ghost recording, but not vice versa). It was further shown that bidirectional coordination supported more consonant harmonization overall between musicians than overdubbing. This finding is interesting, but we are still lacking a grounded mechanistic account for why this is the case.

Why do mutually adaptive partners achieve greater emergent consonance? And what cognitive mechanisms are necessary to support such elaborate and coherent harmonic coordination in improvising musicians in the first place? That is, how is it that expert jazz musicians come to agree upon tonal centers and time-evolving harmonic progressions during freely improvised performance, without any *a prior* plan? One view might be that collaborating musicians have explicit mental representations about the unfolding harmony of their improvisations (i.e., what key signature they are in, where the music ought to go, based on commonly accepted principles of functional harmony), and come to represent one anothers' mental representations through their mutual interactions. This would allow them to make inferences about future notes played by their partners, guiding their own choice of future notes to produce coherent harmony. Such a perspective has roots in cognitivist, theory of mind accounts of group cognition (Baker, Jara-Ettinger, Saxe, & Tenenbaum, 2017; Khalvati et al., 2019; Zhu, Neubig, & Bisk, 2021).

Alternatively, proponents of the "ecological cognition" paradigm argue that intelligent behavior arises not out of central, symbolic information-processing units (i.e.

individual brains); but instead out of ongoing interactions between an agent's brain, body and environment (which might include other agents) (Chiel & Beer, 1997; Kugler & Turvey, 2015; Richardson, Shockley, Fajen, Riley, & Turvey, 2008). In this dynamical systems framing, complex yet regularized behavior arises out of adaptive, mutual, nonlinear interactions among components of a larger, distributed cognitive system (Beer, 2000; Van Gelder, 1998). Thus, the sophisticated coordination of joint musical performance is seen as emerging out of dynamically unfolding mutual interactions amongst collaborating musicians, as opposed to being mediated by explicit mental representations.

Dynamical systems proponents criticize the cognitivist perspective as taking "loans on intelligence", in the sense that internal representations (e.g., about the harmony of an improvised musical piece) are posited, but often without a mechanistic account of how such representations are learned and implemented in the brain (Richardson et al., 2008). A dynamical systems approach benefits from less "loans on intelligence", in the sense that it doesn't require the assumption of internal representations with unknown origins; instead the external environment serves as its own representation.

When taken at their most extreme, the cognitivist and dynamical frameworks seem irreconcilable. But they need not be mutually exclusive. A less radical view is to recognize that cognition does indeed involve internal, symbolic information processing, but that this processing is constrained and supported by dynamical brain-body-environment interactions. In addition to providing constraints, these interactions can also stabilize intelligent behavior, and representations can be offloaded into the environment. The model presented in this paper is motivated by this "representational light" approach. As will be described, agents do indeed possess internal representations of previously played notes and evolving harmony, and they possess a mechanism for selecting future notes to fit in with evolving harmonies. However, these representations are minimal – there is no explicit representation of a key signature, no explicit rules for how harmonic progressions should proceed over time, and no theory of mind model of their partners' internal representation.

The intention is to study how complex dynamics can arise out of different interactions between agents, and the degree to which we can replicate empirically observed human dynamics with a minimal model. In this spirit, we build on a tradition in dynamical systems approaches to cognitive science of seeing how richly structured behavior can be observed in idealized agent-based models due to different kinds of interactions (Beer, 1995; Candadai, Setzler, Izquierdo, & Froese, 2019). This is not to say that more sophisticated representational processing is not happening in the brains of improvising musicians. But to the extent that we are able to replicate patterns of human behavior with simpler internal models, we can conclude that in principle, this behavior can be achievable without relying on such sophisticated representations.

1.1 A dynamical systems framing of improvised tonal coordination

Throughout this paper, we will be referring to “tonal basins”, by which we mean well-defined tonal centers that persist for sustained periods of time. We intentionally use the term “basins” to evoke the idea of a phase space in a dynamical system, which comprises basins of attraction — regions of phase space that the system is attracted to and that it is unlikely to leave, at least not without some strong perturbation, once it arrives there.

Improvising ensembles can be thought of as dynamical systems, and improvised performances can be understood as trajectories through an underlying musical phase space. This phase space corresponds to a high-dimensional tonal space, where different regions correspond to different tonalities (i.e., well-defined key centers, such as C major).[4] Such areas function as basins of attraction: atonal wanderings often converge on structured tonalities, and once improvisers arrive at a given tonality, they become entrenched within it, as they are more likely to play pitches within as opposed to outside of the tonality (e.g., C and G are more likely to be played in a C major tonal

[4] For the purposes of this model, we can think of tonalities as corresponding to the twelve key centers, such as C major/minor or F major/minor, but the dynamical landscape of real-world improvisation is undoubtedly much more complex, and may comprise a more nuanced set of tonal basins (e.g., C blues scale, F melodic minor, etc).

basin than C# and G#).

We will present analyses that quantify whether music produced by our agent-based model exhibits tonal basins, and the degree to which agents are entrenched in tonal basins. On one end of the spectrum, one can imagine atonal music (random walks through tonal space), which never settles into a tonal basin, and where all notes have an equal probability of being played. On the other end, one can imagine scenarios in which agents arrive at a tonal basin and never leave it, because the probability of playing notes outside the established key center is virtually zero. We refer to agents in the latter scenario as being "deeply entrenched" in a tonal basin. Lastly, there are intermediate scenarios on this spectrum, where agents/musicians establish tonal basins and then transition out of them after some period of time – either to a new tonal basin or to atonal wandering. This intermediate, "shifting basins" dynamic is an important source of suspense in improvised music, as can be observed in the Iyer/Taborn video. As we will demonstrate, the dynamical landscape of our model is shaped by interactions between parameters tuning agent memory, entropy in note generation, and the structure of agent coordination.

1.2 The present study

Here we present *TonalEmergence*, an agent-based computational model of tonal coordination, which can help us gain traction on the mechanisms supporting harmonic coordination in improvising musicians. The model is formulated to simulate the experimental conditions of Setzler and Goldstone (2020) (i.e., "coupled" and "one-way"), and to isolate the phenomena of tonal coordination from other musical dimensions that are unconstrained in free improvisation (e.g., rhythm, polyphony, texture/loudness). At each time-step, agents generate a single note biased towards maximizing tonal consonance with their partners' previous notes. Agents are paired up in precisely the same interaction conditions in which human improvisers were paired in the empirical study. Insofar as we can reproduce some of the empirical findings using this computational model, it provides a plausible mechanistic account for tonal coordination in humans.

Furthermore, once given some empirical validation, the intrinsic dynamics of *TotalEmergence* are worth studying for what they reveal about factors that affect coordination and (in)stability in a richly structured behavioral space. The simplicity of the model allows us to go beyond the kinds of analyses reported in Setzler and Goldstone (2020), and to explore other aspects of emergent tonal structure that are difficult to operationalize in naturalistic music, which is complicated by rhythmic variability. Recall the performance of Taborn and Iyer, in which they spontaneously converge on tonal basins. What conditions are necessary to support the emergence of tonal basins in collectively improvising agents? Can agents transition between different tonal basins without any top-down directive to do so? Lastly, how does mutual coupling figure into agents' ability to produce emergent tonal basins, and transition between them?

With regard to this last question, one hypothesis would be that coupling increases agents' propensity to converge on entrenched tonal basins and remain stuck in them for long periods of time. When coupled agents arrive at a tonal basin, each agent is tethered to that basin by the histories of each other's previous notes, and these two note histories might function as a sort of tonal reservoir, anchoring agents to the present tonality. By contrast, in one-way settings, only one agent is responsive to the previous notes of their partner, so the effective size of this tonal reservoir is reduced, and there may be less of a stabilizing force tethering the dyad to the current tonality. Alternatively, it could be that it is difficult to transition between disparate tonal basins *without* mutual coupling, because ghost agents in the one-way condition have no opportunity to respond to outlier notes (i.e., notes falling outside an established tonal basin) played by live agents. In this case, instead of initiating a tonal transition, outlier notes would simply "muddy the waters" and produce dissonance within an unchanging tonal basin.

It is worth emphasizing that this agent-based model is *idealized.* As will be specified, agents are governed by simple rules[5] for generating notes, and their behavior

[5] From a dynamical systems perspective, these "rules" can be thought of as constraints.

is determined by two parameters with intuitive real-world interpretations: memory (how far back in time agents are influenced by their partner's notes) and entropy (degree of randomness in note generation). In this respect, the TonalEmergence model is different in intent from other generative music models that have been gaining popularity in recent years. As part of the deep learning revolution, there has been a proliferation of deep neural network models of music generation (Briot, Hadjeres, & Pachet, 2017; Huang et al., 2019; Oore, Simon, Dieleman, Eck, & Simonyan, 2020); these models are typically trained on large music data sets to predict continuations of musical sequences. While they do a hauntingly good job at generating naturalistic music, these models are typically not intended as mechanistic models of human performance, and they comprise thousands of parameters with no obvious real-world interpretation.

There is no shortage of bells and whistles that could be added to make TonalEmergence more complex and produce more human-like music. But simplicity is a desired property in this model. Our goal here is not to produce music matching the sophistication and nuance of human performance, but rather to isolate and study properties of tonal coordination in a grounded model built on minimal assumptions. In this vein, we use TonalEmergence to examine the minimal conditions necessary to replicate some of the findings of Setzler and Goldstone (2020) by qualitatively comparing trends between the model's behavior and empirically observed human behavior. It also serves as an analytically tractable "toy universe" with which to explore necessary conditions for emergent and possibly shifting tonal basins. Furthermore, insofar as this model is idealized, it can be used as a tool to think about collective phenomena and emergent dynamics in complex systems more generally.

In this spirit, the aims of the present study are twofold: (1) validate TonalEmergence against empirical results, and (2) study behavior of TonalEmergence as its own system. With respect to the former aim: can we reproduce the result that bidirectional, relative to one-way, tonal coordination supports greater emergent consonance in TonalEmergence? What are minimal conditions necessary to do so? Answering these questions will help furnish a plausible mechanistic account of human

JA. With respect to the latter aim: what conditions are necessary to support the emergence of potentially shifting tonal basins? What are the pressures supporting the emergence of tonal basins and transitions between them? And how do the effects of mutual coupling influence how tonality evolves over time?

In the remainder of this paper we provide a fully specified description of TonalEmergence, and report results of experiments simulating TonalEmergence in various parameterizations (of memory and entropy), under the same interaction conditions (coupled versus overdubbed) implemented in Setzler and Goldstone (2020). We conclude by evaluating the merits of this model as a mechanistic account for tonal coordination in co-improvising humans, and discuss connections with related work in complex systems and computer music.

2 Methods

2.1 Model

The model consists of interacting dyads of agents. As depicted in Figure 1a, each agent plays one note (one of the twelve pitches in the chromatic scale) at each step of simulated time. Initial notes are randomly seeded, but as simulations progress, agents infer probability distributions across the twelve pitches, which are biased towards maximizing consonance given their partner's previous notes. (Consonance is evaluated with a measure adapted from the Tonal Spiral Array model (Chew et al., 2014), as specified in Setzler and Goldstone (2020); this is explained further in the Measures of Tonality subsection.) These distributions are independently generated by each agent at every time step, and then sampled from to yield the next note.

There is also a temporal decay in the agents' memory, such that notes from the distant past are weighted as exponentially less important than notes from the recent past. Our decision to incorporate a memory parameter is motivated by past experimental and computational studies demonstrating that working memory is an important factor in improvised musical performance (De Dreu, Nijstad, Baas, Wolsink, & Roskes, 2012; Johnson-Laird, 2002). Modeling memory with an the exponential decay is a longstanding practice in cognitive science (Cowan, 2001; Wickelgren & Norman, 1966); and follows from the minimal assumption that the probability of forgetting is constant at any given moment (e.g., if you have a 10% chance of forgetting, or not being influenced by a past note at any given time, then you get an exponential decay of memory over time in the past).

More concretely, at a given time step, agents compute attention weights for each note in their partner's history according to the equation:

$$attention\_weight = e^{\frac{-m}{\tau}} \tag{1}$$

where m is number of time steps prior to the current time step and τ is a parameter that determines how rapidly attention decays over time (larger τ values correspond to longer-term memory).

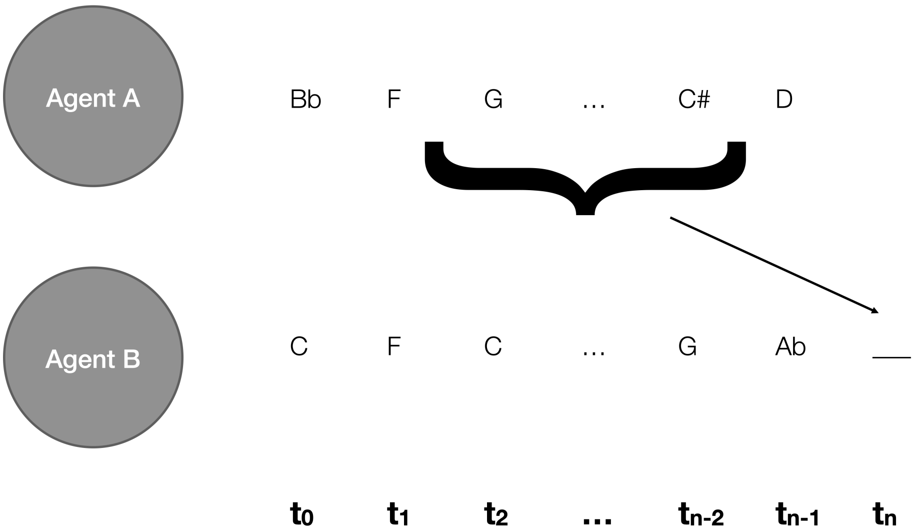
Attention weights are then used to construct an attention-weighted histogram of how frequently each of the twelve pitches occurred (see Figure 1b). This is achieved by summing attention weights for every occurrence of each pitch class in an agent's partner's history. Given this histogram, hypothetical consonance scores[6] are assigned to all of the twelve pitches, by separately measuring the consonance of the histogram with each pitch (weighted by 1). This provides a vector of twelve hypothetical consonance scores, which are then fed through a softmax function, defined below, to yield a well-defined probability distribution across the twelve pitches, biased towards maximizing consonance. The softmax function is defined below:

$$softmax(x)_i = \frac{e^{\gamma x_i}}{\sum_j e^{\gamma x_j}} \tag{2}$$

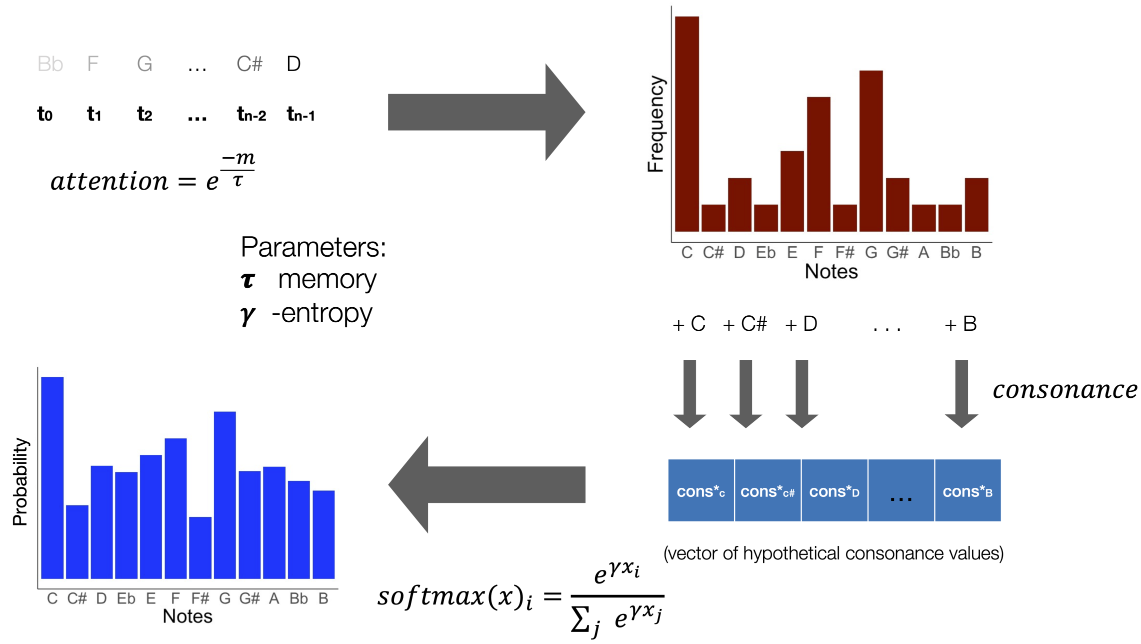
where x_i is the consonance score of pitch i, and γ is a parameter that tunes how entropic the distribution is.

Low γ values yield flatter, more entropic distributions ($\gamma = 0$ produces a completely flat distribution), whereas high γ values produce distributions more sharply

[6] These are *hypothetical* consonance scores because they correspond to the consonance that would result from an agent playing a given pitch in the next time step, given their partner's previous notes.



(a) Interacting agents produce one note at every time step. Note selection is biased towards maximizing consonance given an agent's partners' previous notes. Agent memory decays exponentially over time, represented by the diminishing opacity of Agent A's notes from the perspective of Agent B.



(b) Note generation. At each time step, agents infer probability distributions across the twelve pitches, biased towards maximizing consonance given their partners' previous notes. First, agents create an attention-weighted frequency histogram of how often notes occurred in their partners' history. Next, hypothetical consonance scores are computed for each pitch, using the attention-weighted histogram. Lastly, this vector of consonance scores is fed through a softmax function, to obtain a probability distribution that is sampled from to yield the next note.

Figure 1. Model overview.

biased towards consonance-producing pitches (sufficiently high γ values approach determinism, where the note maximizing consonance is always sampled). Extreme values of γ correspond to degenerate cases: extremely low values produce random walks through tonal space, while extremely high values produce scenarios where agents converge on single note, played in unison throughout the entire simulation. We are primarily interested in a "sweet spot" of γ values that support interesting interactions. Our inclusion of γ is motivated by past experimental studies which have demonstrated that entropy in the performance of improvised (versus rehearsed) musical sequences predicts listener preferences and perception of spontaneity (Engel & Keller, 2011; Keller, Weber, & Engel, 2011; Pearce & Wiggins, 2012; Zeng, Przysinda, Pfeifer, Arkin, & Loui, 2017). In addition to representing potentially aesthetically-motivated decisions, entropy also represents the noisiness in a musician's cognitive-motor system controlling their instrumental performance. The use of softmax in cognitive science settings originated in reinforcement learning and is now ubiquitous in deep learning applications (Goodfellow, Bengio, & Courville, 2016; Sutton & Barto, 2018).

2.2 Simulations

Simulations were run using the same interaction conditions (i.e., coupled versus one-way) and yoked design as were used in the human experiment reported in Setzler and Goldstone (2020). In *coupled* trials, two agents were instantiated. At every time step, each agent generated a note and observed the note generated by its partner. This resulted in two note sequences per performance, one for each agent. Subsequently, individual note sequences from each *coupled* trial provided the recorded sequence for future *one-way* trials. This procedure is depicted in Figure 2.

In *one-way* trials, a single agent played along with a "ghost partner" – a note sequence generated by an individual agent from a previous *coupled* trial. At every time step, the agent generated a note (which was not heard by the "ghost" partner) and heard the note played by its "ghost partner" at the corresponding time step in its pre-generated note sequence. Each coupled trial yoked a corresponding *one-way* trial

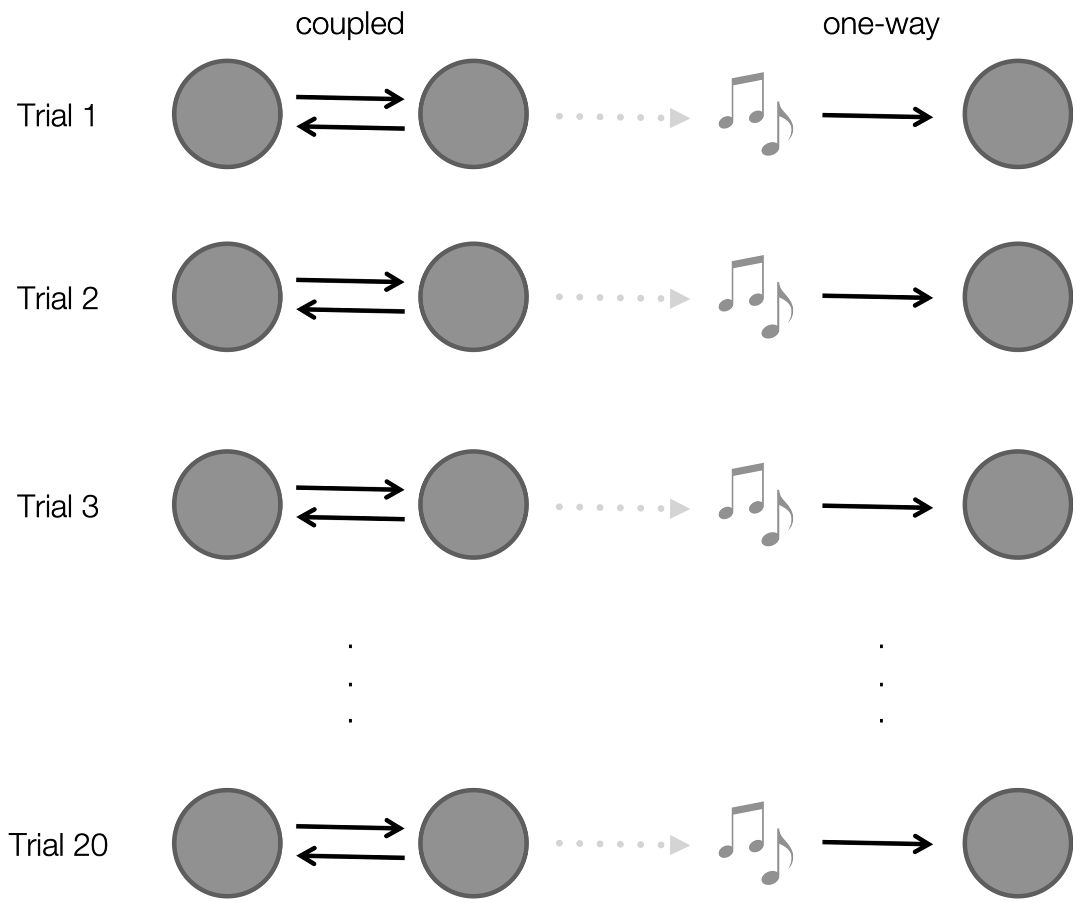


Figure 2. Procedure for simulating agents in coupled and one-way conditions. Twenty coupled trials were first simulated. Subsequently, an individual note sequence (arbitrarily chosen from the two agents) from each coupled trial yoked a corresponding one-way trial.

(the individual note sequence selected from each coupled trial was arbitrarily selected from the two agents, and the other one was not used in one-way trials).

All simulations lasted 500 time steps. In general, 20 trials were run in each condition for every experiment (additional trials were run in certain cases, where a larger sample was necessary to elucidate a systematic trend; these cases are reported in turn). Simulations results are hosted at `https://osf.io/93qzp/` (Setzler & Goldstone, 2021). The code for our model implementation and for running these simulations, which can be used to reproduce these results, can be found in the GitHub repository: `https://github.com/mattsetz/TonalEmergence`. Audio recordings of model output for exemplar trials can be accessed here: `https://mattsetz.github.io/dissertation-media/`.

2.3 Measures of Tonality

The same tonal-consonance measure, defined in Setzler and Goldstone (2020) and used to evaluate degrees of consonance produced by human musicians in the empirical study, was used here. In brief, the measure assigns consonance scores to different intervals. For example, C and G produce a perfect fifth, which is highly consonant, whereas C and F# produce a tritone, which is highly dissonant. Consonance of any arbitrary window of music is a weighted sum of consonance scores for each interval, based on how often those intervals occur. In the present work, consonance is normalized to range between 0 and 1, where 0 represents the lowest consonance level (corresponding to a tritone) and 1 represents the highest consonance level (corresponding to a unison).

As in the empirical study, consonance time series were obtained from note sequences by computing consonance over a sliding window; here we used a window size of five time steps and a hop size of one time step. Individual Consonance (IC) time series were obtained by evaluating consonance across note sequences produced by individual musicians, and Combined Consonance (CC) was computed by merging note sequences of both individual musicians in a trial into a single merged note sequence, and evaluating consonance throughout this merged note sequence.

Lastly, as in the empirical study, Emergent Consonance (EC) was computed as CC minus average IC for a given window. EC measures the consonance arising from the interaction of pitches played by collaborating musicians.[7] The theoretical range of EC is from -1 (minimal EC) to 1 (maximal EC), where 0 represents a situation in which Combined Consonance is equal to average Individual Consonance. This being said, we are less interested in the absolute values of EC, and more interested in how EC varies as

[7] As described in Setzler and Goldstone (2020), "A situation in which each pianist plays self-consonant notes that clash with one another would result in low EC (e.g., C, E, G and F#, A#, C# are consonant on their own but C, E, G, F#, A#, C# is highly dissonant), whereas a situation in which each pianist plays dissonant notes that stabilize one another when sounded together would result in high EC (e.g., C, B and E, G have low average consonance but C, E, G, B has high consonance because it is tonicized to a Cmaj7 chord)."

a function of interaction condition.

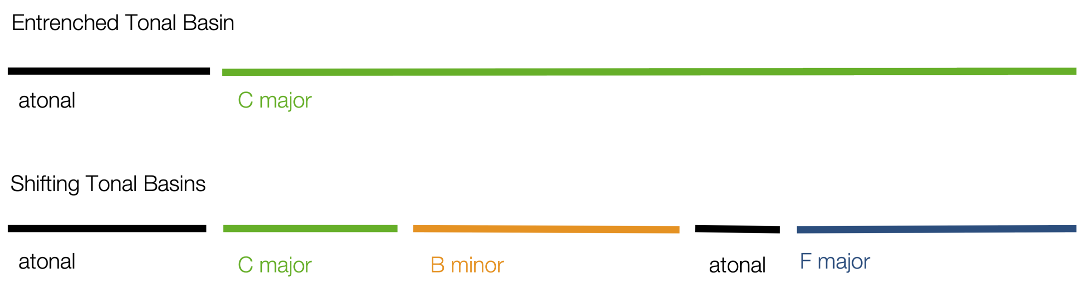
We refer readers to the Supporting Information, and Setzler and Goldstone (2020) for full specification of these measures. Two additional measures were also used in this study: Gapped Consonance and Tonal Novelty, which are described below.

2.3.1 Gapped Consonance. Gapped Consonance was developed to assess the depth of tonal basins (i.e., how robust and long-lasting tonal basins are) in note sequences. Gapped Consonance is essentially a version of autocorrelation, utilizing the consonance measure as a sort of similarity metric. As depicted in Figure 3, consonance is evaluated *between* the notes of two windows (each five time-steps wide) separated by a particular gap length (time duration), as expressed in the equation:

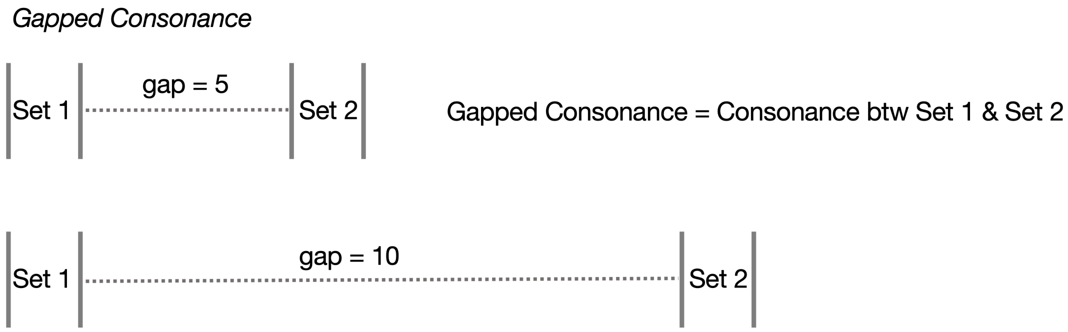
$$GC_{t,gap} = consonance_{btw}(notes_t, notes_{t-gap}) \tag{3}$$

where $GC_{t,gap}$ denotes Gapped Consonance in a particular piece at time t and gap duration gap. For the purpose of the simulation, gaps are expressed in time-steps, but they could be expressed in seconds when examining real-world musical sequences. $notes_t$ denotes the set of notes played by both agents at time t, expressed as a histogram representing how often each pitch was played during a 5-step window (i.e., the frequency count of a given pitch is equal to the number of time-steps that pitch is played during the window). $consonance_{btw}(notes_i, notes_j)$ denotes consonance between two sets of notes, and is computed as the weighted sum of consonance ratings for each interval scaled by how often particular intervals occurred between notes played by the two agents. See the Supporting Information for a formal definition of $consonance_{btw}$.

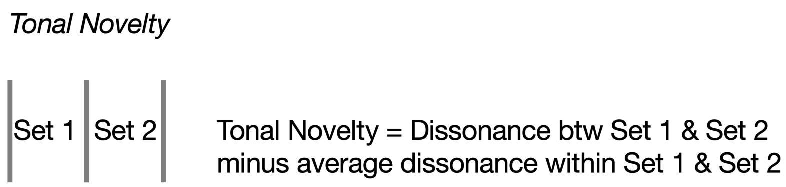
For a given gap value, Gapped Consonance was computed at every time point in the note sequence, and these values were averaged to obtain an overall mean Gapped Consonance at each gap for every trial. This was done over a range of gap sizes. Evaluating Gapped Consonance across a range of gap sizes allows us to measure tonal coherence over a range of timescales. In note sequences comprising long-lasting tonal basins, Gapped Consonance will be relatively high and stable across gap size, because notes played at given time point will be in the same tonality as notes played in the



(a) Toy illustration of two hypothetical performances. In the upper performance, agents converge on a stable tonal basin, C major, and are stuck in that basin indefinitely. In the lower performance, agents converge on the same tonal basin, but subsequently transition to different tonalities, which persist for limited periods of time.



(b) Gapped consonance measures tonal coherence across a range of timescales. It can thus be used to infer the degree to which note sequences comprise robust tonal basins. For example, gapped consonance would be high for small and large gap sizes in the upper performance in (a), which consists of one long-lasting tonal basin. But gapped consonance would decrease with increasing gap sizes in the lower performance, because there is high local tonal coherence (i.e., within basins), but less coherence over longer timescales (i.e., because they span distinct basins).



(c) Tonal novelty measures the degree to which there is a marked change in tonality from one time period to the next. Novelty would be relatively low throughout the upper performance in (a), but high in the lower performance at transitions between different basins.

Figure 3. Operationalizing tonal dynamics.

recent past, and far into the past. By contrast, in note sequences with shifting tonal basins, Gapped Consonance will decrease with increasing gap sizes, because notes played at a given time point will be in a similar tonality to notes played in the recent past, but in a different tonality from from notes played further in the past.

Like Combined Consonance, Gapped Consonance ranges from 0 (minimally consonant) to 1 (maximally consonant), although in general we are less concerned with the absolute value of Gapped Consonance, and more interested in how it varies over gap sizes, as described above. In the following analyses, Gapped Consonance was computed for combined note sequences (i.e., note sequences from both agents in a duo were merged into one), but in principle it could be computed over individual note sequences as well.

2.3.2 Tonal Novelty. Tonal Novelty is inspired by the novelty function in signal processing/audio segmentation put forth by Foote (2000). Here it is used to measure the degree to which tonality changes from one period to the next. As depicted in Figure 3, two adjacent windows (each five time steps wide) are positioned to straddle an arbitrary time-point. Tonal Novelty at that time-point is computed as the dissonance (negative of consonance) between notes in each window minus the dissonance within each window, as expressed in the equation:

$$Novelty_t = diss_{btw}(notes_t, notes_{t+5}) - (diss_{within}(notes_t) + diss_{within}(notes_{t+5})) \quad (4)$$

where $Novelty_t$ denotes Tonal Novelty at time t, $diss_{btw}(notes_i, notes_j)$ denotes *dissonance* between two note sets, and $diss_{within}(notes_i)$ denotes dissonance within a note set. Dissonance is the opposite of consonance and defined as $dissonance = 1 - consonance$. Again, see Supporting Information for full specification of these dissonance terms.

The rationale behind this measure is that Tonal Novelty is high to the degree that (1) music in two successive windows is sufficiently different (i.e., dissonance between them) and (2) music within those windows is relatively homogeneous, or falls within a given tonality (i.e., negative dissonance within them). The second term in Equation 4 is necessary to account for situations of atonality; we don't want to assign high novelty in these cases, because the music is not moving from one well-defined tonality to another. If both windows in question consist of randomly distributed notes (i.e., atonality), there would be high dissonance *between* them but also high dissonance *within* them, so overall novelty would be relatively low.[8] Time series of Tonal Novelty were computed on combined note sequences from each trial.

3 Results

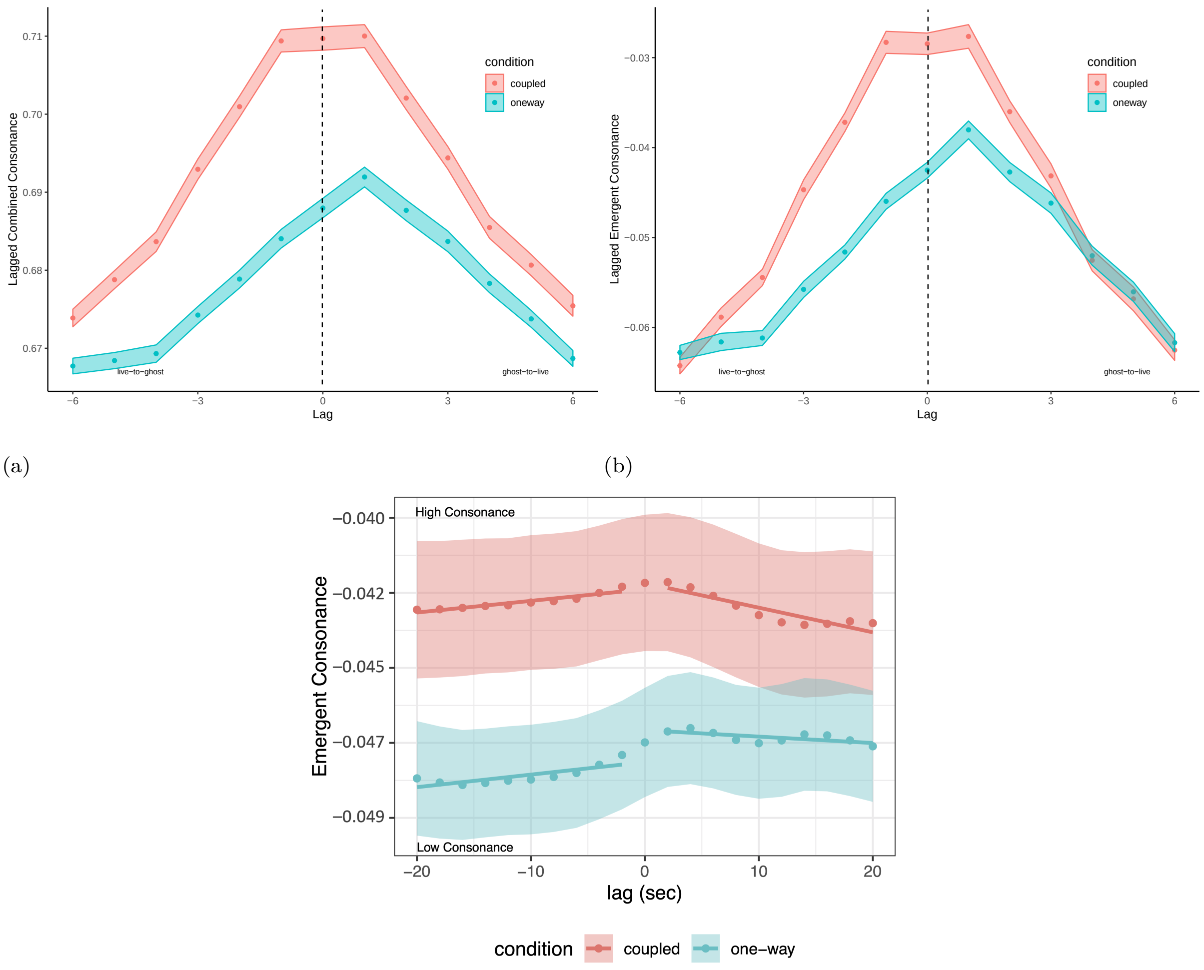
3.1 Mutual coupling supports greater consonance (replication of empirical findings)

In Setzler and Goldstone (2020), a lagged-consonance analysis revealed that human improvisers tend to harmonize with the preceding notes of their partners. This resulted in bidirectional coordination in coupled duos, but in asymmetric, unidirectional coordination in one-way duos, where a live musician harmonized with the preceding notes of their ghost partner, but not vice versa. It was further found that bidirectional coordination supported more consonant harmonization between the notes of mutually adaptive partners, as Emergent Consonance was higher overall in coupled than in one-way duos (Setzler & Goldstone, 2020).

Our first objective here was to see if the simplest version of the TonalEmergence model could replicate these empirical findings. Accordingly, coupled and one-way conditions were simulated with agents parameterized with a relatively low level of entropy (γ=5) and the shortest conceivable memory span: τ=0.1, which effectively meant that agents only had access to their partners' immediately preceding note. 20 trials of 500 time steps were simulated for each condition. We then performed the same lagged-consonance analysis as in Setzler and Goldstone (2020) on the resulting note sequences. Results are depicted in Figure 4.

Lagged consonance of the model simulations (Figure 4b) replicate several important patterns that were empirically observed in the experiment with human jazz musicians (Figure 4c). In both the cases, lagged consonance reveals symmetric,

[8] In theory, Tonal Novelty can range from -1 (minimum novelty) to 1 (maximum novelty), where 0 represents a situation in which dissonance between the notes in adjacent windows is equal to dissonance within each window, and negative values reflect situations in which there is more dissonance within adjacent windows than between them. In general negative values of Tonal Novelty are very rare.



(a) (b)

(c)

Figure 4. Mutual coupling supports greater consonance. Points depict average lagged consonance across all trials in each condition; error bars denote standard error of the mean. Results from model simulations are depicted in (a) and (b). c) depicts results from empirical study with professional jazz pianists (re-printed here from Setzler and Goldstone (2020), the y-axis values have been changed to normalized Emergent Consonance, as described in the Methods.). Lagged Combined Consonance is depicted in (a), and Lagged Emergent Consonance in (b). For both measures, simultaneous consonance is higher in *coupled* trials. Additionally, consonance is symmetric around 0 in *coupled* trials, but higher at positive lags (ghost-to-live; representing consonance of live musicians notes against previous notes from the ghost partner in *one-way* trials) versus negative lags (live-to-ghost; vice versa) for *one-way* trials. This reflects agent coordination inherent to each condition — bidirectional in *coupled* trials and unidirectional in *one-way* trials.

bidirectional tonal coordination in coupled agents (red consonance values are symmetric around zero), but unidirectional coordination in overdubbed trials, in which live agents harmonize with preceding notes from the ghost recording but not vice versa (blue consonance values are higher at positive lags). Additionally, in both the model simulation and human study, we found that bidirectional coordination supports higher emergent consonance overall (red consonance values are higher than blue values). The former result (i.e., asymmetric lagged consonance in *one-way* trials) was expected for the model, because agents were explicitly programmed to harmonize with their partner's preceding notes. However, it was not at all clear to us *a priori* that mutual coupling would support higher consonance in the model, as it did with human participants. Indeed, it is noteworthy that this result was obtained in even the simplest version of TonalEmergence, in which agents only had the capacity to respond to the immediately preceding note played by their partners.

The negative Emergent Consonance (EC) values in Figures 4b and 4c indicate that, in both the empirical study and model simulations, there was more Combined Consonance than Individual Consonance. This being said, more negative values indicate low EC and less negative values indicate higher EC. The difference in absolute values on the y-axis between Figure 4b and Figure 4c indicate that human musicians in the empirical study produced a more narrow range of Emergent Consonance compared to model simulations. However, this was true for both conditions and across all lags, and does not speak to our deeper inquiry. We are less interested in exactly matching absolute Emergent Consonance values from the empirical study, and more interested in replicating robust qualitative trends — such as unidirectional tonal coordination in *one-way* trials, and increased Emergent Consonance in *coupled* trials.

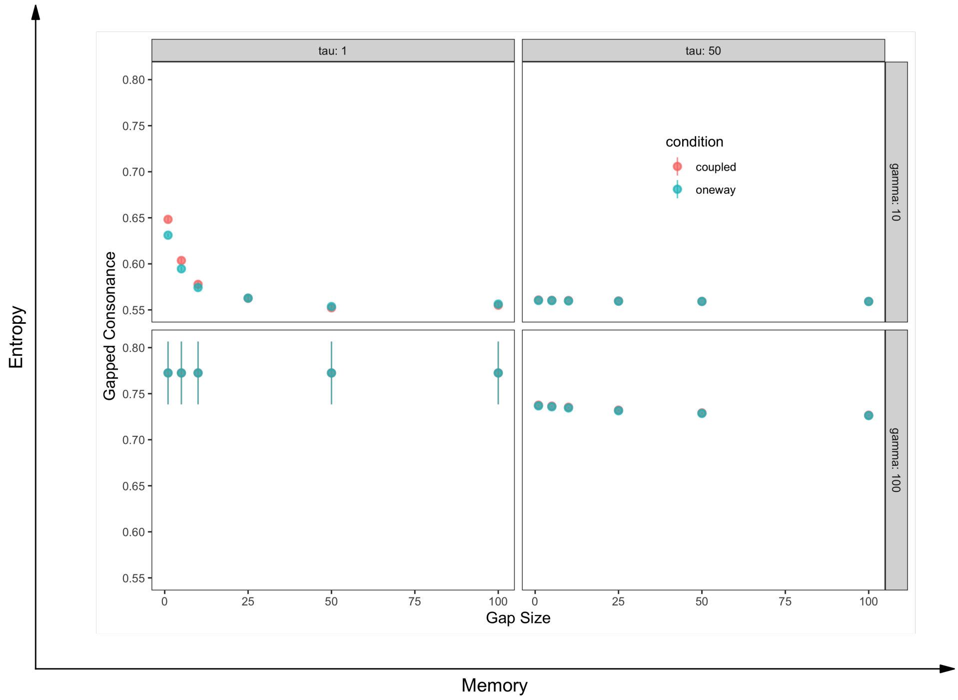
3.2 Emergence of tonal basins

It is remarkable that some of our empirical findings can be reproduced in such a simple model, parameterized at γ=5 and τ=0.1, in which agents have access only to the previous note played by their partners. But there are limitations to this parameterization; namely, it does not support the emergence of tonal basins. The music produced by this model sounds more or less like a random walk through tonal space, which is more biased to comprise consonant intervals in coupled trials. Such atonal drift occurs in free improvised performances, but as can be observed in the Taborn/Iyer video, human improvisers also tend to converge on *tonal basins*: structured tonal centers (e.g., C major, F melodic minor) that cohere over sustained periods of time. Additionally, human improvisers have the ability to transition to different basins over time, as is depicted in Figure 3a.

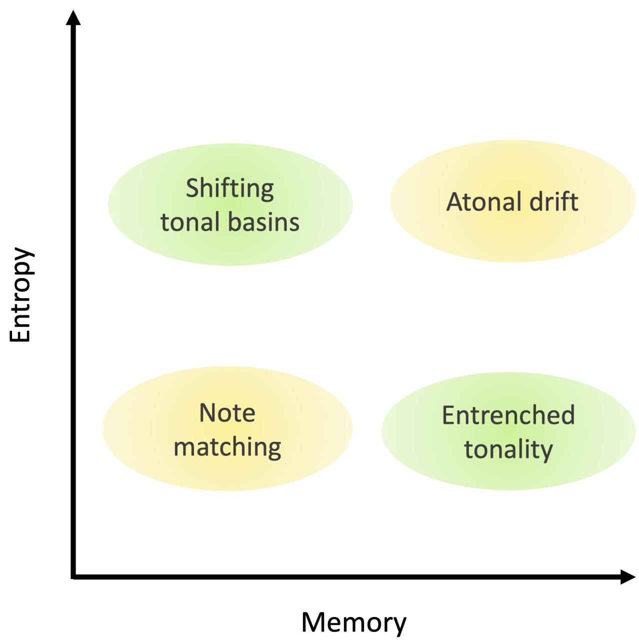
What else is needed for our model to support the emergence of shifting tonal basins? How do the effects of mutual coupling change in models that exhibit these different kinds of tonal dynamics? The former question can be broken down into two sub-questions. First, what is required to support the emergence of tonal basins? Second, what conditions are necessary to support the emergence of tonal basins and changes/transitions between them over time? Transitioning between tonal basins requires the existence of basins, but not the other way around — it is easy to imagine agents capable of arriving at a tonal basin, but being incapable of transitioning out of the basin once they are there.

To address these questions, behavior of the model was contrasted under four parameter settings: (τ=1, γ=10), (τ=1, γ=100), (τ=50, γ=10) and (τ=50, γ=100). This coarse-grained parameter sweep facilitated an examination of model behavior under for combinations of memory (low or high) and entropy (low or high). For each parameterization, 20 simulations of 500 time steps were run in each interaction condition (i.e., *coupled* versus *one-way*). Gapped Consonance was then computed on resulting note sequences. Results are depicted in Figure 5, and Table 1 provides a qualitative summary of the model's behavior at each parameter setting. These qualitative dynamics can be observed in sample audio recordings of the model at each parameterization and condition, which can be found here `https://mattsetz.github.io/dissertation-media/`.

Mean Gapped Consonance is highest at (τ=1, γ=100), which also has the most variability of consonance across trials. In this parameterization of low memory and low entropy, agents are effectively forced to match the immediately preceding note of their partners, which results in a degenerate dynamic: agents simply alternate between two



(a)



(b)

Figure 5. Emergence of (shifting) tonal basins is supported by “sweet spots” in (*entropy*, *memory*) parameter space. (a) Gapped Consonance analysis for both interaction conditions in four model parameterizations. (b) Toy diagram illustrating outcome of the Gapped Consonance analysis. Yellow sections represent regions of parameter space resulting in degenerate dynamics, whereas green sections represent regions that support the emergence of tonal basins.

Parameterization	Memory	Entropy	Dynamics	Effect of Condition
(τ=1, γ=100)	low	low	Degenerate note matching.	None.
(τ=1, γ=10)	low	high	Shifting tonal basins.	coupled > oneway at local timescales
(τ=50, γ=10)	high	high	Atonal drift.	None.
(τ=50, γ=100)	high	low	Entrenched tonal basins.	coupled > one-way; small, but constant

Table 1

Summary of Gapped Consonance over 2×2 parameter sweep. These trends can be observed in sample audio recordings of model at each parameterization and condition `https://mattsetz.github.io/dissertation-media/`*.*

notes – whichever notes agents randomly generated at the first time step – forever. In dynamical systems terms, this is akin to a limit cycle, where agents take on the same state every other time step. Given this dynamic, consonance is entirely determined by the interval between the initially generated notes, which explains the high variance across trials. Consonance is invariant to gap size, because of this fixed limit cycle, tonality never changes throughout improvisations. Lastly there is no effect of condition in this parameterization.

We can see how agents with the same entropy and increased memory behave by analyzing (τ=50, γ=100). In this case mean Gapped Consonance is a bit lower, but still quite high relative to the other parameterizations. Consonance decreases slightly with larger gaps, but still remains relatively high and steady across gap sizes. Lastly, there is significantly less variability in consonance across trials than for (τ=1, γ=100). With low entropy, agents are still heavily biased towards generating notes that maximize consonance, but by virtue of their increased memory, agents are no longer stuck in the degenerate limit cycle described above. Since agents are no longer confined to matching the immediately preceding note of their partner, they instead generate notes that maximize harmony given a long history of their partner's previous notes.

From random initial conditions, agents converge on a tonal basin within the early stages of a simulation, and remain in this basin indefinitely. This basin is more open-ended than the previously described limit cycles, as agents play different notes within (and occasionally outside of) the tonal center. Consonance decreases slightly over larger gaps because of a compounding effect. With some (small) degree of randomness, agents play notes outside of the tonal basin. As time progresses, these outlier notes flatten the histogram of each agent's note history, which further reinforces future outlier notes, in a self-reinforcing cycle. Lastly, there is a small but significant effect of condition (coupled trials produce more consonance), which is robust over gap size. This effect is difficult to see in Figure 5, but is evident in Figure 6.

The setting (τ=50, γ=10) is the most entropic parameterization of the four compared here. These agents have the same long memory as the previously discussed agents, but the decreased γ results in more entropic note generation. And as was just described, this entropy is effectively compounded over time, since flatter pitch histograms result in more entropic note generation, which in turn results in even flatter pitch histograms. As a consequence, consonance is lowest at this parameterization. As in the other parameterizations, consonance is relatively constant over gap size.

This brings us to (τ=1, γ=10); a low-memory, high-entropy agent. This parameterization stands out amongst the four discussed here, because it is the only one in which consonance decreases markedly with increasing gap size. Consonance appears to exponentially decay with gap size, eventually saturating at a low consonance level for sufficiently high gap sizes. Despite this decrease, however, consonance remains relatively high for gap sizes up to 10 (compared to entropic agents). Given these observations, (τ=1, γ=10) appears to support *shifting* tonal basins. In contrast to the previous parameterizations, tonality changes over time, in such a way that local tonal coherence is preserved at small timescales. Another trend to notice here is that there is an interaction between condition and gap size. Consonance is higher in mutually coupled agents at small gaps, but as gap size increases, this effect disappears (i.e., consonance for coupled and one-way agents converges). Thus we see that mutual coupling promotes local tonal coherence (i.e., at small timescales), but not at larger timescales.

In summary, different combinations of τ and γ support different tonal dynamics. Certain parameterizations produce degenerate dynamics, as in the case of (τ=1, γ=100). Others support emergence of complex tonal dynamics, such as in (τ=1, γ=10), which supports shifting tonal basins. The next section will dive deeper into how mutual coupling influences emergent tonal dynamics.

3.3 Interactions between mutual coupling and entrenched versus shifting tonal dynamics

In this section we focus on (τ=50, γ=100) and (τ=1, γ=10) because, out of the above four, they produced the most interesting dynamics. In (τ=50, γ=100), agents quickly converged on entrenched tonal basins that lasted indefinitely. Agents were biased towards playing notes within tonal basins, but still had a enough freedom to play occasional outlier notes, without ever breaking away from the established global tonal basin. (τ=1, γ=10) exhibited "shifting" tonal basins; agents converged on coherent basins that persisted for prolonged periods of time, but they weren't stuck in those basins forever; instead they were free to transition between different tonalities throughout the course of a simulated piece. In what follows, we analyze how the effects of interaction condition vary between these two parameterizations in order to see how mutual coupling constrains tonal dynamics in agents that support more interesting tonal dynamics (i.e., shifting tonal basins).

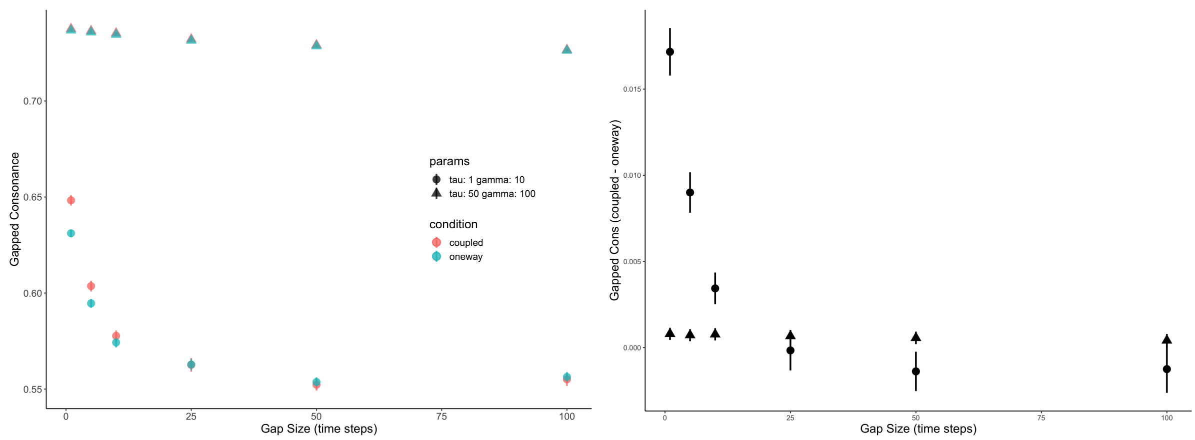
Figure 6 shows Gapped Consonance for just these two parameterizations. This is the same information as was presented in Figure 5a, but with reduced visual clutter, so that the effects of condition across different gap sizes are more salient. For (τ=1, γ=10), there is initially a large effect of condition, such that coupled agents produce more consonance than overdubbed agents, but this effect gradually shrinks as gap size increases.[9] In other words, mutual coupling supports more local tonal coherence across

[9] The effect of condition decreases to 0 at gap=25, and then continues to decrease below zero for larger gap sizes. In other words, at sufficiently large gaps, Gapped Consonance is *higher* in one-way trials.

short timescales, but not at larger timescales. The setting (τ=50, γ=100) also exhibits an effect of condition in the same direction (i.e., mutual coupling supports more consonance), but the effect is much smaller than that observed for small gaps in (τ=1, γ=10), and it is robust across all gap sizes. Because the effect is small, values for coupled trials are hidden behind the blue values for one-way trials in Figure 6a, but the effect is evident in Figure 6b, which shows that the *difference* in consonance between correspondingly yoked coupled minus one-way trials is consistently positive across all gap sizes. In other words, for this parameterization, which produced entrenched tonal basins spanning entire simulations, mutual coupling supported greater consonance across all timescales.

What do we make of the finding that mutual coupling supports greater local tonal coherence in agents producing shifting tonal basins, and more global coherence in agents that produce entrenched tonal basins? One interpretation is that mutual coupling appears to promote consonance *within* tonal basins, but not across disparate basins. For agents producing shifting basins, there is a large effect of mutual coupling at local timescales, because these timescales are likely to encompass a unified basin. But given that tonality changes over time, larger gaps are likely to encompass disparate basins, and when this occurs, the effect of mutual coupling begins to vanish. In contrast, tonality is static when agents produce entrenched tonal basins that span the entire simulation, and in these cases the effect of mutual coupling is insensitive to gap size. This will be further unpacked in the Discussion.

This could be a side-effect of the fact that mutual coupling supports greater consonance at small timescales. Because of this, within small timescales, notes produced by coupled agents are more tightly clustered around a given tonal basin. But since tonality changes over time, at large gaps these tight note clusters will be evaluated against notes clustered around a different tonality, which could result in high dissonance. In contrast, since there is less local tonal coherence in one-way trials, Gapped Consonance at large gaps more resembles consonance for random distributions of pitches (as opposed to distinct pitch clusters).



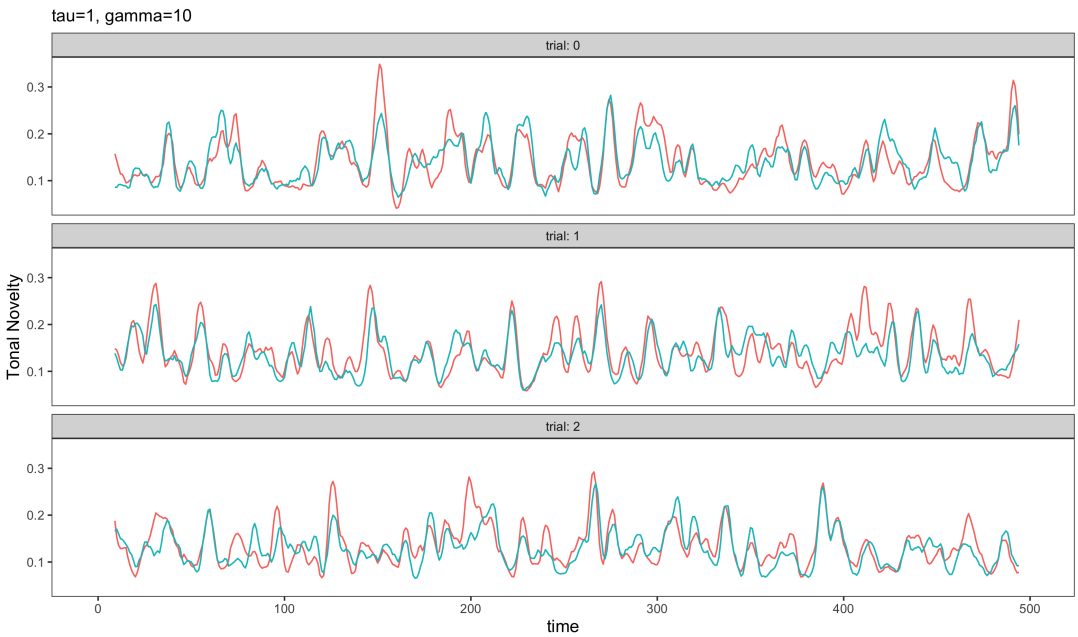
(a) (b)

Figure 6. Interactions between coordination condition, gap size, and memory/entropy combinations. Points in (a) represent mean Gapped Consonance across all trials in each condition. 20 trials per condition were simulated with (τ=1, γ=10). 500 trials per condition were simulated with (τ=50, γ=100), because additional statistical power was needed to confirm a main effect of condition. Coupled trials are shown in red, one-way in blue. Circles denote averages for (τ=1, γ=10) parameterizations and triangles denote averages for (τ=50, γ=100) (red triangles representing averages for coupled trials are hidden from view, because they overlap with values from one-way trials). Error bars, which are barely perceptible, denote standard error of the mean. Part (b) depicts the mean *difference* in Gapped Consonance between coupled trials and the correspondingly yoked one-way trials, with positive values indicating greater consonance in coupled agents.

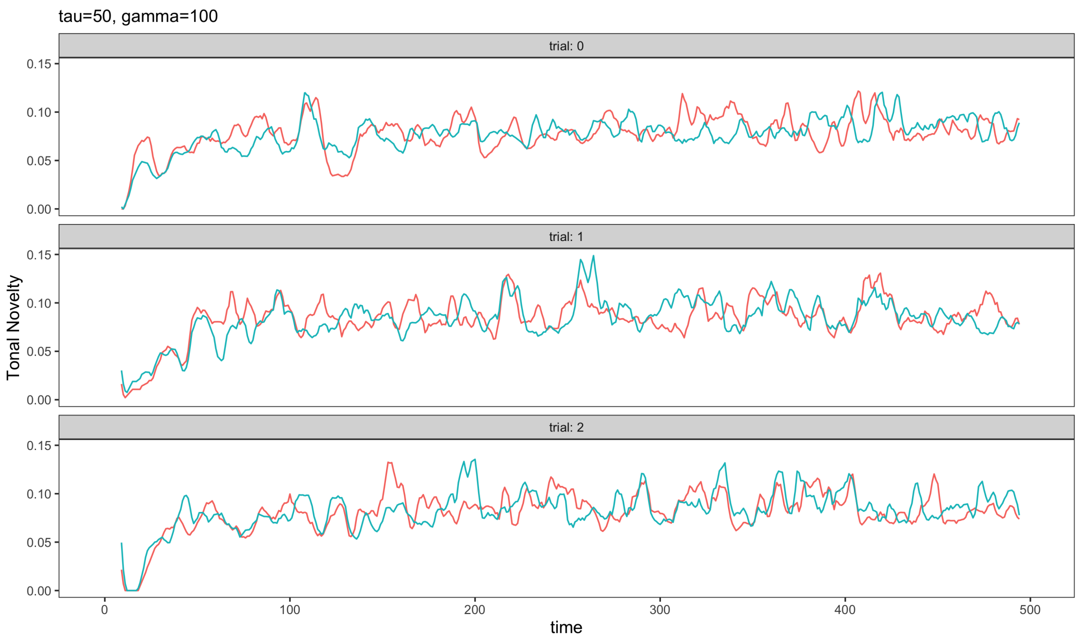
3.4 Mutual coupling and Tonal Novelty

We have just seen evidence supporting the idea that mutual coupling promotes increased consonance within tonal basins. But what effect does mutual coupling have on agents' ability to collectively transition between disparate basins? One hypothesis, presented in the Introduction, is that mutually coupled agents will be less likely to transition out of established tonal basins because there is more "inertia" compared to overdubbed agents. To answer this question, time series of Tonal Novelty were computed with a five step sliding window for all trials in each condition and parameter combination. Figure 7 shows time series for exemplar trials, with correspondingly yoked *coupled* (red) and *one-way* (blue) trials plotted on the same axis.

Peaks in these time series can be interpreted as transitions, or moments of high contrast between the tonality of preceding and future windows, whereas valleys indicate



(a)



(b)

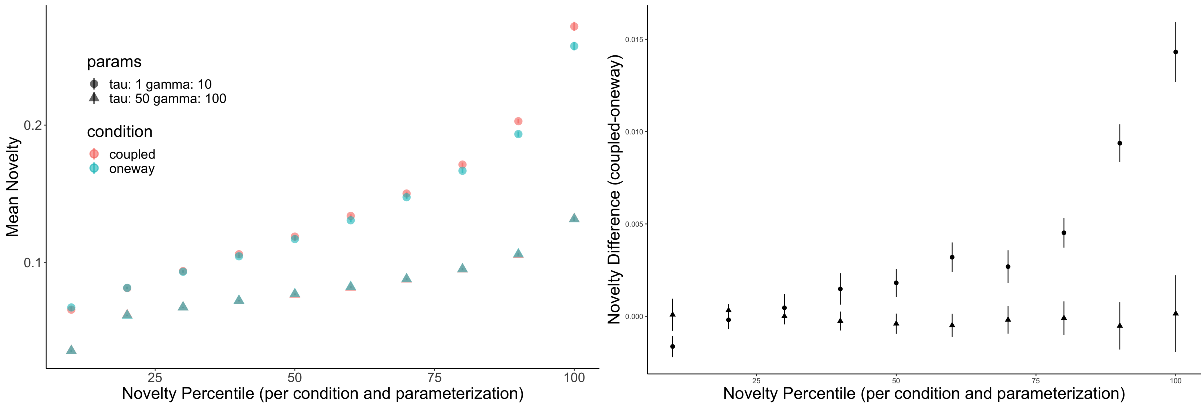
Figure 7. Time series of Tonal Novelty for sample trials. Each axis depicts the rolling average of novelty over time (obtained by taking a five-step rolling average from the raw novelty time series) for a coupled trial (red) and the corresponding yoked one-way trial (blue). For (τ=1, γ=10) trials, depicted in (a), peaks appear to be higher in coupled trials compared to their counterpart overdubbed trials. This does not appear to be the case for (τ=50, γ=100) trials, depicted in (b).

low levels of Tonal Novelty, which would be expected for periods of homogeneous tonality. Based on sampled sessions, there appears to be strong alignment between novelty time series in yoked trials with the (τ=1, γ=10) parameter setting (top panel), except that peaks appear to be generally higher in coupled trials. It is natural to expect time alignment of peaks and valleys in yoked trials, because changes in tonality that originally occurred in coupled trials are encapsulated in the note sequences of these agents, which serve as "ghost partners" in subsequent one-way trials. However, the observation that peaks appear to be higher in coupled trials bears further examination.

For the (τ=50, γ=100) parameterization (lower panel), there does not appear to be the same alignment of yoked trials, nor the same effect of condition on peak height. We can make sense of this by keeping in mind that these trials produced entrenched basins; there were no meaningful changes in tonality throughout these simulations, so the fluctuations in Tonal Novelty reflect fluctuations within a given basin, which can be observed by listening to a recording synthesized from model outputs under this parameterization (`https://mattsetz.github.io/dissertation-media/`).

In order to more rigorously test the qualitative trend seen in individual sessions, that high novelty values are higher in coupled than one-way trials for agents producing shifting tonal basins, we compared novelty values at equivalent percentiles throughout the range of novelty exhibited in each condition. For each trial, novelty values were binned into deciles (i.e., the first decile comprised the lowest 10% of novelty values, and last decile comprised the highest 10%) and average novelty was computed within each decile. This resulted in ten mean novelty values per trial. We then averaged these values across all trials in each condition. Figure 8 displays the results of this analysis, with circles representing mean novelty in (τ=1, γ=10) agents, and triangles representing values for (τ=50, γ=100).

In the left panel (Figure **??**), red marks indicate averages for coupled trials and blue marks for one-way trials. (Values for coupled trials in the (τ=50, γ=100) parameterization are occlued by values for one-way trials, as there was such tight overlap between each condition.) Overall, novelty is higher in (τ=1, γ=10) agents, and these agents also exhibit a greater range of novelty values. This makes sense because this parameterization produced shifting tonal basins, whereas the other



(a) (b)

Figure 8. High novelty values are higher in *coupled* than *oneway* trials for ABMs producing shifting tonal basins. For each trial, novelty time series were binned into deciles and the mean novelty value for each decile was computed. These novelty scores were then averaged across all trials in each condition and parameterization (results from 20 trials in each condition and parameter setting are shown here). In (**??**), circles represent the mean novelty for each decile for (τ=1, γ=10) trials, and triangles represent those values for (τ=50, γ=100) trials (*coupled* values are shown in red and *one-way* values in blue; *coupled* values are hidden from view, because they overlap so closely with *one-way* values).) Error bars represent standard error of the mean. Part (b) shows the average *difference* in Tonal Novelty for each decile between coupled and correspondingly yoked one-way trials, with positive values indicating higher novelty in coupled trials.

parameterization produced entrenched basins. By definition, we expect novelty to be low in cases of unchanging tonality. However, it is important to recognize that (τ=1, γ=10) agents do not *always* produce high novelty; there are also relatively low novelty values as well. This is consistent with the idea that agents in this parameterization produce shifting tonal basins — they do indeed achieve stable basins that persist for sustained periods of time, and in these periods novelty is low. But unlike (τ=50, γ=100), these basins are liable to change over time, and novelty will be high during such transitions.

Now let us turn our attention to how the presence or absence of mutual coupling influences novelty across different deciles in each parameterization. This is best visualized in the right panel (Figure **??**), which shows the *difference* in Tonal Novelty between correspondingly yoked coupled and one-way trials, with positive values indicating greater novelty in the coupled condition. For (τ=50, γ=100) (triangles), there is no effect of condition; triangles hover around zero for each decile. However, there is a compelling interaction between decile and condition in (τ=1, γ=10). The effect of condition starts off negative in the lowest decile (indicating increased novelty in one-way trials for the lowest decile values), and monotonically increases, becoming more positive (indicating more novelty in coupled trials) at higher deciles.

Given the previous gapped-consonance analysis, we shouldn't be surprised that novelty within the first decile is higher in one-way duos. We already know from the gapped-consonance analysis that this parameterization supports short-lived tonal basins that persist for sustained but finite periods of time, before the tonality shifts somewhere else. Thus, low novelty values in these simulations are going to correspond to windows falling within tonal basins — periods of a unified, sustained tonality. We also know from the gapped-consonance analysis that mutual coupling supports consonance within tonal basins, and this increased tonal coherence translates to lower levels of novelty. This explains why novelty is lower in coupled trials at the smallest decile.

The fact that the effect of mutual coupling grows more positive at higher deciles tracks with our previous observations that novelty peaks at this parameterization appear to be higher in coupled trials compared to correspondingly yoked one-way trials. Another way of stating this result is that high novelty values are "higher" in mutually adaptive agents. This is interesting, and runs against the hypothesis that mutual coupling would interfere with transitions between basins by introducing more "inertia". Instead, while mutual coupling does indeed increase tonal coherence within basins, it also creates conditions for sharper, more punctuated transitions between basins.

4 Discussion

In this paper we have presented a novel agent-based model of improvised tonal coordination. The model provides an idealized computational testbed to simulate the same experimental conditions that human improvisers were subject to in Setzler and Goldstone (2020): coupled versus one-way interaction. As in the empirical study, contrasting the music produced in these conditions allows us to isolate the effects of mutual coupling — how it constrains and enables certain patterns of collective musical expression that are not possible in one-way settings. The TonalEmergence model was deliberately formulated to be as simple as possible. It has two parameters: memory (τ) and entropy (γ), which are psychologically motivated and straightforward to interpret.

We reproduced a central finding of the empirical study in one of the simplest imaginable parameterizations of our model: bidirectional coordination increases consonance between notes of improvising musicians. Subsequently, in contrasting model behavior in different parameter settings, we learned that certain "sweet spots" in (memory, entropy) parameter space support the same kinds of tonal dynamics that underpin coherence and suspense in naturalistic improvised music; namely the emergence of tonal basins, and the existence of transitions between them. Lastly, we observed interactions between different parameter settings and the interaction condition, demonstrating that coupling condition (oneway vs. coupled) interacts with memory and entropy to produce different patterns of emergent tonal dynamics.

4.1 Bidirectional coordination increases consonance in professional musicians and formal agent models

We first replicated one of the major findings of the empirical study by showing that coupled agents, mutually harmonizing with each other's previous notes, achieve greater Emergent Consonance than agents in one-way trials, where a single agent harmonized with the previous notes of an unresponsive ghost partner. Given these results, the model provides a "proof of concept" mechanistic account for the finding that bidirectional coordination increases consonance in human improvisation.

As discussed in the Introduction, there are more elaborate accounts for why this might be the case. One is that mutual adaptation fosters alignment of abstract mental representations in human improvisers, such as a shared understanding of the current tonality and where it might be headed. There is evidence suggesting that this kind of thing happens in non-musical contexts, as it has been shown that alignment occurs more in conversation than in lecture settings (Garrod & Pickering, 2009; Pickering & Garrod, 2004). Alternatively (though not mutually so), a "theory of mind" account might say that human improvisers become aware of the fact that they are improvising with either a human or an unresponsive ghost partner, and this awareness influences their motivation to produce consonance with their counterpart.

But agents in the TonalEmergence model have only very primitive internal mental representations[10], and they certainly don't have any theory of mind. Instead, the present finding demonstrates that Emergent Consonance can result as an emergent side-effect of bidirectional coordination in which agents mutually harmonize with each other's previous notes. This is not to say that more elaborate phenomena are not at play in co-improvising humans — more empirical work would be needed to established this — but it does provide a proof of concept sufficient to convince us that we would obtain similar empirical results even without these higher-level considerations. This finding echoes previous modeling work, which has demonstrated that mutually adaptive agents achieve greater synchrony (Demos, Carter, Wanderley, & Palmer, 2017; Demos, Layeghi, Wanderley, & Palmer, 2019; Noy et al., 2011), except that here we show this effect with respect to tonal consonance.

The same result was found with respect to Combined Consonance. In other words, not only did mutual coupling increase consonance *between* agents' notes, it also increased the overall consonance of their collective music making. This differs slightly from results in the empirical study. Humans exhibited the same asymmetry in lagged Combined Consonance shown in overdubbed trials (indicating that live musicians harmonized with preceding notes of the ghost recording, but not vice versa), but overall

[10] Agents do internally represent consonant versus dissonant intervals (as stipulated in the consonance measure), and this knowledge is used to infer note-generation probability distributions biased towards maximizing consonance given their partner's previous notes. This being said, such representations are quite simple compared to the explicit knowledge of key centers and conventions of functional harmony, possessed by human improvisers. Thus, we conclude that the effects of mutual coupling observed here are more an emergent consequence of mutually harmonizing agents as opposed to the result of the alignment of high-level cognitive representations.

there was not a statistically significant main effect of condition on simultaneous Combined Consonance.

This being said, average Combined Consonance trended higher in coupled trials in the empirical study, and this finding was relatively robust across different window sizes even though it didn't reach statistical significance. Given this trend and the present findings with respect to TonalEmergence, it could be that coupling does result in higher Combined Consonance in humans as well, and that we simply didn't have enough statistical power to demonstrate this in our empirical samples. Alternatively, this discrepancy could be due to a fundamental difference in the objectives of TonalEmergence as compared to those of humans. In naturalistic music, some level of tension via tonal dissonance is typically desirable, even if it eventually resolves to a more consonant tonality. Agents in this model, on the other hand, are explicitly programmed to maximize consonance (albeit with some level of randomness) with no direct pressure to introduce dissonance. For now this remains an open question — further experimentation would be needed to disentangle these two possibilities.

4.2 Dynamics of music improvisation emerge out of a minimal set of mechanisms

After validating the model against the empirical lagged-consonance analysis, we next contrasted the model's behavior at different parameterizations to see how different values of memory and entropy supported different kinds of emergent tonal dynamics. The *memory* $\times$ *entropy* parameter space is replete with degenerate instantiations of TonalEmergence that either produced random atonal walks or performances that were fully determined by the initial randomly seeded notes. This being said, certain interesting regions of parameter space, corresponding to "sweet spot" (memory, entropy) combinations, supported the emergence of *tonal basins* — stable tonalities that persisted for sustained periods of time. Certain parameterizations (τ=50, γ=100) produced *entrenched* tonal basins — once agents arrived at a given tonal basin, they were stuck there indefinitely. Other parameterizations (τ=1, γ=10) produced *shifting* tonal basins, in which agents would arrive at a tonal basin, generate notes within that tonality for some continued period of time, and then transition to a qualitatively different tonality.

These two kinds of tonal dynamics underscore an important balance that human improvisers must negotiate, which is akin to the explore/exploit dilemma that drives search tasks in many contexts (Hills et al., 2015). On the one hand, without tonal basins there would be no tonal coherence, and improvisations would sound like atonal drift. In some cases this is desirable, but some amount of coherent tonal structure is necessary to maintain the interest of most audiences. On the other hand, staying too long in a given tonality often becomes monotonous, and defeats the spontaneity of improvised music performance.

The dynamic of "shifting tonal basins" figures importantly into naturalistic improvised performance because it strikes a balance between order and surprise. Listening to the Taborn/Iyer performance, we hear tonality evolving dynamically throughout the improvisation. Well-defined tonal centers are established, exploited for sustained periods of time, transitioned between, and sometimes interspersed with less structured atonal or "quasi-tonal" sections, which eventually converge on a structured tonal center. Remarkably, we observe the same kind of dynamics in (τ=1, γ=10) parameterization of TonalEmergence.

Why does (τ=1, γ=10) produce shifting basins, while (τ=50, γ=100) produces entrenched basins? In both cases, once agents arrived at a given tonal basin, certain notes produce more consonance (e.g., root and fifth) and are thus more likely to occur. However, more dissonant outlier pitches (e.g., augmented) are also liable to occur, depending on the degree of entropy. With enough memory, outlier pitches don't exert any lasting influence on the tonality of the performance, because the memory bank of previous pitches acts as a stabilizing reservoir of mostly within-basin pitches. Conversely, for agents with shorter memory spans, there is a smaller reservoir of past notes anchoring them to the current basin, and outlier pitches are more liable to initiate a transition to a new tonality.

Imagine that Agents A and B are entrenched in C major. If Agent B is playing with some randomness, they might play an outlier pitch (e.g., F#). If memory is sufficiently long, Agent A will still be influenced by many previous pitches within C major, so they will likely not respond to the F#. And even if they do, they will eventually be absorbed back to C major by the stabilizing influence of the large reservoir of past C-major notes. But if agents have low memory (as in $tau1$=1, in which agents just attend to their partners' previous three or four notes), Agent A will respond solely to the recent F#, and a new tonality would be established.[11]

We've been using the term "memory" as a shorthand interpretation of τ, but we don't mean to suggest that literally decreasing the working memory capacity of improvising musicians would make them more produce more dynamic music. Rather, our interpretation of τ is that it reflects how tied a player is to the full history of played notes rather than the most recent ones. A player with a large memory capacity may strategically choose to have a lower value of τ because it will allow them to systematically explore other tonalities.

4.3 Mutual coupling supports more pronounced tonal dynamics

How does mutual coupling play into these different kinds of tonal dynamics? The gapped-consonance analysis revealed that for (τ=50, γ=100), there was a small but statistically significant effect of condition, where there was more consonance in coupled than one-way trials. This effect was robust across all gap sizes, indicating that mutual coupling increased consonance across local and global timescales. We also learned that there was no effect of condition on Tonal Novelty for this parameterization.

For (τ=1, γ=10) there was a large effect of condition (higher consonance in coupled than in one-way) at small gaps, which monotonically decreased with increasing gap sizes. Thus for agents producing shifting tonal basins, mutual coupling supported local tonal coherence across relatively small timescales, but not at larger timescales. Mutual coupling also had a surprising effect on Tonal Novelty: high novelty values were

[11] It should be noted that although τ=1 is reported here, a parameter sweep confirmed that shifting basins also emerge with slightly larger τ values (such as τ=5), as documented in Figure B1.

higher in coupled trials, and low novelty values were lower in coupled trials. In other words, mutual coupling didn't simply produce higher or lower novelty across the board; instead, it supported a greater range of novelty throughout simulated improvisations.

Low novelty values correspond to periods in which agents were entrenched in tonal basins, and we already know from the gapped-consonance analysis that mutual coupling supported more consonance within basins, which is consistent with the fact that low novelty values are lower in coupled trials. High novelty values correspond to potential transitions between qualitatively different tonal basins. The fact that novelty was higher in these periods for coupled trials tells us that mutual coupling supported larger, more pronounced transitions between disparate tonal basins than was possible in overdubbed interaction.

How can we make sense of these latter results pertaining to (τ=1, γ=10)? We might have expected a very different outcome for the tonal-novelty analysis. It could have been that mutual coupling made it *more* difficult for agents to transition between basins, reflecting the same logic about increased memory encouraging entrenched basins. Just as with larger memory, mutual coupling effectively provides a larger reservoir of pitches anchoring dyads to an established basin. But the outcome of the Tonal Novelty analysis tells the opposite story: mutual coupling results in transitions of larger magnitude than those achieved in overdubbed interaction. There is a puzzling tension between this finding and that of the gapped-consonance analysis. The latter tells us mutual coupling produces stronger tonal coherence (at least on relatively short timescales), but the former tells us that mutual coupling also results in sharper transitions between basins.

How can we reconcile the fact that coupled agents produce greater local tonal coherence *and* greater novelty? Imagine that Agent A and Agent B are mutually coupled, and are entrenched in a tonal basin (e.g., C major). With some randomness, Agent B plays a note outside the tonal basin (e.g., F#). Agent A then responds to B's outlier note, B responds to A's response, and so on, until A and B are in a new tonal basin (e.g., F# major) initiated by B's original outlier note. By contrast, imagine Agent

C is playing with Ghost A (i.e., recording of Agent A) in an overdubbed condition, and is entrenched in a tonal basin (e.g., C major). By some randomness, C plays a note outside the tonal basin (e.g., F#). By definition, Ghost A is unresponsive to C's outlier note and continues on unperturbed, such that C eventually gets absorbed back into the enduring tonal basin. In this case, instead of initiating a transition, C's outlier note simply "muddied the waters" and introduced dissonance into the tonal basin.

The above "muddying the waters" account explains why mutual coupling supports greater local tonal coherence and transitions of larger magnitude. Transitions have the potential to be emergent outcomes of positive feedback loops between mutually adaptive agents. These feedback loops enable a randomly generated outlier note to initiate a rapid, coordinated transition to a new tonal center. Additionally, without the muddying effect of overdubbed interactions, in which stray notes introduce transient dissonance in otherwise stable basins, coupled agents are able to produce more consonance within local tonalities. This, in turn, further increases the magnitude of novelty values at transition-points, because novelty partially depends on tonal coherence within windows on either side of a potential transition.

In this way, mutually coupled agents produce more pronounced tonal dynamics – more consonant tonal basins, and sharper transitions between them. To make an analogy with amplitude dynamic range, it would be as if music produced by mutually adaptive agents consisted of very loud sections and very quiet sections with sharp transitions between them, whereas music produced by overdubbed agents was at a more medium volume level throughout. Here we observe this kind of pattern in tonal space, not loudness. In terms of the toy illustration in Figure 3a, it is as if unidirectional coordination results in less distinct tonal segments, representing less coherent local basins and less distinct transitions between them. This finding has not yet been verified empirically with respect to human improvisers. Such a validation is outside the scope of this study, but the data set presented in Setzler and Goldstone (2020) provides an excellent resource for future work to follow up on, because it involves world-class human improvisers playing in the same interaction conditions that are simulated in this agent-based model.

The answer is likely to be more nuanced in human improvisers, who are much more sophisticated than the agents in this model. "Muddying the waters" might be an intentional stylistic device selectively employed by human improvisers. This being said, some of the most compelling moments in freely improvised music occur because of pronounced tonal dynamics — musicians suddenly converging on a structured tonality seemingly out of an abyss of atonality, or surprising an audience by transitioning out of a tonal center to something qualitatively different (Corbett, 2016). It seems remarkable that an ensemble of improvising musicians can collectively achieve such structured and coordinated transitions without the guidance of a musical score, conductor or any advance planning. This model shows that such moments are a special property of the mutual adaptations that bind co-improvising musicians.

4.4 Relation to Existing Work

4.4.1 Connections to complex systems. The emergence of tonal basins examined here bears an interesting resemblance to the phenomenon of punctuated equilibrium, which is the notion that complex systems often give rise to novel structures that emerge quite suddenly, and persist for long periods of time. This idea was first put forth by Gould and Eldredge (1972) in the context of population genetics (where "stable structures" refer to species), and it has since been argued that a similar phenomenon holds for linguistic and technological innovations in cultural evolution (Atkinson, Meade, Venditti, Greenhill, & Pagel, 2008; Valverde & Solé, 2015). A detailed discussion of the connections between emergent tonal basins and punctuated equilibrium in genetic and cultural evolution is beyond the scope of this paper, but it is intriguing to note that the musical dynamics produced by TonalEmergence resemble emergent dynamics in complex systems that seem quite different on the surface.

Past work has shown that punctuated equilibria can occur in cases of neutral drift, where there is no explicit pressure for speciation, simply as a consequence of compounding dynamics of distributed systems that match one another (Gould &

Eldredge, 1972). Here we observe a similar phenomenon. There is no explicit pressure for agents to produce tonal basins, but they emerge as a consequence of a simple bias to harmonize with previously played notes. Interestingly, whereas gene propagation in a population is governed by copying, the generation of tonality in the present model is governed by a complementary tonal coordination, as certain pitch combinations have higher or lower degrees of consonance.

The emergence of tonal basins is also reminiscent of informational cascades in interacting social groups, where small signals (such as preference for a particular kind of vehicle) are amplified by other members of the group, which in turn leads to even more amplification, until the signal suddenly reaches consensus (i.e., everyone ends up owning the same vehicle) (Bikhchandani, Hirshleifer, & Welch, 1992, 1998). In the case of music, one musician (or agent) can signal a particular tonal center, such as C major, by playing a set of notes (e.g., C and G) repeatedly, or somehow emphasizing them over other notes. Another musician in the ensemble is liable to mimic this signal by emphasizing this same set of notes, or playing complementary notes in response. In this way, the signal to establish a new tonal center can become amplified to the point where the entire ensemble is playing in C major.

4.4.2 Connections to other agent-based music models. There is a rich history of computationally-minded composers using agent-based models to generate music (T. Blackwell, 2007; Eldridge & Bown, 2018). One of the first examples was T. M. Blackwell and Bentley (2002), who mapped the self-organizing movements of "boids" in the classic flocking model (Reynolds, 1987) to musical parameters, and synthesized these parameters to sound in real-time. This system was capable not only of generating music, but also of adapting to collaborating human improvisers in realtime. Subsequently, there have been many variants of the idea of using agent-based models for music generation (Beyls, 2007; Eigenfeldt & Kapur, 2008; Eigenfeldt & Pasquier, 2011; Hutchings & McCormack, 2017). This work is fascinating, and is obviously related to the present agent-based model, but it is critical to recognize that these systems were *aesthetically*, not scientifically, motivated. These models were constructed either by experimental composers interested in exploring new ways of generating sound, or by complex systems theorists interested in sonifying concepts of complexity and self-organization. In this sense, they belong to a different class of models than the current agent-based model, though TonalEmergence could be adopted for more artistic purposes.

Scientifically motivated computational models aimed at uncovering the mechanisms of joint music performance have been much sparser. In a notable exception, Demos et al. (2019) modeled synchronization in joint music performance (piano duets) using a delay-coupled dynamic model, where each agent adjusted their frequency based on their partner's previous phase at some time delay. Model simulations predicted outcomes of a human study, namely that asynchronies between co-performers' note onsets would be larger in conditions where auditory feedback was removed compared to baseline conditions in which pianists could mutually adapt to one another. However, this model was formulated to study synchrony in the context of scored music, which is quite different from the purpose of TonalEmergence in the present work — namely, to understand how improvising musicians collectively generate tonality without a written score. Models of collective music improvisation have been even rarer, and previous efforts have suffered from being divorced from empirical data, from lacking specific aims, and from being overly complicated due to an abundance of many confounding parameters (Canonne & Garnier, 2011). To our knowledge, this is the first agent-based model of collective improvisation that has been validated against human improvisers playing in the same conditions.

4.5 Future Directions

There are a number of pathways to extend this work. One would be to perform the same kinds of gapped-consonance and tonal-novelty analyses to empirical data collected in Setzler and Goldstone (2020). One might hypothesize that, as in TonalEmergence, mutual coupling enables more exaggerated tonal dynamics — increased local tonal coherence and more pronounced transitions between different tonalities — in human improvisers. Then again, we found that this effect occurred only with specific parameterizations of TonalEmergence that gave rise to shifting tonal basins, so we would expect to see it only in cases where human improvisers exhibit shifting tonal basins. Perhaps this is just one of several types of dynamics available to human improvisers. In addition to analyzing the existing data set from Setzler and Goldstone (2020), it would be interesting to manipulate entropy and memory in human improvisers — perhaps by instructing them to play more wildly or in a way that is more immediately reactive to what they have just heard — and to see if the same kinds of tonal dynamics observed here are also observed empirically.

Aside from future empirical experimentation, there are several ways in which the model itself can be extended. For example, agents in the present model are purely reactive: their note generation is based entirely on their partners' previous notes. In actuality, note generation in human improvisers is also partially driven by internal forces. One way to incorporate this in the model would be to add a self/other parameter tuned to the degree to which agents are *reactive* or *self-driven*. We know that mutual coupling improves tonal coordination over completely unidirectional coordination, but these are simply two poles of a continuum; perhaps some intermediate degree of influence imbalance stabilizes tonal coordination, or opens new possibilities for evolving tonal dynamics. Lastly, it would be of interest to allow parameters of the model (entropy, memory, self/other) to evolve throughout simulated improvisations, as these parameters are no doubt in flux in human improvisation. In the spirit of minimal modeling however, we should be careful to incrementally examine any extensions to TonalEmergence, and any such extensions should be motivated by clear, operationalized scientific questions.

5 Conclusion

There have been many theoretical papers describing how collectives of improvising musicians can be understood as complex dynamical systems capable of producing emergent structure without any central leader or advanced planning (Borgo, 2005;

Van der Schyff, Schiavio, Walton, Velardo, & Chemero, 2018; A. Walton, Richardson, & Chemero, 2014). Yet despite this abundance of theory, these ideas have yet to be systematically examined in a formal computational setting. Here we have addressed this gap by presenting a novel agent-based model of tonal coordination that was deliberately formulated to simulate the same experimentally controlled interaction conditions that professional improvising musicians were subject to in a previous empirical study. The model successfully reproduced important results from this study, and thus provides a feasible mechanistic explanation for the outcome that mutual coupling supports greater tonal consonance in improvising musicians. Furthermore, the model showed how a minimal set of mechanisms in interacting agents can give rise to a range of complex dynamics essential to naturalistic improvised music. TonalEmergence thus enriches our understanding of musical collectivity in humans and artificial agents alike, and suggests new pathways for empirical investigation that may continue to reveal the mechanisms underpinning the marvel of joint musical improvisation.

6 Acknowledgements

The authors thanks Douglas Hofstadter and Eduardo Izquierdo for their invaluable support and feedback.

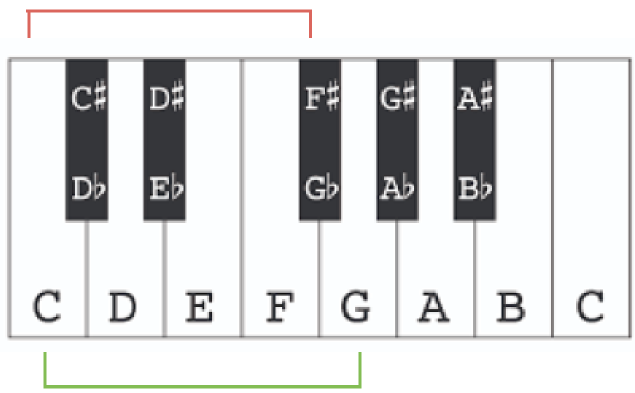
7 Declarations of Interest

Declarations of interest: none.

Appendix A. Consonance Measures

The consonance measures reported in this paper are all either identical to, or derivative of measures used in Setzler and Goldstone (2020) to empirically evaluation consonance produced by human musicians. The measures are adapted from the Tonal Spiral Array model (Chew, 2005; Chew et al., 2014), which has been empirically validated against listener ratings and expert music theory analyses of musical tension. The reader is encouraged to refer to these publications for motivation of these model, and its applicability to assessing consonance in freely improvised music.

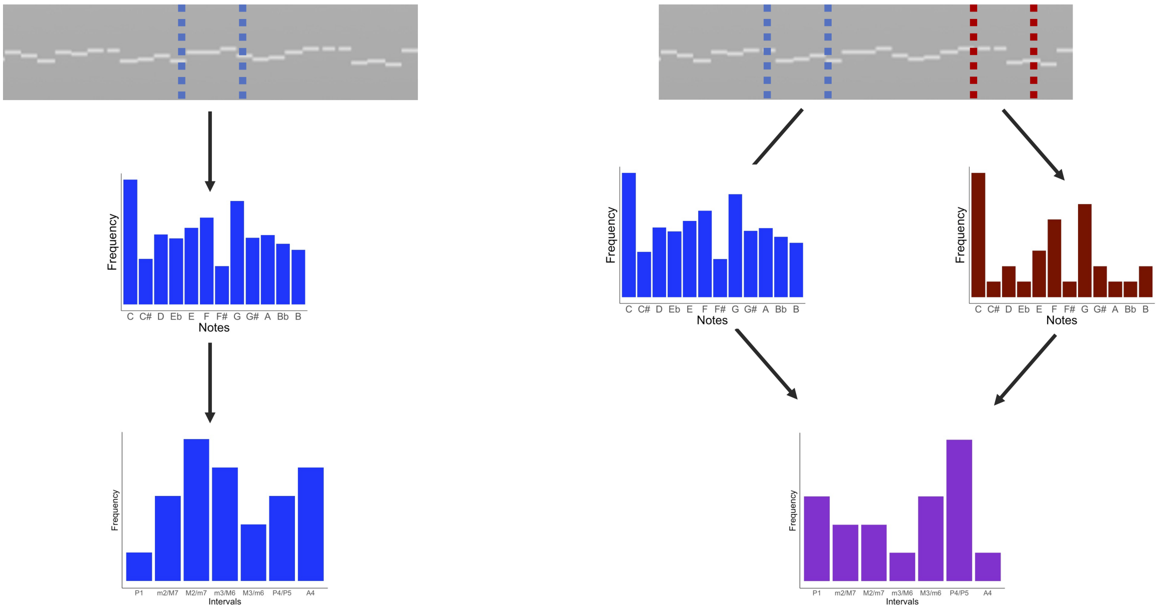
The core intuition behind the consonance measure is that certain pairwise intervals (i.e., intervals between two pitches) are inherently more or less consonance/dissonant. For example, a perfect 5th is a highly consonant interval, whereas a tritone is a highly dissonance interval. Thus, each interval was assigned a dissonance score, as shown in Figure A1. Dissonance scores were obtained by distances in the Tonal Spiral Array model, as in Setzler and Goldstone (2020).



interval	dissonance
unison	1.0
m2/M7	2.6
M2/m6	2.2
m3/M6	1.9
M3/m6	1.7
P4/P5	1.5
A4	3.2

Figure A1. Dissonance scores for pairwise intervals. These scores were taken from the Tonal Spiral Array model (Chew et al., 2014), as described in Setzler and Goldstone (2020). The Tonal Spiral Array model has been empirically validated against listener ratings and expert music theory analyses of musical tension (Chew et al., 2014; Herremans, Chew, et al., 2016).

As depicted in Figure A2, to compute consonance for a given window of music, we first create a frequency histogram of how often each note is played within a time-window. We then convert this note histogram into an interval histogram by iterating over all combinations of pitch pairs, summing the product of their frequency heights, and incrementing the corresponding bins in the interval histogram. We then normalize the interval histogram so that the heights of all bars sum to 1. Consonance is computed as the negative weighted sum[12] of dissonance scores for each interval, scaled by how often they occurred within the window. Finally, consonance is normalized by adding 3.247 (highest dissonance score) and dividing by 2.247 (difference between the highest and lowest dissonance scores). Consequently, the measure has a theoretical range of 0-1, where 0 represents minimal consonance (i.e., a tritone interval) and 1 represents maximal consonance (a unison). We refer to this as $consonance_{within}$, as it denotes consonance of notes *within* a time window.



(a) Consonance within a time window.

(b) Consonance between two time windows.

Figure A2. Computing inner- and inter-window consonance. A given window of music defines a note-set, which can be expressed as a normalized frequency histogram of how often each note is played. As shown in A2a, consonance *within* a given window can be computed by constructing a frequency histogram of how often each interval occurs amongst the notes of said window, and computing a negative weighted sum of interval-dissonance scores scaled by how often each interval occurs. As shown in A2b, consonance can similarly be computed *between* two note-sets. The only difference here is that the interval histogram is computed based on intervals occurring between the notes in both sets. These diagrams are for illustrative purposes only, and the values of these histograms were arbitrarily synthesized.

[12] *Negative* weighted sum, because consonance is the opposite of dissonance.

Similarly, we can compute consonance *between* two windows of music by (i) constructing note-frequency histograms for each time window, (ii) computing an inter-window interval histogram by iterating over all combinations of pitch pairs *between* each window, and (iii) computing the negative weighted sum of dissonance scores for each interval, scaled by their frequency in this inter-window interval histogram.

Appendix B. Gapped Consonance over fine-grained parameter sweep.

In addition to the Gapped Consonance analysis reported in the main text, we also analyzed a finer-grained set of parameterizations around the $(\tau = 1, \gamma = 10)$ parameterization which supported shifting tonal basins. Results, plotted in Figure B1, show that shifting basins were supported across a range of other parameter values in this vicinity. The interaction between gap-size and interaction condition, where mutual coupling increased local (but not global) tonal coherence, was also found to be robust in this region of parameter space.

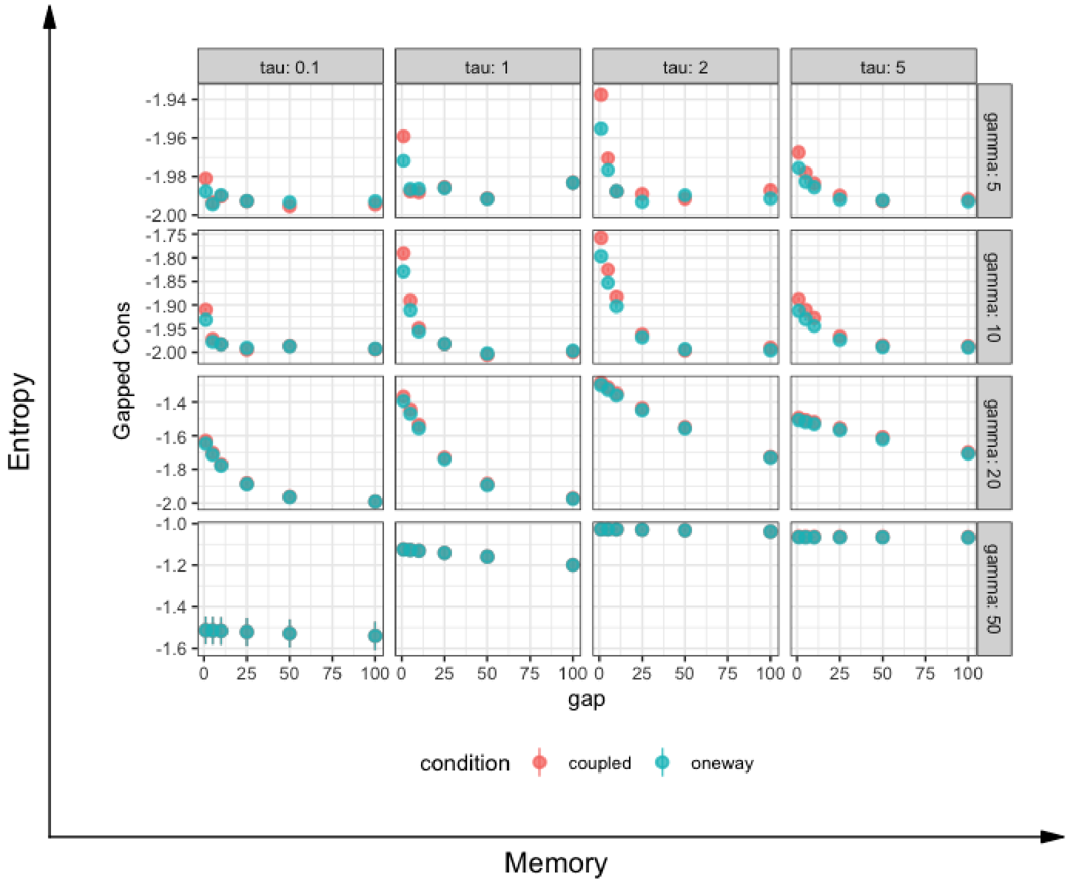


Figure B1. Parameter sweep of Gapped Consonance analysis. Gapped Consonance was computed over a range of (memory, entropy) parameter combinations. 20 trials of 500 timesteps were simulated for both conditions in each parameterization. Points represent mean Gapped Consonance across all trials in a given condition. Error bars denote standard error of the mean.

8 References

Aldwell, E., Schachter, C., & Cadwallader, A. (2018). *Harmony and voice leading.* Cengage Learning.

Atkinson, Q. D., Meade, A., Venditti, C., Greenhill, S. J., & Pagel, M. (2008). Languages evolve in punctuational bursts. *Science*, *319*(5863), 588–588.

Aucouturier, J.-J., & Canonne, C. (2017). Musical friends and foes: The social cognition of affiliation and control in improvised interactions. *Cognition*, *161*, 94–108.

Baker, C. L., Jara-Ettinger, J., Saxe, R., & Tenenbaum, J. B. (2017). Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behaviour*, *1*(4), 1–10.

Beer, R. D. (1995). A dynamical systems perspective on agent-environment interaction. *Artificial intelligence*, *72*(1-2), 173–215.

Beer, R. D. (2000). Dynamical approaches to cognitive science. *Trends in cognitive sciences*, *4*(3), 91–99.

Berkowitz, A. (2010). *The improvising mind: Cognition and creativity in the musical moment.* Oxford University Press.

Beyls, P. (2007). Interaction and self-organisation in a society of musical agents. In *Proceedings of ecal 2007 workshop on music and artificial life (musical 2007).*

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, *100*(5), 992–1026.

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1998). Learning from the behavior of others: Conformity, fads, and informational cascades. *Journal of Economic Perspectives*, *12*(3), 151–170.

Blackwell, T. (2007). Swarming and music. In *Evolutionary computer music* (pp. 194–217). Springer.

Blackwell, T. M., & Bentley, P. (2002). Improvised music with swarms. In *Proceedings of the 2002 congress on evolutionary computation. cec'02 (cat. no. 02th8600)* (Vol. 2, pp. 1462–1467).

Borgo, D. (2005). *Sync or swarm: Improvising music in a complex age.* A&C Black.

Briot, J.-P., Hadjeres, G., & Pachet, F.-D. (2017). Deep learning techniques for music generation—a survey. *arXiv preprint arXiv:1709.01620*.

Candadai, M., Setzler, M., Izquierdo, E. J., & Froese, T. (2019). Embodied dyadic interaction increases complexity of neural dynamics: A minimal agent-based simulation model. *Frontiers in psychology*, *10*, 540.

Canonne, C., & Garnier, N. (2011). A model for collective free improvisation. In *International conference on mathematics and computation in music* (pp. 29–41).

Chew, E. (2005). Regards on two regards by messiaen: Post-tonal music segmentation using pitch context distances in the spiral array. *Journal of New Music Research*, *34*(4), 341–354.

Chew, E., et al. (2014). Mathematical and computational modeling of tonality. *AMC*, *10*, 12.

Chiel, H. J., & Beer, R. D. (1997). The brain has a body: adaptive behavior emerges from interactions of nervous system, body and environment. *Trends in neurosciences*, *20*(12), 553–557.

Christensen, T. (2006). *The cambridge history of western music theory.* Cambridge University Press.

Corbett, J. (2016). *A listener's guide to free improvisation.* University of Chicago Press.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and brain sciences*, *24*(1), 87–114.

De Dreu, C. K., Nijstad, B. A., Baas, M., Wolsink, I., & Roskes, M. (2012). Working memory benefits creative insight, musical improvisation, and original ideation through maintained task-focused attention. *Personality and Social Psychology Bulletin*, *38*(5), 656–669.

Demos, A. P., Carter, D. J., Wanderley, M. M., & Palmer, C. (2017). The unresponsive partner: roles of social status, auditory feedback, and animacy in coordination of joint music performance. *Frontiers in Psychology*, *8*, 149.

Demos, A. P., Layeghi, H., Wanderley, M. M., & Palmer, C. (2019). Staying together: A bidirectional delay–coupled approach to joint action. *Cognitive Science*, *43*(8), e12766.

Eigenfeldt, A., & Kapur, A. (2008). An agent-based system for robotic musical performance. In *Proceedings of the 2008 conference on new interfaces for musical expression (nime08), genova, italy* (pp. 144–149).

Eigenfeldt, A., & Pasquier, P. (2011). A sonic eco-system of self-organising musical agents. In *European conference on the applications of evolutionary computation* (pp. 283–292).

Eldridge, A., & Bown, O. (2018). Biologically-inspired and agent-based algorithms for music. In A. McLean & R. T. Dean (Eds.), *The oxford handbook of algorithmic music* (pp. 209–244). Oxford University Press.

Engel, A., & Keller, P. E. (2011). The perception of musical spontaneity in improvised and imitated jazz performances. *Frontiers in psychology*, *2*, 83.

Foote, J. (2000). Automatic audio segmentation using a measure of audio novelty. In *2000 ieee international conference on multimedia and expo. icme2000. proceedings. latest advances in the fast changing world of multimedia (cat. no. 00th8532)* (Vol. 1, pp. 452–455).

Garrod, S., & Pickering, M. J. (2009). Joint action, interactive alignment, and dialog. *Topics in Cognitive Science*, *1*(2), 292–304.

Goebl, W., & Palmer, C. (2009). Synchronization of timing and motion among performing musicians. *Music Perception: An Interdisciplinary Journal*, *26*(5), 427–438.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). 6.2. 2.3 softmax units for multinoulli output distributions. *Deep learning*(1), 180.

Gould, N. E.-S. J., & Eldredge, N. (1972). Punctuated equilibria: an alternative to phyletic gradualism. *Essential readings in evolutionary biology*, 82–115.

Hasson, U., & Frith, C. D. (2016). Mirroring and beyond: coupled dynamics as a generalized framework for modelling social interactions. *Philosophical*

Transactions of the Royal Society B: Biological Sciences, *371*(1693), 20150366.

Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S., & Keysers, C. (2012). Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Trends in cognitive sciences*, *16*(2), 114–121.

Hennig, H. (2014). Synchronization in human musical rhythms and mutually interacting complex systems. *Proceedings of the National Academy of Sciences*, *111*(36), 12974–12979.

Herremans, D., Chew, E., et al. (2016). Tension ribbons: Quantifying and visualising tonal tension.

Hills, T. T., Todd, P. M., Lazer, D., Redish, A. D., Couzin, I. D., Group, C. S. R., et al. (2015). Exploration versus exploitation in space, mind, and society. *Trends in cognitive sciences*, *19*(1), 46–54.

Huang, C.-Z. A., Vaswani, A., Uszkoreit, J., Simon, I., Hawthorne, C., Shazeer, N., … Eck, D. (2019). Music transformer: Generating music with long-term structure. In *7th international conference on learning representations, ICLR 2019, new orleans, la, usa, may 6-9, 2019.*

Hutchings, P., & McCormack, J. (2017). Using autonomous agents to improvise music compositions in real-time. In *International conference on evolutionary and biologically inspired music and art* (pp. 114–127).

Johnson-Laird, P. N. (2002). How jazz musicians improvise. *Music Perception: An Interdisciplinary Journal*, *19*(3), 415–442.

Kassebaum, G. R. (1987). Improvisation in alapana performance: A comparative view of raga shankarabharana. *Yearbook for Traditional Music*, *19*, 45–64.

Keller, P. E., Weber, A., & Engel, A. (2011). Practice makes too perfect: Fluctuations in loudness indicate spontaneity in musical improvisation. *Music Perception*, *29*(1), 109–114.

Khalvati, K., Park, S. A., Mirbagheri, S., Philippe, R., Sestito, M., Dreher, J.-C., & Rao, R. P. (2019). Modeling other minds: Bayesian inference explains human choices in group decision-making. *Science advances*, *5*(11), eaax8783.

Konvalinka, I., Vuust, P., Roepstorff, A., & Frith, C. D. (2010). Follow you, follow me: continuous mutual prediction and adaptation in joint tapping. *Quarterly Journal of Experimental Psychology*, *63*(11), 2220–2230.

Kostka, S., Payne, D., & Almén, B. (2017). *Tonal harmony: With an introduction to post-tonal music*. McGraw-Hill Higher Education.

Koutsoupidou, T., & Hargreaves, D. J. (2009). An experimental study of the effects of improvisation on the development of children's creative thinking in music. *Psychology of music*, *37*(3), 251–278.

Kugler, P. N., & Turvey, M. T. (2015). *Information, natural law, and the self-assembly of rhythmic movement*. Routledge.

Malvinni, D. (2013). *Grateful dead and the art of rock improvisation*. Scarecrow Press.

Mathieu, W. A. (1997). *Harmonic experience: Tonal harmony from its natural origins to its modern expression*. Simon and Schuster.

Noy, L., Dekel, E., & Alon, U. (2011). The mirror game as a paradigm for studying the dynamics of two people improvising motion together. *Proceedings of the National Academy of Sciences*, *108*(52), 20947–20952.

Oore, S., Simon, I., Dieleman, S., Eck, D., & Simonyan, K. (2020). This time with feeling: Learning expressive musical performance. *Neural Computing and Applications*, *32*(4), 955–967.

Palmer, C., & Zamm, A. (2017). Empirical and mathematical accounts. *The Routledge Companion to Embodied Music Interaction*, 370.

Pearce, M. T., & Wiggins, G. A. (2012). Auditory expectation: the information dynamics of music perception and cognition. *Topics in cognitive science*, *4*(4), 625–652.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, *27*(2), 169–190.

Reynolds, C. W. (1987). Flocks, herds and schools: A distributed behavioral model. In *Proceedings of the 14th annual conference on computer graphics and interactive techniques* (pp. 25–34).

Richardson, M. J., Shockley, K., Fajen, B. R., Riley, M. A., & Turvey, M. T. (2008). Ecological psychology: Six principles for an embodied–embedded approach to behavior. In *Handbook of cognitive science* (pp. 159–187). Elsevier.

Rush, S. (2016). *Free jazz, harmolodics, and ornette coleman.* Taylor & Francis.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in cognitive sciences*, *10*(2), 70–76.

Setzler, M., & Goldstone, R. (2020). Coordination and consonance between interacting, improvising musicians. *Open Mind*, 1–14.

Setzler, M., & Goldstone, R. (2021, Oct). *Tonal emergence simulation results and analysis.* OSF. Retrieved from `osf.io/93qzp`

Solis, G., & Nettl, B. (2009). *Musical improvisation: Art, education, and society.* University of Illinois Press.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction.* MIT press.

Touma, H. H. (1971). The maqam phenomenon: An improvisation technique in the music of the middle east. *Ethnomusicology*, *15*(1), 38–48.

Valverde, S., & Solé, R. V. (2015). Punctuated equilibrium in the large-scale evolution of programming languages. *Journal of The Royal Society Interface*, *12*(107), 20150249.

Van der Schyff, D., Schiavio, A., Walton, A., Velardo, V., & Chemero, A. (2018). Musical creativity and the embodied mind: Exploring the possibilities of 4e cognition and dynamical systems theory. *Music & Science*, *1*, 2059204318792319.

Van Gelder, T. (1998). The dynamical hypothesis in cognitive science. *Behavioral and brain sciences*, *21*(5), 615–628.

Vougioukalou, S., Dow, R., Bradshaw, L., & Pallant, T. (2019). Wellbeing and integration through community music: the role of improvisation in a music group of refugees, asylum seekers and local community members. *Contemporary Music Review*, *38*(5), 533–548.

Walton, A., Richardson, M. J., & Chemero, A. (2014). Self-organization and semiosis in jazz improvisation. *International Journal of Signs and Semiotic Systems (IJSSS)*, *3*(2), 12–25.

Walton, A. E., Richardson, M. J., Langland-Hassan, P., & Chemero, A. (2015). Improvisation and the self-organization of multiple musical bodies. *Frontiers in Psychology*, *6*, 313.

Walton, A. E., Washburn, A., Langland-Hassan, P., Chemero, A., Kloos, H., & Richardson, M. J. (2018). Creating time: Social collaboration in music improvisation. *Topics in Cognitive Science*, *10*(1), 95–119.

Wickelgren, W. A., & Norman, D. A. (1966). Strength models and serial position in short-term recognition memory. *Journal of Mathematical Psychology*, *3*(2), 316–347.

Zeng, T., Przysinda, E., Pfeifer, C., Arkin, C., & Loui, P. (2017). White matter connectivity reflects success in musical improvisation. *BioRxiv*, 218024.

Zhu, H., Neubig, G., & Bisk, Y. (2021). Few-shot language coordination by modeling theory of mind. In *International conference on machine learning* (pp. 12901–12911).