Expert Musical Improvisations Contain Sequencing Biases Seen in Language Production

Roger E. Beaty[1*#], Klaus Frieler[2,3*], Martin Norgaard[4], Hannah Merseal[1], Maryellen C. MacDonald[5], & Daniel J. Weiss[1]

[1]Department of Psychology, Pennsylvania State University, USA

[2]Institute for Musicology, Franz Liszt Weimar, Germany

[3]University of Music, Franz Liszt Weimar, Germany

[4]School of Music, Georgia State University, USA

[5]Department of Psychology, University of Wisconsin-Madison, USA

*authors contributed equally

[#]corresponding author (rebeaty@psu.edu)

Author Contributions

R.B., K.F, M.N., H.M., M.M., and D.W. designed the research. K.F. performed the research and analyzed data. R.B, K.F, M.N., H.M., M.M, and D.W. wrote the paper.

Competing Interest Statement: The authors declare no competing interests.

Abstract

Language production involves action sequencing to produce fluent speech in real-time, placing a computational burden on working memory that leads to sequencing biases in production. Here we examine whether these biases extend beyond language to constrain one of the most complex human behaviors: music improvisation. Using a large corpus of improvised solos from eminent jazz musicians, we test for a production bias observed in language termed *easy first*—a tendency for more accessible sequences to occur at the beginning of a phrase, allowing incremental planning later in the same phrase. Our analysis shows consistent evidence of easy first in improvised music, with the beginning of musical phrases containing both more frequent and less complex sequences. The findings indicate that expert jazz musicians, known for spontaneous creative performance, reliably retrieve easily accessed melodic sequences before creating more complex sequences, suggesting that a domain-general sequencing system may support multiple forms of complex human behavior, from language production to music improvisation.

Keywords: Creativity; Language; Music

Spontaneous Melodic Productions of Expert Musicians Contain Sequencing Biases Seen in Language Production

Charles Darwin famously observed that "the capacity of producing musical notes…must be ranked amongst the most mysterious with which (man) is endowed" (Darwin, 1896). While Darwin was concerned with understanding the adaptive significance of music, a fundamentally similar state of uncertainty also shrouded the cognitive underpinnings of music production. Insight into both of these issues would finally arrive 80 years later when Karl Lashley suggested that all complex sequential actions require hierarchical planning (Lashley, 1951; Rosenbaum, Cohen, Jax, Weiss, & van der Wel, 2007). Thus, the underpinnings of both music and language production share common origins, both phylogenetic and cognitive, with hierarchical action plans. Although this view has generally been endorsed (Rosenbaum et al., 2007), there has been virtually no evidence supporting the notion that common cognitive processes underlie the spontaneous *production* of music and language. Instead, the abundant research on commonalities between language and music has tended to focus on comprehension and structural similarities (e.g. Patel, 2007). Studying the potential link between language and music production may yield insights into the nature of syntactic systems and the extent to which such systems are domain general. We begin to tackle this longstanding issue by investigating whether one of the cognitive biases known to influence word ordering choices in language production also manifests to constrain sequencing in the context of spontaneous music production.

Producing language requires the retrieval of words from long-term memory that match the ideas the producer wants to express, sequencing those words into a sentence, and ordering the articulatory gestures that allow the words to be articulated (Levelt et al., 1999). Like all complex action, language production is guided by an internal hierarchical plan (Rosenbaum et al., 2007), and development of the plan is thought to be incremental and concurrent, such that speakers begin executing the plan (articulating early parts of an utterance) while continuing to plan later parts (Levelt, 2008). This incremental planning is advantageous because it minimizes the amount of information that must be maintained in working memory until it can be executed. The ability to plan incrementally is most valuable when there is flexibility to order subcomponents that are ready to be articulated earlier in the utterance plan, allowing for articulation to begin and providing more time to plan portions of the utterance that are not ready to be uttered. Languages have this flexibility at some levels, including word order. Even in languages like English with fairly rigid word orders, most ideas can be expressed using many different words and word orders (Brennan, 1990; MacDonald, 2013).

Both language production experiments and analyses of spontaneous conversations show robust tendencies for speakers to utter easier words and phrases—more frequent, shorter, recently mentioned, or more salient—before harder ones, a bias termed *accessibility* or *easy first* (Bock, 1982; MacDonald, 2013). Easy first biases may also be seen at other levels; for example, consonants that are easier to articulate in the infant vocal tract tend to precede more difficult ones in the babbling sequences that infants produce (MacNeilage & Davis, 2000). Several computational accounts of motor planning, task optimization, and cognitive control contain mechanisms that would tend to generate an easy first bias in serial ordering (Botvinick & Cohen, 2014; Grossberg, 1978), suggesting that these biases are an adaptive consequence of the need to sequence behaviors in any incremental action system. Nevertheless, evidence for sequencing biases is scarce outside of language production research, and therefore there has been little attention to the potential domain generality of these biases.

Here we aim to provide a strong test of the hypothesis that sequencing biases seen in language production can also constrain *music improvisation*, one of the most complex spontaneous human actions. Our focus is on improvisation because the creation of musical sequences aligns most straightforwardly to the creation of word and phrase orders that have been intensively studied in spontaneous language production. Music improvisation and spontaneous language production also have similar time constraints, in that the performer/speaker must produce output for an audience listening in real time. Because incremental production and easy first biases are essential components of fluent language production, an extension to music improvisation forms a well-matched but distinct analog to test whether sequencing biases transcend language production.

Analyses of large corpora of spontaneous speech and written language is an important method in language research, as large corpora contain naturalistic data on a scale that is not easily replicated in a laboratory. For example, Bresnan et al. (2007) examined factors underlying speakers' use of different sentence structures in the Switchboard corpus (Godfrey, et al. 1992), which contains about 260 hours of phone conversations (about 3 million words). Bresnan et al. investigated the factors underlying speakers' use of different word orders, specifically in phrases that can appear in two different orders in English, such as *the girl* and *a book* in sentences like *I gave the girl a book* vs. *I gave a book to the girl*. They found about a dozen factors affected speakers' word orders, most of which can be seen as components of easy first biases. Some examples include the tendency to put shorter phrases before longer ones, and the tendency to put phrases with recently mentioned (and thus practiced) words earlier.

Just as words and phrases are reused in language, an influential theoretical framework asserts that musical improvisation includes the insertion of stored melodic patterns (Pressing, 1988). Through practice, these melodic patterns likely include auditory information linked to the corresponding motor movements needed to produce the sounds on the performer's particular instrument (Baumann et al., 2007). This claim is supported by corpus research of improvised jazz solos which show a high degree of pattern use (Finkelman, 1997; Norgaard, 2014; Weisberg et al., 2004). Furthermore, experimental research shows artist-level jazz improvisers use more patterns in their improvisations when their attention is diverted by a secondary task, suggesting that the use of patterns is sensitive to computational or memory burdens. Using patterns during an improvisation may thus allow musicians to achieve high virtuosity levels and to focus attention on higher-order goals, e.g., interacting with other ensemble members (Norgaard et al., 2016; Pressing, 1988). Similar to sentence planning in language production, an improvisor needs to plan ahead what to play next, which musical idea to express, and how to shape that idea with sounding tones. As this goal is primarily aesthetic and artistic, there is a higher degree of freedom for the improvisor compared to language, even though the creation of melodic material during improvisation is not completely unconstrained but is guided by stylistic and tonal rules (Johnson-Laird, 2002), which result in cognitive demands. Indeed, hierarchical relationships between notes, ideas, and longer phrases can be described using a tree structure initially developed for syntactic language analysis (Jackendoff & Lerdahl, 1982).

In the context of music improvisation, "ideas" have recently been conceptualized and formalized as midlevel units (MLU; Frieler, Pfleiderer, Zaddach, & Abeßer, 2016), which reside on a level between single notes and phrases (typical duration ~2-3 seconds). The most common MLUs are 'licks' (short and rhythmically diverse sequences) and 'lines' (long and rhythmically uniform sequences; see Figure 1). While music and language are clearly not identical, these mid-level licks and lines may be seen as similar in level to sentences (or partial sentences) in language.

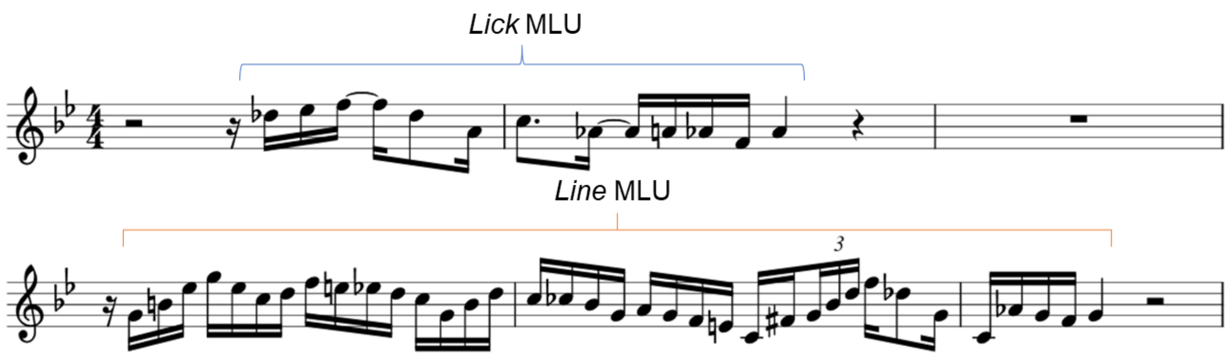

Figure 1. Examples of a lick and a line midlevel units (MLUs) from a solo by Sonny Rollins in the WJD. Lick MLUs are typically short and rhythmically diverse sequences, whereas line MLUs are longer melodic extensions that are rhythmically uniform.

Because of the different musical characteristics of licks and lines, and thus different cognitive and motoric demands, it seems natural to include licks and lines in an investigation examining easy first in music. Given the longer duration of lines, and the corresponding cognitive burden of their real-time execution, one might expect these musical ideas to show a pronounced easy first bias when produced by jazz improvisers, as they should contain a larger content of pre-configured motor patterns. By contrast, the shorter licks may not yield such tendencies. To date, however, it is unknown whether any such production biases influence complex motor behaviors like music improvisation, or if they are limited to language production.

Following in the corpus analysis tradition of language, we analyzed a corpus of several thousand improvised musical phrases, extracted from popular jazz recordings produced by eminent jazz musicians—an expert group defined by their successful spontaneous creative performance. We tested whether spontaneous improvisations follow the easy first bias observed in language production. To test for an easy first bias in music production, we examined multiple definitions of 'difficulty' analogous to those seen in language production, focusing on the production of musical ideas (i.e., licks and lines). This approach allowed us to examine whether production biases consistently seen in language influence complex creative behavior in music.

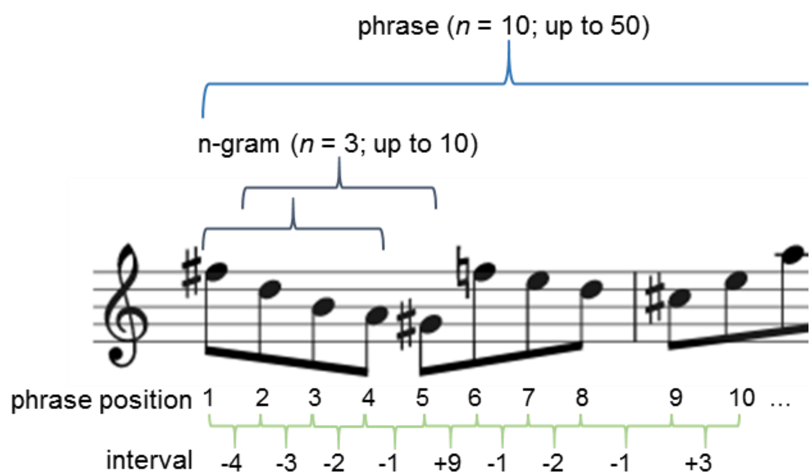

Figure 2. Example melodic improvisation phrase and corresponding metrics used for analysis. N-grams (n = 3, up to 10 notes) consist of interval patterns (differences in pitch between two notes) embedded within these longer phrases (n = 10; up to 50 notes). Phrase position provided temporal information to test the easy first hypothesis.

Method

To test for an easy first production bias in music improvisation, we analyzed the corpus data at two levels: 1) all melodic sequences ("n-grams") and 2) sequences identified as the midlevel licks and lines. We therefore extracted all n-grams (length 3 to 10 intervals, see Figure 2) and their corresponding phrase positions, as well as all sequences previously identified as licks and lines (Frieler et al., 2016a; see *Midlevel analysis*), from the Weimar Jazz Database (WJD). The WJD corpus is a collection of 456 annotated improvisation recordings[1] from well-known jazz musicians (Pfleiderer, 2017). For each n-gram and midlevel unit (lick/line), we calculated measures of within-phrase complexity and across-corpus frequency (termed *intrinsic* and *extrinsic* difficulty, respectively) against serial position within phrases of the improvisation, hypothesizing that more difficult n-grams and licks/lines would come later in the phrase (i.e., easy first). To control for effects due to stylistic convention, as well as possible statistical artifacts of the analysis process,

[1] In jazz circles, it is a strong taboo to play a pre-composed solo, except for very few situations, e.g., as part of a larger arrangement in the form of a solo break or interlude. This is further evidenced by examining alternate "takes" (i.e., recordings of the same tune), where two solos—by the same performer, on the same tune, during the same recording session—are typically distinctively dissimilar. So, the assumption that a solo is truly improvised "on the spot" is well-justified for jazz because it is a centerpiece of the jazz "ethics" or style.

we created a Markov-simulated control corpus to compare to the WJD corpus. All data and code are available at Open Science Framework (https://osf.io/svm2z/?view_only=95b8ab8ec485424399c08d8b343ee183).

Music corpus. The WJD includes solo transcriptions from recordings of renowned jazz musicians (e.g., Charlie Parker, John Coltrane, Miles Davis) covering a time-span from 1925 to 2009 (see SI Method, Figure S1). The solos were previously transcribed (manually) as pitch, onset, and duration tuples by expert transcribers using Sonic Visualiser (Cannam et al., 2010) and carefully cross-checked several times to ensure high quality (Pfleiderer, 2017). The solos in the WJD are annotated for a variety of features, such as beats, bars, metrical information, chords, and form parts. Phrases and midlevel units were also annotated manually by the transcribers. For this study, we use semitone intervals (the difference between consecutive pitches) as well as phrase and mid-level lick/line annotations. As relative distances between pitches, semitone intervals have been used in previous research on patterns in jazz (e.g., Finkelman, 1997) and reflect the fact that most humans hear musical notes relative to each other (Krumhansl, 1979), i.e., a melody or a song played in different keys is still identified as being the same. The WJD contains 200,809 tone events, which amounts to 200,353 intervals, as intervals between the end of one solo and the start of the next were not counted. There are 11,802 phrases with a mean length of 18.1 tones (median = 13, SD = 15.4). The distribution of phrase lengths is heavily right-skewed towards longer phrases (skewness = 2.5) with a range from 1 to 203. Across-phrase n-grams were excluded due to interruptions in phrases serving as a perceptual interruption in the improvisation (similar to punctuation in language).

Midlevel analysis. In addition to examining all melodic sequences in the corpus (i.e., n-grams), we also focused specifically on the production of melodic “ideas” using midlevel analysis

(MLA; Frieler et al., 2016a). MLA is a qualitative annotation system that classifies different types of aesthetic phrase forms commonly used in monophonic jazz solos. MLA aims at classifying ideas (i.e., MLUs) in jazz solos based on intrinsic features of the phrase, such as timbre, length, and rhythm. MLUs are hand-annotated with a very good interrater agreement with respect to segment boundaries (Frieler et al., 2016a). To accommodate some labeling uncertainties, all units were cross-checked by an independent rater.

Our analysis focused on two types of MLUs with theoretical links to language and relevance for easy first: *lines* and *licks* (Frieler et al., 2016a; see Figure 2). *Line* and *lick* MLUs comprise 77.2 % of all MLUs in the WJD and cover 77.1 % of all solos' duration, thus representing a majority of all midlevel units in the corpus (Frieler et al., 2016). *Lines* are longer than licks and mostly consist of extended uniform rhythmical movement in pitch space, whereas *lick*s are shorter and rhythmically more diverse, motif-like units. Hence, *line* MLUs are 19.4 notes on average (SD = 33.8) with a mean duration of 2.9 s, and *lick* MLUs are 8.3 notes on average (SD = 5.0) with a mean duration of 1.8 s (Frieler et. al., 2016: Tab. 2). The *line* puts high demands on planning and the instrumental technique of a player, as the notes must fit the given harmonies and are typically played very fast. We expect to see a much wider use of preconceived patterns in *lines* than in *licks,* since the latter have lower note-density (about 5.4 tones/s for *lick* MLUs vs. 7.7 notes/s for *line* MLUs, Frieler et al., 2016: Tab. 2) and leave more room for explicit planning. Moreover, given the cognitive burden of executing lines—long and often fast sequences, produced in real-time with little time to plan—we expected these melodic ideas to show a more pronounced easy first effect.

Difficulty measures. In order to test the hypothesis that difficulty affects order in music improvisation, it was necessary to develop measures of difficulty in musical performance. For the set of all extracted interval n-grams and licks/lines, we calculated a set of *extrinsic* and *intrinsic* difficulty measures. Extrinsic difficulty is based on the entire WJD corpus, whereas intrinsic difficulty captures features of a sequence per se. Extrinsic difficulty was quantified by the negative logarithm of occurrence probability (i.e., surprisal, information content)—a metric of difficulty also used in language—as estimated by the relative frequency of a sequence across the corpus (based on all sequences of the same length). Thus, the less frequent the sequence (n-gram, lick, or line), the higher the extrinsic difficulty value.

To calculate intrinsic difficulty, we computed several measures that reflect different conceptualizations of within-phrase complexity, using the following intuitions for interval sequences (see SI Method for computational method):

i. *On average, larger intervals are harder to play.* For a given sequence, this can be operationalized by either *mean interval size* or *maximum interval size* within the phrase.

ii. *A sequence should be more complex when it has a larger variety of intervals and a larger variety of corresponding pitches.* This can be operationalized by either the *mean interval variety* (number of different intervals) or *mean pitch variety* (number of different pitches).

iii. *A sequence should be more complex when it often changes directions.* This can be operationalized by the *mean number of direction changes* or the *mean run length* in a direction before changing (cf., Boltz, 1998).

The six intrinsic measures (interval size, maximum interval size, interval variety, pitch variety, direction changes, and run length) describe different (though not orthogonal) aspects of difficulty. Because some measures were highly correlated with one another (SI Results), and to simplify the analysis, we constructed a combined measure of intrinsic difficulty. First, we calculated all difficulty measures for all extracted n-grams, z-transformed these values for each n-gram length separately (N-scaling), and summed the values. The measures of interval size were excluded due to large-sized outliers. The sum of N-scaled variables was again N-scaled for ease of interpretation, as z-values are equivalent to Cohen's d values. Formally, we have $intrinsic\_difficulty_N(\mathrm{I}) = z_N(z_N(interval\ variety) + z_N(pitch\ variety) + z_N(direction\ changes) + z_N(run\ length))$, where z_N denotes scaling per n-gram length (N-scaling).

Control corpus. To provide a control corpus, we simulated interval sequences of 456 solos by using a first-order interval Markov model based on the WJD data. The rationale behind using a control corpus is to have a baseline to control for effects solely from typical interval and interval transition probabilities, which prefer small intervals[2]. This control corpus has exactly the same solo, phrase, and MLU annotations as the actual WJD corpus and contains the same statistical distribution of interval and interval transitions; effectively, the simulated corpus is a "sophisticated scrambling" of the order of tones in the original corpus (Herremans et al., 2017).

Results

Correlations between intrinsic and extrinsic difficulty. We began by computing correlations between the extrinsic and intrinsic measures of melodic difficulty for all sequences in the corpus to test whether the frequency of a sequence (extrinsic difficulty) tracks its complexity (intrinsic difficulty). Note that due to the large N, all correlations are statistically significant ($p < .001$). Results showed positive correlations between extrinsic and intrinsic difficulty: $r = .58$ for shorter sequences ($n = 3$) and $r = .30$ for longer sequences ($n = 10$). The same pattern of positive correlation was found between extrinsic difficulty and the individual difficulty measures that contributed to the intrinsic difficulty composite measures (SI Results). The simulated corpus

[2] In addition to controlling for small intervals in the Weimar Jazz Database (as is common in melodic corpora), the simulated corpus controls for potential random fluctuations in improvised sequences, particularly for smaller n-gram sizes (given that such corpora can be noisy, similar to language corpora). The simulated corpus also serves as a "visual null-hypothesis" to test against, allowing us to visualize and compare the difficulty curves in a dataset with the exact same interval structure.

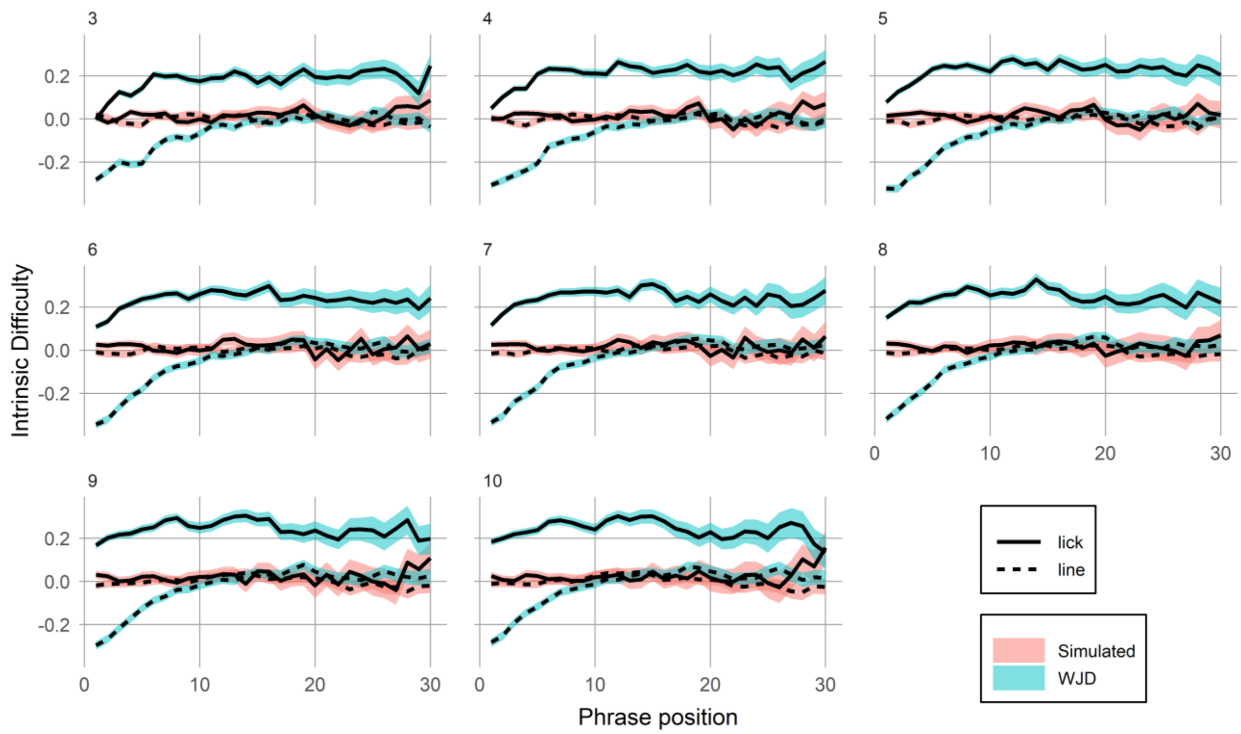

Figure 3. Plot of standardized intrinsic difficulty (y-axis) against phrase position (x-axis) in the improvisation corpus (WJD) and simulated corpus for *lick* and *line* MLUs. The eight panels correspond to the different sequence (n-gram) lengths (3 to 10); ribbons indicate standard error; color indicate corpus and line-type indicates MLU type.

showed consistently lower correlations between extrinsic and intrinsic difficulty, particularly for longer n-grams (r = *.54* for n = *3* and r = *.05* for n = 10). Critically, the simulated corpus showed that the difficulty effect cannot be explained by a general preference for smaller (i.e., easier) intervals in music (SI Results). The correlation analysis thus suggests that, overall, expert musicians tended to produce easier melodic sequences more frequently: as the complexity of those sequences decreased (lower intrinsic difficulty), the frequency of melodic sequences in the corpus increased (lower extrinsic difficulty).

Easy first analysis. We next tested the question of *when* easier sequences are likely to occur. The easy first hypothesis posits that sequence complexity should be lower at the beginning of a phrase, which would allow incremental planning of more complex sequences at later phrase positions. To visualize the effects, we created separate plots for intrinsic (Figure 3) and extrinsic

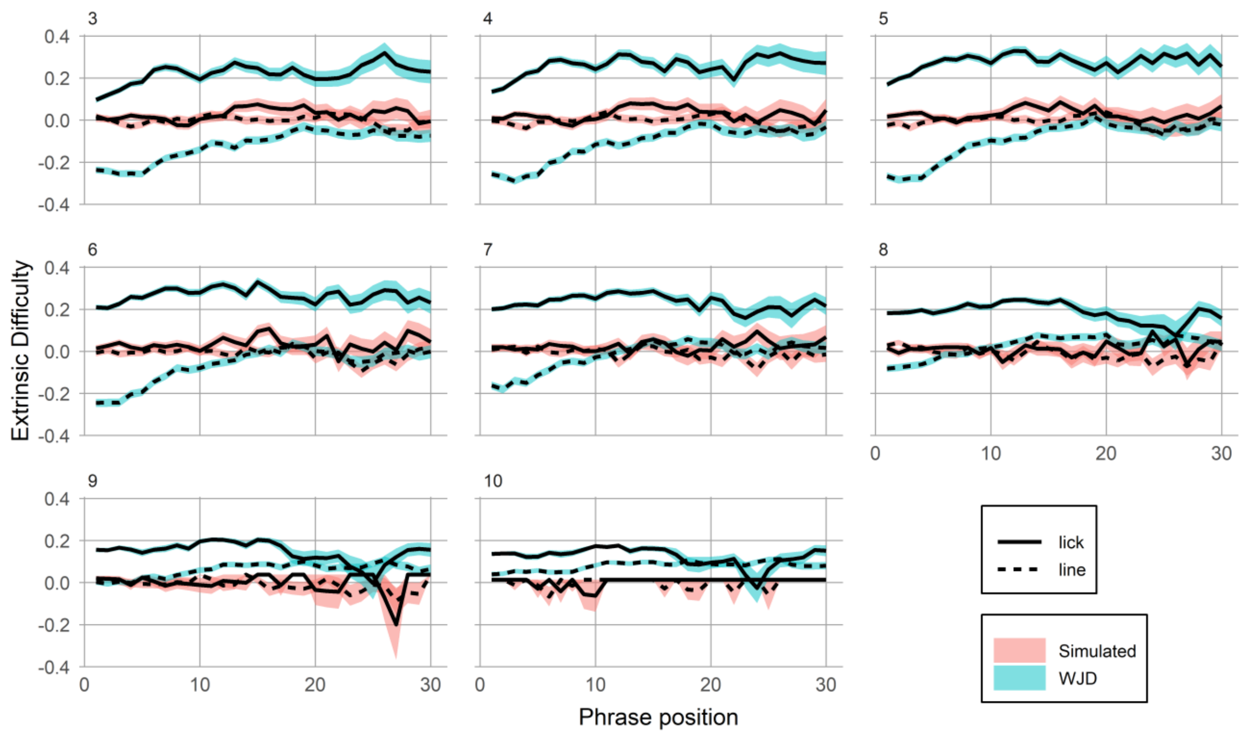

Figure 4. Plot of standardized extrinsic difficulty (y-axis) against phrase position (x-axis) in the improvisation corpus (WJD) and simulated corpus for *lick* and *line* MLUs. The eight panels correspond to the different sequence (n-gram) lengths (3 to 10); ribbons indicate standard error; color indicate corpus and line-type indicates MLU type.

difficulty (Figure 4), combining n-gram lengths 3 to 10 (panels), corpus type (colors), and MLU types (licks and lines; solid and dashed lines), with phrase position on the x-axes and difficulty values on the y-axes. We hypothesized that *line* MLUs (longer and faster musical ideas) would show more pronounced easy first effects than *lick* MLUs (shorter and rhythmically diverse ideas), given the higher cognitive burden of lines on motor execution.

For intrinsic difficulty (Figure 3), *line* MLUs showed consistent evidence of easy first that is robust across n-gram length (3 to 10); for *lick* MLUs, similar but less consistent easy first trends can be seen. Notably, the curves for *lick* and *line* MLUs are strongly separated for both metrics, with *line* MLUs showing overall easier n-grams than *lick* MLUs (cf. Figures S3 and S4). For *line* MLUs, the mean intrinsic difficulty values increase monotonically from the first phrase position to the mean around phrase position 10 to 15, depending on n-gram length. For *lick* MLUs, greater overall difficulty was observed, irrespective of easy first trends[3]. Critically, the simulated data in the control corpus is similar for both types of MLUs, and the difficulty measures remain relatively flat around the mean. This result demonstrates an easy first effect in improvised music when music sequences are defined by intrinsic difficulty (i.e., complexity), similar to the classic easy first effect seen in language production.

For extrinsic difficulty (Figure 4), a similar easy first pattern can be observed, particularly for shorter n-gram lengths. For n-gram lengths greater 8, the easy first effect for extrinsic difficulty essentially disappears, likely because longer n-grams are less likely to be played more than once by musicians in the corpus. For *lick* MLUs, and for both difficulty measures, a small and fast-rising easy first effect can be observed, with mean difficulties increasing from the phrase beginning to about the seventh or eighth position before attenuating. Thus, an easy first effect was also found for extrinsic difficulty (frequency), but the effect was specific to *line* MLUs (longer, faster melodic ideas) up to an n-gram length of 8.

Notably, the plots suggest a nonlinear relationship between difficulty and phrase position. This nonlinear relationship was confirmed, as a polynomial regression analysis over the mean difficulty values per phrase position showed a consistently better fit, with convex shapes that rise at the beginning of phrases compared to linear models (Table S1); due to the large amount of data, all regression models were significant at $p < .001$. Critically, the simulated data did not show easy first effects for both intrinsic (Figure 3) and extrinsic difficulty (Figure 4). Slopes in the simulated data were largely flat and hovered around the midpoint of zero, providing evidence that the

[3] Higher difficulty for licks compared to lines relates to the MLU classification system. Lines are faster than licks, and thus require reusable patterns to be executed, yielding low extrinsic and intrinsic difficulty. In contrast, licks may be more improvisatory because they allow more (cognitive) space for ornamentation, without a need to rely on pre-learned patterns. However, note the absolute effect size for licks is fairly small (~.2), so the difference might appear larger due this systematic trend in the corpus.

difficulty effects in the real data were not due to a preference for small-step motion. To further examine the easy first effects, we ran a set of linear regressions (i.e., trend analyses) on averaged difficulties against increasing maximal phrase positions, showing that the linear trend (as measured by the beta coefficient) is maximal at the beginning position and then gradually declines, corroborating the visual impression from Figs. 3 and 4 showing an increasing trend that saturates after ~8 to 10 notes, depending on N-gram length (see SI Trend Analysis, Figure S4 and S5 for more details).

Effect sizes. To estimate effect sizes, we used three different measures: 1) the *range of difficulty values* (difference between maximum and minimum values) across the entire set of phrase positions, with range measured in standard deviation units akin to Cohen's d; 2) the *difficulty values* at the very phrase beginnings; and 3) the *mean difficulty* across all phrase positions. Full results can be found in Table S2.

For *line* MLUs, we found ranges between 0.351 and 0.391 for intrinsic difficulties and between 0.07 and 0.298 for extrinsic difficulties. Mean and first position difficulties were consistently negative, demonstrating a robust easy first effect. For *lick* MLUs, the values range between –0.138 and –0.249 for intrinsic difficulties and between –0.123 and –0.223 for extrinsic difficulties. Mean and first position difficulties were consistently positive (close to zero for first position and small N). Thus, *line* MLUs for intrinsic difficulty show the largest effect size, consistent with the cognitive demand of planning and executing longer and faster melodic sequences in real-time.

Discussion

In language production, sequencing biases that alleviate memory constraints create statistical regularities in patterns of word and phrase difficulty in word orders (MacDonald, 2013).

Analyses of these patterns lend critical insights into the architecture of language production, as well as language comprehension (MacDonald, 2013). Here we explored whether one of several biases hypothesized to constrain language production can also provide insights into musical improvisation, often considered one of the most complex creative behaviors. Specifically, we sought evidence that the *easy first* bias observed in language production, in which more easily accessible sequences tend to be produced before more complex sequences, might also constrain sequences produced in the context of spontaneous jazz improvisation. Indeed, we found that the easy first bias extends beyond language production to constrain artist-level, creative musical performance.

In large-scale corpus analysis of several hundred recordings from eminent jazz musicians, we found that musicians tended to produce easier melodic sequences—defined both as more frequent and less melodically complex—at the beginning of improvised phrases. This was particularly true when different types of musical ideas (*line* and *lick* midlevel units) were taken into consideration (cf. Figure S2 and S3 for global effects). Critically, simulated data with the same interval structure did not show this pattern, suggesting that the easy first effects reflect production biases for improvised music and are not simply a result of a preference for smaller (and thus easier) intervals, or other stylistic conventions of jazz music. Taken together, our results indicate that similar production biases may constrain the spontaneous production of both language and music.

To our knowledge, this is the first evidence of a domain-general sequencing bias influencing spontaneous melodic production, which has largely been documented in language production (MacDonald, 2013). In language, sequencing biases such as easy first are thought to minimize the computational burden of language production via incremental planning. Beginning with more easily accessible words and phrases allows the speaker to simultaneously begin producing easy material and continue planning more complex material, freeing working memory and allowing speech to initiate more quickly than if the complex material had to be planned first. Such temporal constraints similarly affect musical improvisation, which requires the real-time expression of aesthetically-appealing music with minimal time to plan (Pressing, 1984, 1988). In improvised jazz, musicians must meet the cognitive and physical constraints of performing fast and complex melodies. Having expert knowledge of common motor patterns (melodic sequences) can minimize these computational burdens (Norgaard, 2011, 2014; Norgaard et al., 2016; Pressing, 1984, 1988), allowing the improviser to focus on other aspects of performance and plan more novel and interesting melodic sequences.

Our finding of easy first also extends prior research on pattern analysis showing recurring melodic sequences in improvised musical performance (Norgaard, 2014; Weisberg et al., 2004) as well as an influential theory of improvisation emphasizing the importance of schemas in automating aspects of performance to meet the temporal, physical, and psychological constraints of improvisation (Pressing, 1988). The present work also extends previous corpus-based analyses (e.g., Finkelman, 1997) by considering overlapping patterns in melodic sequences, circumventing the problem of defining discrete melodic units (within continuous strings of notes) using arbitrary criteria (Norgaard, 2014). We also extend previous corpus work by examining positions of n-grams within individual phrases, allowing the detection of fine-grained temporal trends in improvised performance. Taken together, the findings suggest that high-level creative performance is partly facilitated by domain-general sequencing biases that also support language production, and that the substantial physical, temporal, and psychological constraints of spontaneous creativity can be mitigated by first producing less complex and easily accessible melodic sequences. Thus, an insight from the language production literature (i.e., easy first) reveals constraints on spontaneous musical creativity.

An important step for future research is investigating whether similar easy first patterns are found in other music corpora. For example, Bresnan et al. (2007) investigated whether the patterns of word order found in the phone conversations of the Switchboard corpus were replicated in a corpus of newspaper articles. They found very similar word order biases across these quite different language genres, and it will similarly be interesting to investigate whether different musical genres replicate the patterns seen here—for example, whether easy first is seen in folk songs that were originally communicated orally prior to the invention of musical notation.

The notion of common constraints in planning for spontaneous language and music production also finds support in production studies that focus on incremental planning of memorized sequences (e.g., Palmer & Pfordresher, 2003). Unlike the present study's focus on spontaneous creation of music and its relationship to spontaneous language production, the goal of this incremental planning research has been to identify parallels in patterns of errors and interference across domains in the context of well-practiced patterns of speech and music, such as exploring how producers monitor feedback on their (speech or musical) articulatory accuracy (e.g., Keller & Koch, 2008; Levelt, 1983). Consistent with our claims of a common planning bias across spontaneous language and music, so too there is evidence for a similar time course of planning and interference effects in memorized sequences spanning these domains (Mathias et al., 2019; Palmer & Pfordresher, 2003). Likewise, a common anticipatory bias in sequencing errors has been found in both language and music (Dell, Burger, & Svec, 1997; Drake & Palmer, 2000). Together, these lines of research lend evidence in support of Lashley's (1951) assertion that complex sequential behaviors, be they spontaneously generated or well-practiced, are governed by hierarchical plans.

It is worth noting that easy first is not the only bias that shapes the serial order of words and phrases in language production. Other sequencing biases, including *plan reuse*, operate on the level of syntax, promoting repetition of higher-order linguistic structure (Bock, 1982; MacDonald, 2013). The present work establishes the existence of one sequencing bias in music production, but whether and how other known biases such as plan reuse influence melodic sequencing remains an open question. Another key question for future research concerns the role of production biases in shaping perception/comprehension. According to the Production-Distribution-Comprehension model, sequencing biases such as easy first shape language perception through the statistical regularities created by speakers following the biases: the more a listener perceives language with an easy first structure over their lifetime, the easier such structures will be to comprehend, leading to greater preferences for sequences with easy first ordering (MacDonald, 2013). To what extent does such statistical learning influence musical preferences? One possibility is that, given the high frequency of the easy first pattern in music improvisation, expert musicians may develop an expectation of, or a preference for, melodic sequences that follow easy first (and other sequencing biases).

More generally, our findings promote further investigations of other domains with respect to sequencing biases that align with linguistic biases, with potential implications for distinguishing uniquely human behaviors like music and language from behaviors found in other species. Sequencing biases such as easy first are well-documented in language (Bock, 1982; Bresnan et al., 2007; MacDonald, 2013), and researchers have begun to speculate about their relevance in other types of actions, such as reaching, grasping, and more complex motor behaviors that are shared with other species (Rosenbaum, Chapman, Weigelt, Weiss, & van der Wel, 2012; MacDonald & Weiss, 2017). To our knowledge, the current study is among the first empirical investigations testing the easy first bias outside of language, and further work is needed to establish the scope of such biases. Sequencing biases may be limited to producing highly complex hierarchical structures like language and music; if so, they could be seen as a hallmark of uniquely human behaviors. Alternatively, sequencing biases may be part of a domain-general "syntax of action" (Rosenbaum et al., 2007), not limited to language and music. Further pursuit of the degree of sequencing commonalities in music and language, and in action more broadly, may further advance our understanding of the cognitive and evolutionary underpinnings of an essential component of human behavior.

References

Baumann, S., Koeneke, S., Schmidt, C. F., Meyer, M., Lutz, K., & Jancke, L. (2007). A network for audio-motor coordination in skilled pianists and non-musicians. *Brain Research, 1161,* 65–78. https://doi.org/10.1016/j.brainres.2007.05.045

Berliner, P. F. (1994). *Thinking in Jazz. The Infinite Art of Improvisation.* University of Chicago Press.

Bock, J. K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review*, *89*(1), 1–47. https://doi.org/10.1037/0033-295X.89.1.1

Boltz, M. G. (1998). Tempo discrimination of musical patterns: Effects due to pitch and rhythmic structure. *Perception & Psychophysics, 60*, 1357-1373. https://doi.org/10.3758/BF03207998

Botvinick, M. M., & Cohen, J. D. (2014). The computational and neural basis of cognitive control: Charted territory and new frontiers. *Cognitive Science*, *38*(6), 1249–1285. https://doi.org/10.1111/cogs.12126

Brackett, D., & DeVeaux, S. (1998). The Birth of Bebop: A Social and Musical History. *Notes*, *55*(1), 74. https://doi.org/10.2307/900349

Brennan, S. E. (1990). Conversation as direct manipulation: An iconoclastic view. *The art of human-computer interface design*, 393–404.

Cannam, C., Landone, C., & Sandler, M. (2010). Sonic visualiser: An open source application for viewing, analysing, and annotating music audio files. *MM'10 - Proceedings of the ACM Multimedia 2010 International Conference*, 1467–1468. https://doi.org/10.1145/1873951.1874248

Darwin, C. (1896). *The descent of man and selection in relation to sex* (Vol. 1). D. Appleton.

Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review*, *104*(1), 123-147. https://doi.org/10.1037/0033-295X.104.1.123

Drake, C., & Palmer, C. (2000). Skill acquisition in music performance: relations between planning and temporal control. *Cognition*, *74*(1), 1-32. https://doi.org/10.1016/S0010-0277(99)00061-X

Finkelman, J. (1997). Charlie Christian and the role of formulas in jazz improvisation. *Jazzforschung/Jazz Research*, *29*, 159–188.

Frieler, K. (2018). A Feature History of Jazz Solo Improvisation. In W. Knauer (Ed.), *Jazz @ 100. An alternative to a story of heroes* (Vol. 15). Wolke Verlag.

Frieler, K., Pfleiderer, M., Abeßer, J., & Zaddach, W.-G. (2016a). Midlevel analysis of monophonic jazz solos. A new approach to the study of improvisation. *Musicae Scientiae*, *20*(2), 143–162. https://doi.org/10.1177/1029864916636440

Frieler, K., Pfleiderer, M., Abeßer, J., & Zaddach, W.-G. (2016b). "Telling a story“. On the dramaturgy of monophonic jazz solos. *Empirical Musicology Review*, *11*(1). http://emusicology.org/article/view/4959

Grossberg, S. (1978). A Theory of Human Memory: Self-Organization and Performance of Sensory-Motor Codes, Maps, and Plans. In *Progress in Theoretical Biology* (S. 233–374). https://doi.org/10.1016/b978-0-12-543105-7.50013-0

Herremans, D., Chuan, C. H., & Chew, E. (2017). A functional taxonomy of music generation systems. *ACM Computing Surveys (CSUR)*, *50*(5), 1-30. https://doi.org/10.1145/3108242

Jackendoff, R., & Lerdahl, F. (1982). A grammatical parallel between music and language. In *Music, mind, and brain* (pp. 83-117). Springer, Boston, MA. https://doi.org/10.1007/978-1-4684-8917-0_5

Keller, P. E., & Koch, I. (2008). Action planning in sequential skills: Relations to music performance. *Quarterly Journal of Experimental Psychology*, *61*(2), 275–291. https://doi.org/10.1080/17470210601160864

Krumhansl, C. L. (1979). The psychological representation of musical pitch in a tonal context. *Cognitive Psychology*, *11*, 346–374. https://doi.org/10.1016/0010-0285(79)90016-1

Lashley, K. S. (1951). The problem of serial order in behavior. In *Cerebral mechanisms in behavior*. Bobbs-Merrill Oxford, United Kingdom.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, *14*(1), 41–104. https://doi.org/10.1016/0010-0277(83)90026-4

Levelt, W. J. M. (2008). The architecture of normal spoken language use. In *Linguistic Disorders and Pathologies: An International Handbook* (S. 1–15). De Gruyter Mouton.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1–75. https://doi.org/10.1017/S0140525X99001776

London, J. (2004). *Hearing in Time: Psychological Aspects of Musical Meter*. Oxford University Press.

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, *4*, 26. https://doi.org/10.3389/fpsyg.2013.00226

MacNeilage, P. F., & Davis, B. L. (2000). On the origin of internal structure of word forms. *Science*, *288*(5465), 527–531. https://doi.org/10.1126/science.288.5465.527

Mathias, B., Gehring, W. J., & Palmer, C. (2019). Electrical brain responses reveal sequential constraints on planning during music performance. *Brain Sciences*, *9*(2). https://doi.org/10.3390/brainsci9020025

Norgaard, M. (2011). Descriptions of improvisational thinking by artist-level jazz musicians. *Journal of Research in Music Education*, *59*(2), 109–127. https://doi.org/10.1177/0022429411405669

Norgaard, M. (2014). How Jazz musicians improvise: The central role of auditory and motor patterns. *Music Perception*, *31*(3), 271–287. https://doi.org/10.1525/MP.2014.31.3.271

Norgaard, M., Emerson, S. N., Dawn, K., & Fidlon, J. D. (2016). Creating under pressure: Effects of divided attention on the improvised output of skilled jazz pianists. *Music Perception*, *33*(5), 561–570. https://doi.org/10.1525/MP.2016.33.5.561

Palmer, C., & Pfordresher, P. Q. (2003). Incremental planning in sequence production. *Psychological Review*, *110*, 683–712. https://doi.org/10.1037/0033-295X.110.4.683

Patel, A. D. (2007). Music, language, and the brain. New York: Oxford University Press.

Pfleiderer, M. (2017). *Inside the Jazzomat: New Perspectives for Jazz Research* (M. Pfleiderer, K. Frieler, J. Abesser, W.-G. Zaddach, & B. Burkhart, Eds.). Schott Campus.

Pressing, J. (1984). Cognitive Processes in Improvisation. *Advances in Psychology*, *19*(C), 345–363. https://doi.org/10.1016/S0166-4115(08)62358-4

Pressing, J. (1988). Improvisation: Methods and models. In J. A. Sloboda (Ed.), *Generative processes in music* (S. 129–178). Oxford.

Rosenbaum, D. A., Chapman, K. M., Weigelt, M., Weiss, D. J., & van der Wel, R. (2012). Cognition, action, and object manipulation. *Psychological Bulletin*, *138*(5), 924–946. https://doi.org/10.1037/a0027839

Rosenbaum, D. A., Cohen, R. G., Jax, S. A., Weiss, D. J., & van der Wel, R. (2007). The problem of serial order in behavior: Lashley's legacy. *Human Movement Science*, *26*(4), 525–554. https://doi.org/10.1016/j.humov.2007.04.001

Weisberg, R. W., Brinkman, a R., Folio, C. J., Dick, a S., Fleck, J. I., Niederberger, B., & Bassett, F. (2004). Toward a cognitive analysis of creativity: Improvisation in jazz. *Proceedings of the Conference on Interdisciplinary Musicology Annekatrin Kessler and Frank Zimmer Graz Austria Conference on Interdisciplinary Musicology*, 1–9.