Journal of Memory and Language

Global Matching in Music Familiarity: How Musical Features Combine across Memory Traces to Increase Familiarity with the Whole in which they are Embedded
--Manuscript Draft--

Corresponding Author:	Katherine L. McNeely-White Colorado State University Fort Collins, CO United States
First Author:	Katherine L. McNeely-White, M.S.
Order of Authors:	Katherine L. McNeely-White, M.S.
	David G. McNeely-White, M.S.
	Anne M. Cleary, Ph.D.

| Abstract: | Across three experiments, the present study tested specific hypotheses regarding how the encoding of different isolated musical feature-types (rhythm and pitch) affects perceived familiarity with later piano song clips in which those features are embedded. The results are broadly consistent with global matching approaches to the computation of the familiarity signal for a musical piece. Further, the results suggest that during the global matching process, the feature-match assessments of separately encoded instances of isolated rhythm and isolated pitch information combine additively across memory traces to increase the familiarity of the test song clip in which they are embedded. As we show through simulations, this additive combination of feature-match levels across memory traces adheres to an assumption present in the MINERVA 2 model of familiarity signal computation. |

Katherine L. McNeely-White, M.S.
Department of Psychology
Colorado State University

PHONE (970) 491-7701
FAX (970) 491-1032
E-MAIL Kat.White@colostate.edu

Colorado
State
University

Department of Psychology
1876 Campus Delivery
Fort Collins, CO 80523-1876
970-491-6363 Phone
970-491-1032 Fax

Dear Dr. Rastle:

Thank you for the opportunity to revise this manuscript for JML. We are grateful for the latest round of suggestions from both you and the reviewer, as we feel that making these changes has dramatically improved the paper and increased its potential theoretical impact. As we realize that this is a somewhat lengthy paper, we not only include the page numbers of our responses to the reviewer suggestions and concerns, but we highlight these sections in the manuscript as well (in yellow) to make it easier to find them.

The largest, most substantial change to the paper is our inclusion of additional simulations at the suggestion of the reviewer. Specifically, we now report simulations that better align with our data. The way that we incorporated these additional simulations into the paper is by keeping our initial simulations in the introduction, but clarifying in this revision that those simulations are proof of principle simulations, intended to illustrate that the additivity principle exists within the MINERVA 2 model's mechanisms with respect to mean echo intensities (see pages 23-24). This proof of principle demonstration then helps to lead into our experiments, where we then turn to the simulations of our actual data. These simulations of our actual data can be found in the results sections of Experiment 2 (pages 57-63) and Experiment 3 (pages 69-71). We followed the reviewer's suggestion to build baseline familiarity into the model (via pre-existing knowledge traces) and that strategy was effective at simulating our approximate pattern of results in each experiment, as described on the aforementioned pages and also as shown in Figures 7 and 9.

With the addition of these new simulations, we felt that the assistance of David G. McNeely-White, who we previously thanked in the author note for assisting with our first round of

simulations, now warranted adding him as a co-author with this second round of simulations. David is a computer science PhD student whose assistance with the simulation code and the simulation runs was invaluable. All authors approve of the current authors and author order.

In response to another of the Reviewer's suggestions, we now mention in the General Discussion the limitation of low discriminability in our data, and the possibility that low discrimination can force a linear pattern. We note that having our simulations set at slightly higher levels of discrimination helped to address this at least somewhat within the simulations of the model's yes-no responses but that future research should aim to search for additivity in situations of higher discrimination. This discussion occurs on pages 77-79 along with some other insights obtained from the new simulations, such as the possible importance of the variances in whether or not additivity will be shown in the model's yes-no responses.
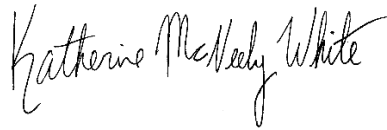
The reviewer also suggested that we make clearer the fact that trials were only removed if they were identified at study (not at test). We now make this clearer on page 43 where we say "Based on these criteria, songs were later separated into songs identified-at-study and songs unidentified-at-study for later analysis." (And also on page 51 and on 65 where we use similar clarifying wording). We also came to realize while addressing this comment that our method sections for Experiments 2 (pp. 50-51) and 3 (pp. 64-65) could benefit from the inclusion of more detail, which we have now added.

We are also extremely grateful to you for checking our data and noting a few minor discrepancies when attempting to re-run our analyses. We checked through the data and indeed were surprised to find some minor discrepancies. These have now been corrected in the manuscript and do not change the patterns. We suspect that the minor discrepancies may have reflected rounding differences when copying and pasting from different files (such as from Excel to SPSS), but in any case, these have now been corrected so that the data on the OSF are exactly in line with the analyses should anyone try to reproduce the analyses.

We have also now made further changes at your suggestion, including making the stimuli available on the OSF, making the abstract more concise, and making sure the manuscript adheres to APA style and that our figures are publication quality.

Thank you again for this opportunity to revise the manuscript and for the many helpful suggestions for improving it. We feel that this manuscript has a lot of potential to advance theory and current theoretical thinking.

Sincerely,

Katherine L. McNeely-White
Cognitive Psychology Graduate Student

Katherine L. McNeely-White, M.S.
Department of Psychology
Colorado State University

PHONE (970) 491-7701
FAX (970) 491-1032
E-MAIL Kat.White@colostate.edu

Department of Psychology
1876 Campus Delivery
Fort Collins, CO 80523-1876
970-491-6363 Phone
970-491-1032 Fax

Dear Dr. Rastle:

Thank you for the opportunity to revise this manuscript for JML. We are grateful for the latest round of suggestions from both you and the reviewer, as we feel that making these changes has dramatically improved the paper and increased its potential theoretical impact. As we realize that this is a somewhat lengthy paper, we not only include the page numbers of our responses to the reviewer suggestions and concerns, but we highlight these sections in the manuscript as well (in yellow) to make it easier to find them.

The largest, most substantial change to the paper is our inclusion of additional simulations at the suggestion of the reviewer. Specifically, we now report simulations that better align with our data. The way that we incorporated these additional simulations into the paper is by keeping our initial simulations in the introduction, but clarifying in this revision that those simulations are proof of principle simulations, intended to illustrate that the additivity principle exists within the MINERVA 2 model's mechanisms with respect to mean echo intensities (see pages 23-24). This proof of principle demonstration then helps to lead into our experiments, where we then turn to the simulations of our actual data. These simulations of our actual data can be found in the results sections of Experiment 2 (pages 57-63) and Experiment 3 (pages 69-71). We followed the reviewer's suggestion to build baseline familiarity into the model (via pre-existing knowledge traces) and that strategy was effective at simulating our approximate pattern of results in each experiment, as described on the aforementioned pages and also as shown in Figures 7 and 9.

With the addition of these new simulations, we felt that the assistance of David G. McNeely-White, who we previously thanked in the author note for assisting with our first round of

simulations, now warranted adding him as a co-author with this second round of simulations. David is a computer science PhD student whose assistance with the simulation code and the simulation runs was invaluable. All authors approve of the current authors and author order.

In response to another of the Reviewer's suggestions, we now mention in the General Discussion the limitation of low discriminability in our data, and the possibility that low discrimination can force a linear pattern. We note that having our simulations set at slightly higher levels of discrimination helped to address this at least somewhat within the simulations of the model's yes-no responses but that future research should aim to search for additivity in situations of higher discrimination. This discussion occurs on pages 77-79 along with some other insights obtained from the new simulations, such as the possible importance of the variances in whether or not additivity will be shown in the model's yes-no responses.
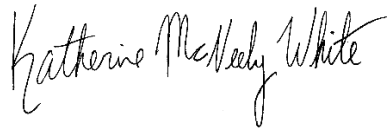
The reviewer also suggested that we make clearer the fact that trials were only removed if they were identified at study (not at test). We now make this clearer on page 43 where we say "Based on these criteria, songs were later separated into songs identified-at-study and songs unidentified-at-study for later analysis." (And also on page 51 and on 65 where we use similar clarifying wording). We also came to realize while addressing this comment that our method sections for Experiments 2 (pp. 50-51) and 3 (pp. 64-65) could benefit from the inclusion of more detail, which we have now added.

We are also extremely grateful to you for checking our data and noting a few minor discrepancies when attempting to re-run our analyses. We checked through the data and indeed were surprised to find some minor discrepancies. These have now been corrected in the manuscript and do not change the patterns. We suspect that the minor discrepancies may have reflected rounding differences when copying and pasting from different files (such as from Excel to SPSS), but in any case, these have now been corrected so that the data on the OSF are exactly in line with the analyses should anyone try to reproduce the analyses.

We have also now made further changes at your suggestion, including making the stimuli available on the OSF, making the abstract more concise, and making sure the manuscript adheres to APA style and that our figures are publication quality.

Thank you again for this opportunity to revise the manuscript and for the many helpful suggestions for improving it. We feel that this manuscript has a lot of potential to advance theory and current theoretical thinking.

Sincerely,

Katherine L. McNeely-White
Cognitive Psychology Graduate Student

- Familiarizing isolated musical features increased subsequent familiarity for songs

- Separately encoded rhythm and pitch sequences increased familiarity additively

- Multiple exposures to a rhythm increased song familiarity additively

- An additivity principle was found in MINERVA 2's familiarity computations

- Music feature familiarization across traces exhibited MINERVA 2-type additivity

*Conflict of Interest

The authors declare that we have no conflict of interest.

1
2
3
4
5
6
7
8
9
10   Running Head: MUSICAL FEATURES
11
12
13
14
15
16
17
18
19
20
21
22
23   **Global Matching in Music Familiarity: How Musical Features Combine across Memory Traces**
24
25   **to Increase Familiarity with the Whole in which they are Embedded**
26
27
28   Katherine L. McNeely-White, David G. McNeely-White, & Anne M. Cleary
29
30
31   Colorado State University
32
33
34
35
36   Correspondence should be addressed to:
37
38   Katherine L. McNeely-White
39
40   Department of Psychology
41
42   Colorado State University
43
44
45   1976 Campus Delivery
46
47   Fort Collins, CO 80523-1876
48
49
50   Phone: (970) 491-7701
51
52   Fax: (970)-491-1032
53
54
55   Email: Kat.White@colostate.edu
56
57
58
59
60
61
62
63
64
65

Abstract

Across three experiments, the present study tested specific hypotheses regarding how the

encoding of different isolated musical feature-types (rhythm and pitch) affects perceived

familiarity with later piano song clips in which those features are embedded. The results are

broadly consistent with global matching approaches to the computation of the familiarity signal

for a musical piece. Further, the results suggest that during the global matching process, the

feature-match assessments of separately encoded instances of isolated rhythm and isolated

pitch information combine additively across memory traces to increase the familiarity of the

test song clip in which they are embedded. As we show through simulations, this additive

combination of feature-match levels across memory traces adheres to an assumption present

in the MINERVA 2 model of familiarity signal computation.

Key words: Familiarity-detection; Familiarity; Music Cognition; Features; Global Matching

Global Matching in Music Familiarity: How Musical Features Combine across Memory Traces

to Increase Familiarity with the Whole in which they are Embedded

The Feature Assumption in Memory Research

A central question in cognitive psychology concerns how information is represented in

the human mind, and how these representations contribute to cognitive processes. A long-held

assumption among memory researchers is the idea that memory traces—often referred to as

engrams—consist of "feature bundles," or collections of independent elements from the

original encoding episode that are connected together to form the record of the experience

(Tulving & Watkins, 1975; Estes, 1950). The assumption that memory traces are sets of tied-

together features from the original experience has even been described by some as a "basic

pretheoretical assumption" in memory theory (Tulving & Bower, 1974, p. 269):

> "A rather general and atheoretical conception of the memory trace of an
> event regards it as a collection of features or a bundle of information. This
> view has been proposed and elaborated by many writers (e.g., Anisfeld &
> Knapp, 1968; Bower, 1967; Bregman & Chambers, 1966; Underwood, 1969;
> Wickens, 1970) and is now generally accepted as one of the basic
> pretheoretical assumptions."

The feature assumption is apparent in most mathematical models of human memory (Nairne,

1990; Plaut, 1995; Seidenberg, 2007; Smith, Shoben, & Rips, 1974), and is especially so in

models of recognition memory (Brockdorff & Lamberts, 2000; Clark & Gronlund, 1996; Cox &

Shiffrin, 2017; Hintzman, 1988; McClelland & Chappell, 1998; Shiffrin & Steyvers, 1997).

A logical question to follow the feature assumption is: What is a feature? And, following from

this, how is it exactly that a feature can contribute to a memory process?

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

What is a Feature? The Systematic Identification of Features

Some of the most direct evidence for a role of separable features in the memory trace

for a previously experienced event comes from research on familiarity-detection during

retrieval failure. The task that is used to focus on familiarity-detection during retrieval failure

involves either isolating features from a studied item and using those isolated features as a test

cue (e.g., Cleary & Greene, 2000, 2001; Cleary, Langley, & Seiler, 2004; Cleary, Ryals, & Wagner,

2016; Cleary, Winfield, & Kostic, 2007; Kostic & Cleary, 2009), or manipulating the potential

feature overlap between a novel holistic test cue and a studied item (e.g., Cleary, 2004; Cleary,

Ryals & Nomi, 2009; Cleary et al., 2012; Ryals & Cleary, 2012; Ryals, Cleary, & Seger, 2013). The

focus is on the ability to detect familiarity with the cue during instances of retrieval failure

(failure to recall the study item to which the cue corresponds). Although the participant cannot

specifically retrieve the target to which the cue corresponds, a sense of familiarity signals to the

participant that the cue corresponds to something held within memory. This method has been

used to systematically identify many different types of features that are present within memory

traces (see Cleary, 2014, for a review). Many of the identified feature-types are described

below.

Geons. Cleary et al. (2004) isolated geometric shape features (geons) using picture

fragments that came from potentially studied black and white line drawings. The method used

for isolating geons stemmed from Biederman's (1987) approach of leaving junction points

intact in the fragments to allow for some extraction of geon information from the fragment.

Participants could discriminate between unidentified geometric fragments of studied drawings

and unidentified geometric fragments of unstudied drawings, but only when geon information

was included in the fragments. In a comparison condition in which the same number of pixels

from the original line drawing were present (10%) but no junction points were included, no

ability to discriminate unidentified studied from unidentified unstudied fragments occurred.

This pattern suggests that geons—basic geometric shapes found in everyday objects—are a

type of feature retained in memory traces for recently studied images.

Spatial Configuration. Cleary et al. (2009) showed that the overall configuration of

elements within a scene—or its overall gestalt—can actually be a type of feature retained in

memory traces for recently viewed events. Using the feature overlap method, Cleary et al.

showed that participants could discriminate between novel test drawings (of new unstudied

scenes) that configurally mapped onto unrecalled studied scenes and test drawings that did

not.

Cleary et al. (2012) later showed this pattern in an immersive virtual reality situation

with scenes. Novel scenes that spatially mapped onto earlier viewed but unrecalled scenes in

their configuration of elements were found to be more familiar than novel scenes that did not

map onto earlier viewed scenes. This pattern has since been obtained in a number of studies

involving the first-person perspective through virtual tours (Cleary & Claxton, 2018; Cleary,

McNeely-White, Huebert, & Claxton, 2018; Cleary, Huebert, McNeely-White, & Spahr, 2019).

Overall, the pattern suggests that the particular spatial arrangement of elements on a grid

constitutes a type of feature for an experienced event that is retained in the memory trace for

that event.

Graphemic Wordform Features. By manipulating feature overlap between a test cue

(e.g., bashful) and a potentially studied item (e.g., bushel), Cleary (2004) demonstrated that

when cued recall failed (e.g., *bashful* failed to be recalled in response to the cue *bushel*),

participants were still able to discriminate between cues resembling and not resembling

studied words. This finding of an ability to detect increased familiarity with cues that overlap in

graphemic features with an unrecalled studied word suggests that graphemic features

(orthographic and phonological components of the word) are present in the memory trace for

that studied word.

Semantic Features. Though semantic features are much more abstract in nature than

the features discussed thus far, there is an abundance of evidence pointing toward the likely

existence of semantic features in the human knowledge-base (Chang, Mitchell, & Just, 2011;

Griffiths, Steyvers, & Tenenbaum, 2007; Landauer & Dumais, 1997; McCrae, Cree, Seidenberg,

& McNorgan, 2005; Plaut, 1995; Seidenberg, 2007; Smith, Shoben, & Rips, 1974). Using the

feature overlap method with semantic feature norms, Cleary (2004) showed that test cues (e.g.,

bashful) that shared semantic overlap with unrecalled studied words (e.g., shy) led to greater

perceived familiarity than test cues that did not share such semantic resemblance. Cleary et al.

(2016) followed up on this finding using the semantic feature overlap norms of McCrae et al.

(2005) and found that increasing the level of semantic feature overlap between a test cue (e.g.,

cedar) and its corresponding unrecalled target words from the study phase (e.g., birch, oak,

pine, willow) increased the level of perceived familiarity with that cue. These findings suggest

that semantic features are yet another kind of feature retained in memory traces for recently

studied items.

Letter Positions. Other research studies suggest that letter location information (or

relative letter position information) is another type of feature retained in memory traces for

studied words. This research has come from the approach of feature isolation, rather than

feature overlap, at test. Specifically, when letters in their relative positions are isolated at test

(e.g., R_ I _ _R_P) from their potentially unrecalled studied words (e.g., RAINDROP), participants

can discriminate between isolated letter fragments that came from unrecalled studied words

and those that came from unidentified unstudied words (Cleary & Greene, 2000, 2001;

Peyniorcioğlu, 1990). This pattern suggests that relative letter position information is a type of

isolable feature retained in the memory trace for a recently studied word.

Phonemes. Yet other research points toward phonemes as another feature present in

memory traces for studied words. Cleary (2004) found evidence for this using the feature

overlap method. When a cue rhymed with but did not look like an unrecalled studied word

(e.g., the cue *raft* for the unrecalled studied word *laughed*), participants could detect increased

familiarity with that cue relative to when it did not rhyme with a studied word.

Cleary et al. (2007) additionally found evidence for the presence of phonemes in

memory traces for recently studied words using the feature isolation method. When phonemes

were digitally spliced from spoken word recordings, those isolated phoneme fragments at test

were perceived to be more familiar when they came from recently heard or recently seen

unrecalled studied words than when they came from unstudied words. Taken together, these

results suggest that phonemes are a type of isolable feature that is retained in memory traces

for recently studied words.

Rhythm and Pitch. Kostic and Cleary (2009) found evidence that rhythm and pitch are

two types of isolable features that are stored in memory traces for recently heard musical

pieces. Using the feature isolation method, they found that isolating a piano song's rhythm (by

having the exact rhythm tapped out in a single note on a wood block instrument) led to greater

perceived familiarity when the unidentifiable rhythm came from an unrecalled piano song clip

that was recently heard at study than when it came from an unidentified unstudied song clip. In

short, people could discriminate between rhythms that came from recently heard songs and

rhythms that had not, based solely on their feeling or sense of familiarity about the rhythm

itself. The same was found for isolated pitch sequences. Isolating a piano song's pitch sequence

(by retaining the note order but attaching those notes to an arbitrary unstudied rhythm) led to

greater perceived familiarity when the unidentifiable pitch sequence came from an unrecalled

piano song clip that was heard at study than it when it came from an unidentified unstudied

song clip. In short, people could discriminate pitch sequences that came from recently heard

songs and pitch sequences that had not, based solely on their feeling or sense about the pitch

sequence itself.

The Feature Assumption in the Computation of Familiarity

The consistent finding that isolable features are involved in familiarity-detection during

retrieval failure fits well with global matching models of familiarity (see Clark & Gronlund, 1996,

for a review). Global matching models are a class of computational models that specify how the

familiarity signal may operate in human memory. According to these models, the familiarity

signal is a feeling or sense that can vary in intensity and, through relatively high signal intensity,

potentially alert a person to the fact that the current situation that is being experienced has

been experienced in some way previously.

In global matching models, features play a central role in the computation of the

familiarity signal. Every item that was encoded from the encoding phase is represented as a set

of features. Whether the familiarity signal for a given test item (test probe) is higher or lower in

intensity depends on the degree of match between the features present in the test probe and

the features that were stored in memory from the encoding phase. A higher degree of match

results in a higher intensity familiarity signal and vice versa. The term "global' in the global

matching models stems from the fact that the familiarity signal's intensity is not just a function

of the degree of feature match between a given encoded item and the test probe, but the

degree of feature-match *combined* across all items in memory in relation to the test probe. In

short, the level of feature match between the probe and all items in memory determines the

familiarity intensity that results from that probe.

A particularly straightforward example of a global matching model is Hintzman's (1988)

MINERVA 2 model depicted in Figure 1. In this model, each encoded item from the study phase

is laid down as a vector of features. The vector of features constitutes the memory trace, and

the features themselves are represented as a series of +1s, -1s, and 0s (0s are non-encoded or

missed features). The test probe is also a vector of features (see Figure 1). Thus, the assumption

is that the test item itself is decomposed into features by the mind. These test probe features

are matched, on a feature-by-feature basis, with features that exist within the memory traces

that have been stored in memory.
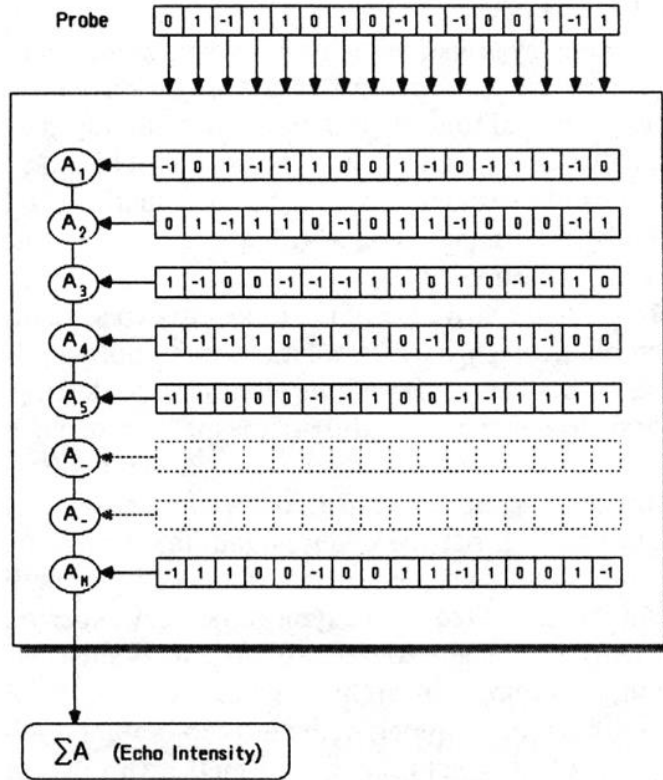
Figure 1. The feature-matching process in the MINERVA 2 model, from Hintzman (1988).

Mathematically, the feature matching process that determines the intensity of the familiarity signal (labeled echo intensity in this model) is carried out as follows. Each feature in the test probe is matched with the feature in that corresponding location in a given memory trace by multiplying the two (e.g., the first feature in the test probe vector is multiplied with the first feature in the memory trace vector). Each product of the multiplication indicates whether there was a feature match or not, as a positive product is an indication of a feature match and a negative product is an indication of a mismatch. For that particular memory trace, each product computed for each feature of that trace is then summed across the memory trace to provide what is called an Activation Value for that memory trace (represented as each A value in Figure 1). More specifically, the value of the sum of the products across the memory trace is cubed, to preserve the sign, and that cubed sum constitutes the Activation Value for that memory trace. The Activation Value serves as an index of the degree of feature match between that particular memory trace and the test probe.

The way that this model is global in its feature matching process is that the Activation Values for each memory trace are then summed to provide the numerical value of the familiarity signal (the echo intensity) that occurs in response to that test item. This means that if multiple memory traces have a high degree of feature overlap with the test probe (hence high Activation Values), the resulting familiarity signal will be higher than if only one does.

Although it is relatively straightforward how the MINERVA 2 model can describe, at a theoretical level, what may be occurring to allow the aforementioned cases of familiarity-detection during retrieval failure (the familiarity signal is higher in intensity among test cues that have high feature overlap with a memory trace than among test cues that have lower feature overlap with any memory traces), that pattern was not what this type of model was originally devised to describe. Global matching models in general were originally devised as theoretical accounts of recognition memory, or old-new discrimination between exact repeats of studied items and new items. These models largely fell out of favor when it became clear that many of their predictions were problematic and did not hold up in studies of old-new recognition memory (e.g., see Clark & Gronlund, 1996, for a review) or of other variants of recognition memory tasks like associative recognition tasks (Gronlund & Ratcliff, 1989). Their downfall was also possibly partly due to the fact that they are single process accounts of recognition memory when it is now widely accepted that old-new recognition very likely involves more than a single process (e.g., Diana, Reder, Arndt, & Park, 2006; Mandler, 2008; Mickes, Wais, & Wixted, 2009; Onyper, Zhang, & Howard, 2010; Wais, Mickes, & Wixted, 2008; Yonelinas, 2002).

However, the fact that the feature-matching method of computing a familiarity signal's

intensity does not adequately describe old-new recognition in standard old-new recognition

memory paradigms does not mean that the process itself does not exist in human cognition. In

fact, it may be that a task designed specifically to probe familiarity-detection in isolation may be

well-suited to testing global matching models' ability to describe the computation of the

familiarity signal that presumably enables familiarity-detection during retrieval failure in the

first place.

Familiarity-detection as a Metacognitive Phenomenon Well-described by Feature Matching

Familiarity-detection as a Metacognitive Phenomenon. Familiarity-detection—the

ability to sense something as familiar despite no ability to identify the source of that familiarity

or the previous experience responsible for it—is a common experience. As Mandler (1991) once

eloquently argued,

> "It is an experience all of us have had at some time or another: We meet somebody at a
> party, know them to be familiar but do not know who they are; we recognize a melody,
> but fail to remember its name or when or where we have heard it before; we read a line
> of a poem, know it, but do not know where we have read it before, much less the title
> or author of the poem" (Mandler, 1991, p. 207).

Thus, familiarity as a subjective experience of human memory appears to exist, even if perhaps

not necessarily well-captured in standard old-new recognition memory paradigms (for reviews

of debates surrounding attempts to separate familiarity from recollection in standard old-new

recognition memory paradigms, see Diana, Reder, Arndt, & Park, 2006; Hintzman, 2011; Wais,

Mickes, & Wixted, 2008; Yonelinas, 2002). Most people can relate to having the experience of

familiarity now and then, in much the same way that most people can relate to having other

metacognitive subjective experiences of memory, like tip-of-the-tongue states (Brown, 2012;

Schwartz, 2002), feelings of déjà vu (Brown, 2004; Cleary et al., 2012; Cleary & Claxton, 2018),

or feelings-of-knowing (Koriat, 1993; Metcalfe, Schwartz, & Joaquim, 1993; Schwartz &

Metcalfe, 1992). Moreover, stimulus familiarity is thought to play a role in other aspects of

metacognition, such as metacognitive control (Malmberg, 2008; Metcalfe, 1993; Reder, 1987).

It may be that, as with these other subjective memory states (tip-of-the-tongue and déjà

vu experiences), the key to studying the sensation of familiarity is to use a task that can induce

the experience in the lab. Indeed, Hintzman (2011) argues that memory researchers often

become too focused on debating theoretical explanations for particular laboratory tasks (using

old-new recognition tasks as one particular example), and lose sight of the real-world memory

phenomena that cognitive scientists should be seeking to understand. Familiarity feelings are

one such cognitive phenomenon.

To understand the subjective metacognitive feeling of familiarity, it may be necessary to

focus on the sense of familiarity during instances of recall failure by using a task designed to

elicit or focus on recall failure. Similar to tip-of-the-tongue experiences, experiences of

familiarity in the world seem to be examples of the subjective sense of memory during retrieval

failure. In the case of tip-of-the-tongue states, there is a subjective sense that a word is in

memory even though the word fails to be retrieved. In the case of familiarity, there is a

subjective sense that the current situation was experienced before, without being able to recall

or identify that prior experience responsible for the sense of familiarity. Just as tasks aimed at

studying tip-of-the-tongue states involve focusing on instances of retrieval failure, attempting

to study the subjective sensation of familiarity, too, might best be arrived at by focusing on

instances of retrieval failure. Focusing on instances of retrieval failure may reveal that the

feature matching process specified in global matching models like MINERVA 2 are a useful

account of how familiarity-detection occurs. That is, the feature-matching process specified in

MINERVA 2 may be a good account of how the familiarity signal arises from a cue to allow a

sense of familiarity with the cue during a failure to consciously recall the previous experience

responsible for that sense of familiarity.

   A Role of Global Feature Matching in Familiarity-detection. Indeed, focusing on

familiarity-detection during instances of retrieval failure has revealed that, besides the general

assumption of segmentation of the cue and the memory traces into isolable features, another

one of the general assumptions of the feature matching process specified in MINERVA 2 holds

up in describing the process likely to underlie such familiarity-detection. This assumption is that

of *global* matching—the idea that it is the combined level of feature-match across all items in

memory that produces the familiarity signal for the test cue, not just any one item. According to

the MINERVA 2 model depicted in Figure 1, if multiple memory traces have a high degree of

feature match to the test probe, the resulting echo intensity (the intensity of the familiarity

signal) should be higher than if only one memory trace has a high degree of feature match to

the test probe. In turn, if only one memory trace has a high degree of feature match to the test

probe, that trace should still lead to a higher echo intensity value than if no memory traces

have a high degree of feature match to the test probe.

   Ryals and Cleary (2012) empirically investigated whether this pattern would hold up

among instances of familiarity-detection during retrieval failure. Using the aforementioned

feature overlap method, they examined the level of reported familiarity intensity for nonword

test cues (e.g., POTCHBORK) that overlapped in graphemic features with four different but

unrecalled studied words that had been scattered across the encoding phase (e.g., PITCHFORK,

PATCHWORK, POCKETBOOK, PULLCORK, all going unrecalled at test) versus cues that

overlapped in graphemic features with only one studied word that happened to go unrecalled

(e.g., only PITCHFORK was studied and it went unrecalled at test), versus cues that did not

overlap in graphemic features with any studied words. They found that participants' reported

familiarity intensity with test cues increased with an increasing number of unrecalled studied

items that shared graphemic features with the test cue. In short, during instances of retrieval

failure, the test cue was perceived to be more familiar if it overlapped in features with more

than one studied item (all of which were unrecalled) than if it overlapped with only one (that

went unrecalled). This pattern suggested that the computation of the familiarity signal followed

the principle of global matching. Interestingly, the pattern was less pronounced during

instances of retrieval success (when any of the targets were successfully recalled in response to

the cue), suggesting that other processes besides feature-matching are likely at work during

instances of retrieval success, and feature-matching may be the primary driver of decisions

made during instances of retrieval failure.

Cleary et al. (2016) later found the same pattern with semantic features, rather than

graphemic features. They found that when a test cue (e.g., cedar) overlapped in semantic

features with four unrecalled studied words (birch, oak, pine, willow), the reported cue

familiarity level during retrieval failure was significantly higher than when the cue overlapped

with only two studied words that happened to go unrecalled (e.g., only *birch* and *oak* were

studied and both happened to be unrecalled). In turn, when the cue overlapped semantically

with two unrecalled studied words, reported familiarity was significantly higher than when it

overlapped with none. This follows from what the MINERVA 2 model depicted in Figure 1 would predict.

The Present Study

General Goals. Familiarity-detection during retrieval failure appears to be well-described by some of the basic assumptions of global matching models like MINERVA 2. For one, familiarity-detection during retrieval failure appears to involve isolable features, and in fact, the basic methodology for studying familiarity-detection during retrieval failure has been used to systematically identify the types of features that are present in memory traces for recently experienced events (see Cleary, 2014, for a review). For another, familiarity-detection during retrieval failure appears to be at least roughly well-described by a global feature matching process, whereby the degree of feature overlap between each and every memory trace and the cue is factored into the intensity of the familiarity signal (Cleary et al., 2016; Ryals & Cleary, 2012).

A limitation to the studies that have investigated the global matching assumption of the MINERVA 2 model in its ability to account for familiarity-detection during retrieval failure is that these studies did not allow for a precise quantification of a single feature-type. For example, Ryals and Cleary (2012) relied on general graphemic overlap between the cue (e.g., POTCHBORK) and items in memory (e.g., PITCHFORK, PATCHWORK, POCKETBOOK, PULLCORK). This method allows for a very general investigation of the global matching and feature overlap assumptions of MINVERA 2, but does not allow for isolation of a specific feature-type or an assessment of how different feature-types might combine across memory traces in a

quantifiable way. The same held true for the study of semantic feature overlap by Cleary et al.

(2016).

If different feature-types from a whole stimulus could be isolated—an aim of

Experiment 1 of the present study—would those different isolated feature-types combine

across different study episodes in a predictable fashion to increase the later overall level of

perceived familiarity with the whole stimulus item in which they were embedded at test? Note

from Figure 1 how the MINERVA 2 model would predict that they should. Different feature sets

encoded in isolation during separate study episodes should combine to contribute to the

familiarity signal in predictable ways. Specifically, if one set of features was laid down in one

trace and another different set of features was laid down in another trace, the contribution of

those two memory traces to the intensity of the familiarity signal should be additive. This would

occur because the Activation Values are added together to contribute additively to the

familiarity signal for that test item (as shown in Figure 1). This was the aim of Experiment 2 of

the present study.

Moreover, if one particular isolated feature-type was repeated multiple times across

study episodes, there should also be a corresponding quantifiable increase in the level of

perceived familiarity for the whole stimulus in which that feature-type is embedded at test. This

is because the Activation Value for every encoding instance of that isolated feature-type at

study would be expected to be very similar (not necessarily identical given the random variation

in which features are attended to and successfully encoded as opposed to missing and laid

down as zeros in the memory trace) and to combine additively toward increasing the overall

intensity of the familiarity signal from the test cue. This means that the average familiarity

signal should increase in a predictable, additive fashion when increasing repetitions of a specific

isolated feature-type at encoding. This was the aim of Experiment 3 of the present study.

Two candidate feature-types that can be isolated and separated from their whole

stimulus in a relatively straightforward way are rhythm and pitch from a musical piece. Rhythm

and pitch can both be digitally extracted from their original musical piece using music

composition software (Kostic & Cleary, 2009), making them great candidate feature-types for

exploring global matching predictions about the role of isolable features in the computation of

the familiarity signal. For this reason, the focus in the present study is on musical stimuli, with

the features of rhythm and pitch as the two isolable feature-types under investigation.

A clear assumption present in the MINERVA 2 model shown in Figure 1 is that the test

probe itself is decomposed into its constituent features. If memory traces are stored as sets of

features that are then combined to compute a familiarity signal for a test item, it should not

matter if the whole test item itself was never presented at encoding. If some of its component

features were studied, that should increase the test item's familiarity. Those features should be

detectable from the whole. Although this assumption has been generally supported in the

aforementioned feature overlap studies (Cleary, 2004; Cleary et al., 2012; Cleary et al., 2016;

Ryals & Cleary, 2012), it has not yet been directly shown with musical stimuli.

With music, the reverse has been shown: Isolated features in the test probe (either

rhythm or pitch sequences in isolation) felt more familiar if they had been embedded within a

whole song clip heard at the time of study, suggesting that those features were retained in the

memory traces for song clips heard at study (Kostic & Cleary, 2009). However, the reverse

should also hold true: Features presented in isolation at encoding should lead to later

familiarity with wholes at test that contain those features. Thus, Experiment 1 of the present

study sought to test the hypothesis that studying isolated sets of musical features at encoding

will increase the perceived familiarity of a whole test song clip in which those features happen

to be embedded. It should not matter if the entire test song clip itself was not heard at study; if

some of the song's features were encoded in isolation (when the song itself could not be

consciously identified from those features), that should be enough to increase the perceived

familiarity with the whole song clip at test. In Experiment 1, therefore, the encoding phase

consisted of sets of isolated rhythms and isolated pitch sequences from piano song clips. For

each piano song clip heard at test, either its isolated rhythm or its isolated pitch sequence

(without its rhythm) were potentially heard at study. Furthermore, each test song clip's identity

was unable to be accessed at the time of that feature exposure (i.e., the song was not identified

at encoding from the features, making it unlikely that an encoding episode could be consciously

retrieved in response to the test item, leaving the decision to be based solely on the sense of

familiarity with the test item). This logic has been used in studies of word fragment recognition

without identification in a reverse form of the feature isolation procedure (e.g., Cleary &

Greene, 2000, Experiments 3a and 3b).

Experiment 1 not only sought to demonstrate that embedding familiarized features

(either rhythm or pitch) into a piano song clip increases the sense of familiarity with that piano

song clip, but also to investigate whether the different features of rhythm and pitch are

weighted comparably in the computation of the familiarity signal. Though not directly

investigated in their study, Kostic and Cleary's (2009) results seem to point toward the

possibility that pitch information carries more weight in the computation of the familiarity

signal than rhythm information. Thus, it is possible that pitch is weighted more heavily in the computation of musical familiarity than rhythm. If so, then reported familiarity experiences should be more likely among test song clips for which the isolated pitch was studied than among test song clips for which the isolated rhythm was studied. Experiment 1 allowed for a direct investigation of this hypothesis.

Experiments 2 and 3 then sought to test a specific idea embedded within the MINERVA 2 model's mechanisms regarding how features from different memory traces combine across traces in the computation of the familiarity signal--what we refer to from hereafter as the *additivity principle*. The additivity principle is the idea that the Activation Values shown in Figure 1 combine additively in the computation of the familiarity signal. Below, we demonstrate the additivity principle using simulations of our general experimental paradigm. We include simulations for scenarios in which pitch carries more weight in the computation of the familiarity signal for a whole song clip than rhythm, as well as scenarios in which pitch and rhythm contribute comparably to the familiarity signal computation for the whole song clip.

MINERVA 2 Simulations. Given our goal of testing whether music familiarity-detection in our paradigm adheres to the additivity principle present in the MINERVA 2 model regarding how different feature-types should combine across separate memory traces in the computation of the familiarity signal, we carried out a set of simulations to first demonstrate the additivity principle within the MINERVA 2 model. It was important to also consider whether the simulations would robustly show additivity if pitch is indeed a larger contributing feature-type than rhythm to the familiarity signal computation. To simulate a scenario in which pitch information is more heavily weighted within the memory trace than rhythm information in the

MINERVA 2 model, when creating the memory traces, a larger proportion of the test probe's

features can be dedicated to being "pitch-features" than "rhythm-features." For example, for a

given "whole song segment" test probe [1, -1, 0, 1, -1, -1, 0, 1, 1, 1], a pitch memory trace might

consist of the first 70% of the features, with the remaining features being zeroed out [1, -1, 0, 1,

-1, -1, 0, 0, 0, 0]. Conversely, the rhythm memory trace of the above test probe might consist of

the last 30% of the features: [0, 0, 0, 0, 0, 0, 0, 1, 1, 1]. The ratio of rhythm-to-pitch features can

be varied across the simulations to examine how varying the ratio affects the outcome in terms

of whether the additivity principle still holds as this is varied. In our simulations, we examined

the rhythm-to-pitch ratios of 1:1, 2:3, and 3:7.

Our critical simulations, taking into account possible varying ratios of pitch versus

rhythm feature-types, were to demonstrate MINERVA 2's additivity principle, which our

Experiments 2 and 3 were designed to examine. Specifically, Experiments 2 and 3 sought to

examine how feature-types contained within a single test cue but present within different

memory traces combine across traces to contribute to the resulting familiarity signal. Our initial

proof of principle MINERVA 2 simulations were set up to mimic the general design of these two

experiments in order to demonstrate the mathematical existence of the additivity principle

among mean echo intensities for these types of designs. Therefore, before describing each set

of simulations, we will begin by giving an overview of the experimental design that it was meant

to simulate.

In Experiment 2, four conditions were compared. In one condition, a song clip's isolated

rhythm was separately encoded from that song clip's isolated pitch sequence, then both

feature-types were embedded within the whole song clip at test. We will refer to this as the

Rhythm+Pitch condition. Note that in the *Rhythm+Pitch* condition, the two feature-types

(rhythm and pitch) were still studied in isolation, just in different study episodes—one where

rhythm was studied in isolation (without the pitch), and one where pitch was studied in

isolation (without the rhythm). In a second condition (*Rhythm-Only*), only the isolated rhythm

was studied. In a third condition (*Pitch-Only*), only the isolated pitch sequence was studied.

Finally, a fourth control condition involved having no studied features for a given test song clip.

The question was how the isolated features of rhythm and pitch would combine across study

episodes to contribute to the familiarity signal. That is, exactly how much would the familiarity

of the cue increase in the *Rhythm+Pitch* condition compared to the *Rhythm-Only* or the *Pitch-*

Only conditions?

The additivity principle within the MINERVA 2 model allows for specific predictions

regarding how the separately encoded traces in the *Rhythm+Pitch* condition should combine to

boost the familiarity signal. As shown in Figure 1, having two separate memory traces, each

with a different feature-type—one memory trace for the song's rhythm and another separate

memory trace for the song's pitch sequence—should lead to greater perceived familiarity with

the test song clip in which *both* feature-types are embedded than when the test song clip only

contains one studied feature (either rhythm *or* pitch instead of rhythm *and* pitch). Because of

the aforementioned additivity principle, the level of test song clip familiarity elicited by having

separately encoded both rhythm *and* pitch features of the cue should equal that of the

combined level of familiarity for studying *only* rhythm and studying *only* pitch. That is, the level

of familiarity elicited by a cue for which only one feature-type was encoded (e.g., rhythm),

when added to the level of familiarity elicited by a cue for which only the other feature-type

was encoded (e.g., pitch), should approximately equal the level of familiarity elicited by a test

song clip that contains both encoded feature-types (both rhythm and pitch separately encoded

across different study episodes). This is the principle of additivity. Logically, this means that,

relative to the control condition, the level of familiarity increase for *Rhythm+Pitch* cues should

equal the level of familiarity increase for *Rhythm-Only* cues plus the level of familiarity increase

for *Pitch-Only* cues. Thus, the model predicts the following:

$$(Rhythm+Pitch \text{ cues}) = (Rhythm\text{-}Only \text{ cues}) + (Pitch\text{-}Only \text{ cues})$$

To demonstrate that this principle exists within MINERVA 2, we carried out a set of

simulations that are depicted in Figure 2. This set of simulations was intended to demonstrate

that, when added together, the mean echo intensities for the *Rhythm-Only* and the *Pitch-Only*

conditions approximately equal that of the *Rhythm+Pitch* condition. We performed our

simulations using Python 3.7, including NumPy (Oliphant, 2006) and Jupyter (Kluyver et al.,

2016), and the code for these simulations can be found at (https://github.com/dwhite54/MINERVA

2).

So that our initial proof of principle simulations would closely mimic our actual

experimental design while also adhering to precedent regarding how test probes corresponding

to "new" or unstudied items are typically simulated (e.g., Hintzman, 1988), we included an

unstudied condition in which the test probes corresponded to zero items in the memory store

in these initial proof of principle simulations. However, despite existing precedent for

simulating probes for unstudied items as corresponding to zero items in the memory store, as

will be shown later, even unstudied items in our paradigm are not completely novel—they are

features of known songs—and thus likely still overlap somewhat with musical memory

representations in the knowledge-base at some baseline level; we will address this when simulating our actual data from our experiments in the individual experiment sections. For the purposes of these initial proof of principle simulations, given that our primary theoretical focus is on the comparison between the echo intensity values across the three conditions of *Rhythm-Only*, *Pitch-Only,* and *Rhythm+Pitch* (specifically, the question of whether mean echo intensities adhere to the formula [*Rhythm+Pitch* cues] = [*Rhythm-Only* cues] + [*Pitch-Only* cues]), we only depict the echo intensity density graphs for these three conditions in Figure 2.

To closely follow our actual experimental design, for each cue condition (*Rhythm-Only, Pitch-Only, Rhythm+Pitch,* and *unstudied features*), 30 test probes were created; thus, there were a total of 120 test probes per run. We completed 120 runs (each run being a hypothetical participant) using the 120 test probes each for a total of 14,400 data points, with 3,600 data points in each of the four conditions. Thus, 3,600 data points contributed to each echo intensity density graph in Figure 2. Each test probe was randomly generated then used to create a corresponding memory trace (or set of memory traces) according to otherwise pre-specified criteria for each condition, described below. Thus, no two test probes or memory traces were identical across simulation runs.

Setting the number of features per memory trace and test probe is a somewhat arbitrary process, as it is unclear how many individual features are contained within a given stimulus in actuality. This uncertainty also applies to the question of how many features are contained within a song's rhythm or a song's pitch structure in actuality. We set each test probe and each memory trace to contain 1,000 features each.
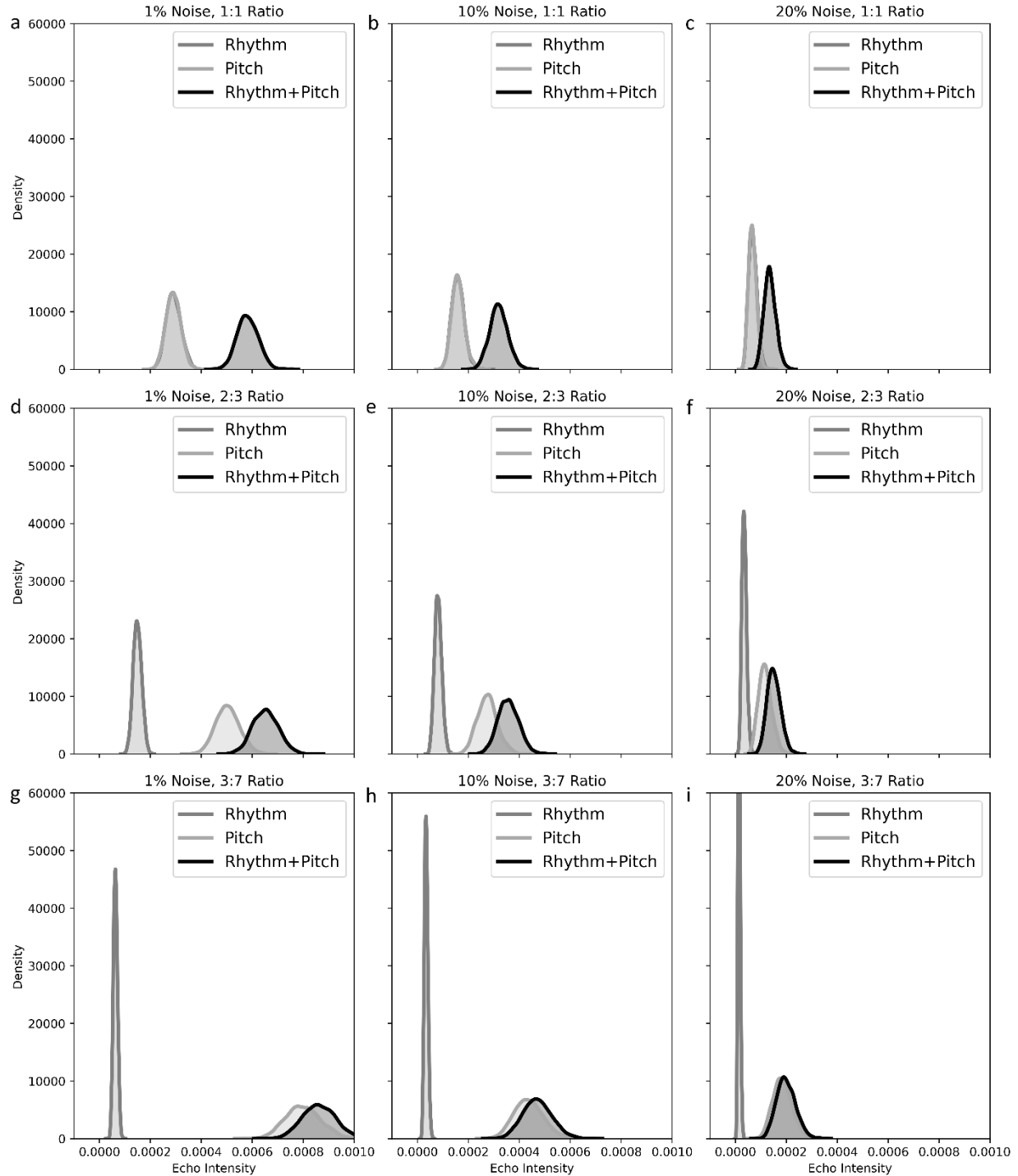
Memory traces were created for the *Rhythm-Only, Pitch-Only,* and *Rhythm+Pitc*h

conditions using the randomly generated test probes, such that a proportion of the test probe's

features were present in the memory trace as if they had been studied in isolation as in our

experimental paradigm. Across different simulations, we varied the rhythm-to-pitch ratios

within the test probes and their corresponding memory traces. A 1:1 rhythm-to-pitch ratio

would mean that 50% of the features within each test probe were considered "pitch" features

and 50% were considered "rhythm" features. A test probe corresponding to the *Rhythm-Only*

condition would have the "rhythm" half of the features contained within it map onto a memory

trace containing only those particular features whereas a test probe corresponding to the *Pitch-*

Only condition would have the "pitch" half of the features contained within it map onto a

memory trace containing only those particular features. A test probe corresponding to the

Rhythm+Pitch condition would have both sets of features contained within it each mapping

onto a different memory trace (the "rhythm" half would map onto to the memory trace

containing only the "rhythm" features, and the "pitch" half would map onto the memory trace

containing the "pitch" features).

To simulate a scenario in which pitch information is more heavily weighted within the

memory trace than rhythm information, when creating the memory traces, a larger proportion

of the test probe's features were dedicated to being "pitch-features" than "rhythm-features" in

some of the simulations. For example, in a given test probe [1, -1, 0, 1, -1, -1, 0, 1, 1, 1], a pitch

memory trace might consist of the first 70% of the features, with the remaining features being

zeroed out [1, -1, 0, 1, -1, -1, 0, 0, 0, 0]. Conversely, the rhythm memory trace of the above test

probe might consist of the last 30% of the features: [0, 0, 0, 0, 0, 0, 0, 1, 1, 1]. We varied the

ratio of rhythm-to-pitch features across the simulations to examine how this affects the

outcomes; specifically, we examined the following ratios: 1:1, 2:3, and 3:7.

Our method is otherwise similar to prior MINERVA 2 simulations (e.g., Hintzman, 1988,

p. 532). To mimic the imperfection of human memory, noise was incorporated into the memory

traces, such that a proportion of the features' signs were flipped (e.g., a feature of +1 might be

changed to -1 to create noise; Hintzman, 1986). The amount of noise was also varied to

examine if the additivity principle holds up across different noise levels (1%, 10%, and 20%).

Because the test probes for each of the 3,600 data points per condition were each

randomly generated and because it is unclear how one hypothetical "participant" would differ

in any meaningful way from the next (the way that actual human participants would), rather

than attempting to separate data points by "participants" for the purposes of statistical

analysis, all data points were entered into independent samples *t*-tests to assess for differences

in mean echo intensities. As the priors are all equal in the model (unlike with human

participants), treating all data points (individual trials) the same should not violate statistical

assumptions the way that they would if individual human trials were all entered into the same

analysis (as between-subject variability does not occur in the model). This also allowed for a

high-powered investigation of proposed null effects between simulation conditions.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Figure 2. Density graphs for the echo intensities obtained for the simulated feature-type conditions of *Rhythm-Only, Pitch-Only*, or *Rhythm+Pitch*, their relative rhythm-to-pitch ratios (1:1, 2:3, or 3:7), and differing noise-levels (1%, 10%, or 20%). Note that when the rhythm-to-pitch ratio is 1:1 the *Rhythm-Only* and *Pitch-Only* distributions are perfectly overlapping.

Figure 2 represents the density graphs for each of the simulations for the conditions of theoretical interest (*Rhythm-Only, Pitch-Only,* and *Rhythm+Pitch*). The mean echo intensities are depicted in Table 1. To assess whether a given mean echo intensity differed significantly from that of another condition, we performed independent-samples *t*-tests. The outcomes of the statistical analyses are reported in Table 2, including the Bayes Factors analyses. Because Levene's Test for Equality of Variances revealed unequal variances (which are also depicted in Figure 2), *t*-tests for unequal variances were used (hence the varying degrees of freedom across analyses).

Across all of the noise-levels that we examined, when the rhythm-to-pitch ratio was 1:1 (see panel a in Figure 2), as expected, the average echo intensities for the *Rhythm-Only* probes or *Pitch-Only* probes were not significantly different (see Tables 1 and 2 for the descriptive and inferential statistics, respectively). However, across all noise levels, when the ratio of rhythm-to-pitch features was adjusted to either 2:3 or 3:7, a significant difference was found between the echo intensities, such that the average echo intensities for the *Pitch-Only* probes were then larger than those of the *Rhythm-Only* probes. This mimics what we would expect to find in our data from human participants if pitch information is weighted more heavily than rhythm in the computation of the familiarity signal from the whole song clip cue (i.e., a higher instance of reported familiarity experiences for cues corresponding to an isolated pitch sequence than to cues corresponding to an isolated rhythm). If pitch carries approximately the same weight as rhythm in the computation of the familiarity signal, then we should find that familiarity elicited by either feature-type does not significantly differ (as when the rhythm-to-pitch ratio is 1:1). Whether encoded pitch information carries more weight in the computation of the familiarity

signal for a whole song clip than encoded rhythm information is an empirical question that will

be borne out by our data. The above simulation results demonstrate how it is possible to factor

differing rhythm-to-pitch ratios into our simulations. Below, we discuss the simulations that

demonstrate the primary theoretical question of interest; we include simulations that

demonstrate the various outcomes when using the above different rhythm-to-pitch ratios and

the above varying noise levels.

Table 1.
Mean Echo Intensities Across Experiment 2 Simulation Conditions

Noise	Ratio	Rhythm-Only		Pitch-Only		Rhythm+Pitch	
		M	*SD*	*M*	*SD*	*M*	*SD*
1%	1:1	2.92e-04	2.93e-05	2.91e-04	2.89e-05	5.82e-04	4.13e-05
1%	2:3	1.50e-04	1.66e-05	5.03e-04	4.65e-05	6.52e-04	5.02e-05
1%	3:7	6.32e-05	8.30e-06	7.99e-04	6.78e-05	8.64e-04	6.68e-05
10%	1:1	1.59e-04	2.44e-05	1.60e-04	2.46e-05	3.19e-04	3.55e-05
10%	2:3	8.17e-05	1.41e-05	2.74e-04	3.90e-05	3.57e-04	4.10e-05
10%	3:7	3.46e-05	6.98e-06	4.37e-04	5.75e-05	4.71e-04	5.70e-05
20%	1:1	6.84e-05	1.62e-05	6.77e-05	1.58e-05	1.36e-04	2.31e-05
20%	2:3	3.50e-05	9.26e-06	1.17e-04	2.57e-05	1.51e-04	2.65e-05
20%	3:7	1.48e-05	4.47e-06	1.85e-04	3.73e-05	2.00e-04	3.73e-05

We now turn to our primary theoretical question of interest, which is that of how

familiarizing both rhythm *and* pitch (*Rhythm+Pitch)* affects the echo intensities compared to

familiarizing either alone (*Rhythm-Only* or *Pitch-Only*). As shown in Tables 1 and 2, at all

rhythm-to-pitch ratios and all noise levels, the mean echo intensities for *Rhythm-Only* and

Pitch-Only probes were significantly lower than those for the *Rhythm+Pitch* probes. These

results demonstrate the generic prediction of MINERVA 2 that when a test cue contains a blend

of two different feature-types that were each separately familiarized across two different

memory traces (*Rhythm+Pitch* cues*)*, the overall familiarity intensity produced by the test cue is

higher than when the test cue contains only one familiarized feature-type among its blend of features (*Rhythm-Only* cues or *Pitch-Only* cues). More importantly, the results also demonstrate MINERVA 2's additivity principle: The separately encoded memory traces for the two feature-types (rhythm vs. pitch) combine additively to boost the cue familiarity of the full song clip in the *Rhythm+Pitch* condition. To reveal this, the echo intensities were examined to determine if they adhered to the following equation:

$$(Rhythm+Pitch \text{ cues}) = (Rhythm\text{-}Only \text{ cues}) + (Pitch\text{-}Only \text{ cues})$$

Toward this end, we compared the obtained echo intensities for the *Rhythm+Pitch* condition to the values obtained by adding together the *Rhythm-Only* and the *Pitch-Only* echo intensities for each of the different rhythm-to-pitch ratios and noise levels under examination. There was no significant difference found in any of these cases, and the evidence favored the null, as shown in Tables 1 and 2 (see the [*Rhythm-Only*]+[*Pitch-Only*] vs. *Rhythm+Pitch* comparisons in Table 2). These simulations demonstrate that the separately encoded memory traces for each feature-type (rhythm vs. pitch) combine additively in the computation of the global familiarity signal produced by a test probe containing both feature-types in MINERVA 2. This additivity occurred regardless of the rhythm-to-pitch ratio or noise level used in the simulations. Thus, the additivity principle held up robustly. Based on these simulations, we examined whether the experimental data from human participants given isolated rhythms and/or pitch sequences across encoding episodes will also adhere to the following formula when both feature-types are contained within the test cue:

$$(Rhythm+Pitch \text{ cues}) = (Rhythm\text{-}Only \text{ cues}) + (Pitch\text{-}Only \text{ cues})$$

Table 2.
Statistical Comparisons of Echo Intensities Across Experiment 2 Conditions

		Rhythm-Only vs. Pitch-Only
Noise	Ratio	Statistical Result
1%	1:1	$t(7198) = -1.12$, $SE = 6.86e\text{-}7$, $p = .26$, $BF_{01} = 20.06$
1%	2:3	$t(4498) = 430.13$, $SE = 8.23e\text{-}7$, $p = < .001$, $BF_{01} < .001$
1%	3:7	$t(3707) = 646$, $SE = 1e\text{-}6$, $p = < .001$, $BF_{01} < .001$
10%	1:1	$t(7198) = -.76$, $SE = 5.77e\text{-}7$, $p = .45$, $BF_{01} = 28.29$
10%	2:3	$t(4528) = 278.74$, $SE = 6.92e\text{-}7$, $p < .001$, $BF_{01} < .001$
10%	3:7	$t(3705) = 417.54$, $SE = 9.65e\text{-}7$, $p < .001$, $BF_{01} < .001$
20%	1:1	$t(7194) = 1.93$, $SE = 3.78e\text{-}7$, $p = .05$, $BF_{01} = 5.86$
20%	2:3	$t(4519) = 180.65$, $SE = 4.55e\text{-}7$, $p < .001$, $BF_{01} < .001$
20%	3:7	$t(3715) = 272.11$, $SE = 6.26e\text{-}7$, $p < .001$, $BF_{01} < .001$
		Rhythm-Only vs. Rhythm+Pitch
Noise	Ratio	Statistical Result
1%	1:1	$t(6486) = 344.47$, $SE = 8.44e\text{-}7$, $p < .001$, $BF_{01} < .001$
1%	2:3	$t(4375) = 571.25$, $SE = 8.80e\text{-}7$, $p < .001$, $BF_{01} < .001$
1%	3:7	$t(3710) = -713.93$, $SE = 1e\text{-}6$, $p < .001$, $BF_{01} < .001$
10%	1:1	$t(6379) = 223.19$, $SE = 7.17e\text{-}7$, $p < .001$, $BF_{01} < .001$
10%	2:3	$t(4441) = 381.13$, $SE = 7.23e\text{-}7$, $p < .001$, $BF_{01} < .001$
10%	3:7	$t(3707) = 455.90$, $SE = 9.58e\text{-}7$, $p < .001$, $BF_{01} < .001$
20%	1:1	$t(6460) = 143.45$, $SE = 4.70e\text{-}7$, $p < .001$, $BF_{01} < .001$
20%	2:3	$t(4463) = 248.68$, $SE = 4.68e\text{-}7$, $p < .001$, $BF_{01} < .001$
20%	3:7	$t(3715) = 295.03$, $SE = 6.27e\text{-}7$, $p < .001$, $BF_{01} < .001$
		Pitch-Only vs. Rhythm+Pitch
Noise	Ratio	Statistical Result
1%	1:1	$t(6439) = 346.94$, $SE = 8.40e\text{-}7$, $p < .001$, $BF_{01} < .001$
1%	2:3	$t(7157) = 130.70$, $SE = 1e\text{-}6$, $p < .001$, $BF_{01} < .001$
1%	3:7	$t(7198) = -40.49$, $SE = 2e\text{-}6$, $p < .001$, $BF_{01} < .001$
10%	1:1	$t(6415) = 221.83$, $SE = 7.20e\text{-}7$, $p < .001$, $BF_{01} < .001$
10%	2:3	$t(7180) = 87.91$, $SE = 9.44e\text{-}7$, $p < .001$, $BF_{01} < .001$
10%	3:7	$t(7198) = 24.99$, $SE = 1e\text{-}6$, $p < .001$, $BF_{01} < .001$
20%	1:1	$t(6373) = 147.18$, $SE = 4.66e\text{-}7$, $p < .001$, $BF_{01} < .001$
20%	2:3	$t(7190) = 55.66$, $SE = 6.15e\text{-}7$, $p < .001$, $BF_{01} < .001$
20%	3:7	$t(7198) = 16.58$, $SE = 8.79e\text{-}7$, $p < .001$, $BF_{01} < .001$
		[Rhythm-Only] + [Pitch-Only] vs. Rhythm+Pitch
Noise	Ratio	Statistical Result
1%	1:1	$t(7198) = .06$, $SE = 9.73e\text{-}7$, $p = .95$, $BF_{01} = 37.57$
1%	2:3	$t(7197) = .44$, $SE = 1e\text{-}6$, $p = .66$, $BF_{01} = 34.17$
1%	3:7	$t(7194) = .50$, $SE = 2e\text{-}6$, $p = .50$, $BF_{01} = 29.93$
10%	1:1	$t(7194) = -.21$, $SE = 8.26e\text{-}7$, $p = .83$, $BF_{01} = 36.79$
10%	2:3	$t(7196) = 1.30$, $SE = 9.75e\text{-}7$, $p = .20$, $BF_{01} = 16.28$
10%	3:7	$t(7198) = .64$, $SE = 1e\text{-}6$, $p = .52$, $BF_{01} = 30.65$
20%	1:1	$t(7198) = .49$, $SE = 5.38e\text{-}7$, $p = .63$, $BF_{01} = 33.43$
20%	2:3	$t(7198) = 1.17$, $SE = 6.34e\text{-}7$, $p = .24$, $BF_{01} = 19.11$
20%	3:7	$t(7198) = .31$, $SE = 8.83e\text{-}7$, $p = .76$, $BF_{01} = 35.89$

Whether this additivity assumption holds up was examined in Experiment 2, where we also present simulations of our behavioral data patterns.

The additivity principle would not be shown in the human data if, for example, the *Rhythm+Pitch* condition did not lead to any increase in the perceived familiarity of the song in which those features are embedded over and above either of those features in isolation. That is, if combining the two feature-types across separate encoding episodes leads to no further increase in the level of perceived familiarity with the song clip in which they are embedded, the additivity assumption of the MINERVA 2 model would not hold up.

The additivity principle would also not be supported by the data if rhythm and pitch were to combine multiplicatively in the generation of the familiarity signal; for example, if the overall perceived level of familiarity with the test song clip showed an even larger increase in familiarity than would be expected by either feature-type alone simply added together, it would suggest that there is added familiarity value to having more than one familiar feature-type in the cue. Such a pattern might exist if the whole were greater than the sum of its parts when it came to the computation of the familiarity signal—that is, if the bundling of multiple familiarized feature-types adds familiarity value over and above the familiarization of those individual feature-types. Such a multiplicative combining of features in the computation of familiarity might make evolutionary sense, insofar as the presence of multiple familiar feature-types might, in and of itself, suggest an even greater likelihood that this current situation has been experienced in the past. Thus, in actuality, unlike in MINERVA 2, there might be added value to having more than one familiarized feature-type in the cue. Experiment 2 also served as a test of this idea.

In Experiment 3, we used a slightly different approach to demonstrate the additivity

principle in MINERVA 2 regarding how features combine across memory traces to produce the

familiarity signal. Instead of examining how two different isolated feature-types combine across

separate encoding episodes to produce the familiarity signal with a whole song clip at test, the

same specific instance of a feature-type (i.e., the same rhythm) was simply repeated multiple

times throughout the encoding phase. The goal was to examine the level of increase in the

perceived familiarity with the whole song clip at test. As shown in Figure 1, even a single

feature-type like a song's rhythm, if repeated across separate study episodes, should

systematically and predictably boost the level of perceived familiarity with the whole in which

that feature is embedded at test, specifically in an additive fashion. Our next set of simulations

was aimed at demonstrating this manifestation of the additivity principle within MINERVA 2.

In Experiment 3, we examined if the repetition of a single feature-type—rhythm—across

separate study episodes would additively increase the level of perceived familiarity with the

whole song clip in which that rhythm was embedded at test. Toward this end, piano song clips

presented at test had their rhythms familiarized either zero times (the *Rhythm0x* condition),

one time (the *Rhythm1x* condition), or three times (the *Rhythm3x* condition) throughout the

study phase. As shown in Figure 1, relative to the control condition (*Rhythm0x*), the level of

perceived familiarity increase for test songs containing a rhythm that was repeated three times

at study (*Rhythm3x*) should be three times the level of familiarity increase caused by a single

rhythm presentation at study (*Rhythm1x*). Thus, it should adhere to the following formula:

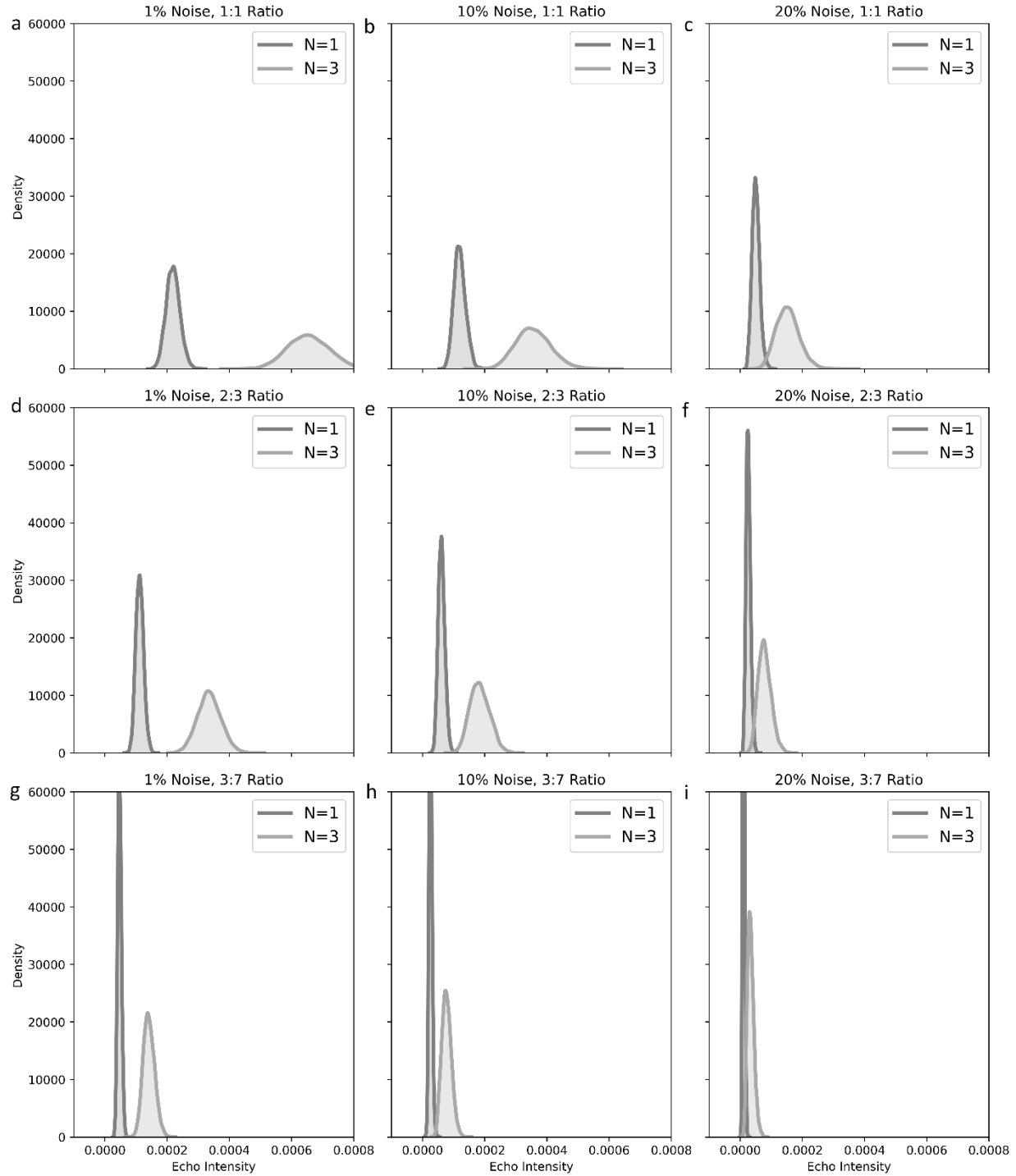$$Rhythm3x \text{ cues} = 3[Rhythm1x \text{ cues}]$$

Figure 3. Density graphs for the echo intensities obtained for the simulated feature-type conditions of *Rhythm1x* and *Rhythm3x*, their relative rhythm-to-pitch ratios (1:1, 2:3, or 3:7), and differing noise-levels (1%, 10%, or 20%).

To demonstrate this manifestation of the additivity principle within the MINERVA 2 model (regarding how the same feature set repeated across memory traces should combine across those separate memory traces to affect the overall echo intensity elicited by a cue containing that subset of features), we carried out a set of simulations that are depicted in Figure 3. This set of simulations was intended to demonstrate, in a proof of principle, that repeating an isolated feature set from a particular song (e.g. the song's rhythm) across three different memory traces would boost the echo intensity elicited by a cue containing those features by a factor of three as described in the above equation.

For our Experiment 3 simulations, 40 test probes were randomly generated for each of the critical conditions under examination (*Rhythm0x, Rhythm1x, Rhythm3x*), each consisting of 1,000 features, for a total of 120 test probes. Like for our previous simulations, memory traces were created from each randomly generated test probe, this time for the *Rhythm1x* and *Rhythm3x* conditions, by using a proportion of the corresponding test probe's features. Although pitch sequences were not the focus of Experiment 3, a full piano song clip as the test cue would hypothetically still contain both pitch and rhythm features; therefore, we examined the same rhythm-to-pitch ratios and noise levels in our Experiment 3 simulations as for Experiment 2's simulations described above (and as depicted in Tables 3 and 4).

Table 3. Mean Echo Intensities Across Experiment 3 Simulation Conditions

Noise	Ratio	Rhythm1x		Rhythm3x	
		M	*SD*	*M*	*SD*
1%	1:1	2.19e-4	2.21e-5	6.55e-4	6.62e-5
1%	2:3	1.12e-4	1.26e-5	3.36e-4	3.77e-5
1%	3:7	4.72e-5	6.14e-6	1.42e-4	1.84e-5
10%	1:1	1.20e-4	1.87e-5	3.56e-4	5.53e-5
10%	2:3	6.13e-5	1.07e-5	1.84e-4	3.15e-5
10%	3:7	2.59e-5	5.15e-6	7.76e-5	1.54e-5
20%	1:1	5.08e-5	1.22e-5	1.53e-4	3.65e-5
20%	2:3	2.63e-5	7.12e-6	7.90e-5	2.08e-5
20%	3:7	1.12e-5	3.43e-6	3.39e-5	1.04e-5

Our simulations were meant to mimic the isolation of rhythm features from the full song

clip that would serve as the test cue, and the presentation of those isolated rhythms at study,

as was done in our Experiment 3. As in our previous proof of principle simulation, our focus was

again on the two conditions of theoretical interest--the comparison of the *Rhythm1x* condition

to the *Rhythm3x* condition to test for adherence to the above formula. Simulations for our full

behavioral data patterns will be presented in Experiment 3's section. For each memory trace

assigned to the *Rhythm3x* condition, three duplicates (total) of that memory trace were

created. Once the test probes and their memory traces were created, overall echo intensities

for each condition were calculated and the process was repeated again for a total of 120

simulations to mimic the actual experimental design.

The density graphs created from the simulations are shown in Figure 3. The mean echo

intensities are depicted in Table 3. To assess whether a given mean echo intensity differed

significantly from that of another condition, we performed independent-samples *t*-tests. The

outcomes of the statistical analyses are reported in Table 4, including the Bayes Factors

analyses. Because Levene's Test for Equality of Variances revealed unequal variances (which are

also depicted in Figure 3), *t*-tests for unequal variances were used (hence the varying degrees of

freedom across analyses).

We first compared the echo intensities obtained for the *Rhythm1x* condition with the

echo intensities obtained for the *Rhythm3x* condition. As can be seen in Table 4, across all

rhythm-to-pitch ratios and noise levels examined, the echo intensities were significantly higher

for the *Rhythm3x* condition than the *Rhythm1x* condition, as expected. When a memory trace is

repeated three times, the resulting familiarity signal is greater in intensity than if the memory

trace occurs only once. Turning now to our primary theoretical question, we compared the

mean echo intensity for the *Rhythm3x* condition with the mean value obtained by multiplying

the echo intensity in the *Rhythm1x* condition by a factor of three to determine if these values

adhere to the following formula:

$$Rhythm3x \text{ cues} = 3[Rhythm1x \text{ cues}]$$

Indeed, as shown in Table 4, the evidence across all of our rhythm-to-pitch ratios and

noise levels favored the null according to Bayes Factors analyses. Possibly due to the extremely

high power resulting from the many iterations of our simulation runs, some *p*-values

approached or reached significance; however, even in these cases, the Bayes Factor analyses

indicated that the evidence favored the null.

Table 4. Statistical Comparisons of Echo Intensities Across Experiment 3 Conditions

Noise	Ratio	Statistical Result
Rhythm 1x vs. Rhythm3x		
1%	1:1	$t(5861) = -432.79$, $SE = 1.00e{-}6$, $p < .001$, $BF_{01} < .001$
1%	2:3	$t(5853) = -389.78$, $SE = 5.74e{-}7$, $p < .001$, $BF_{01} < .001$
1%	3:7	$t(5858) = -337.59$, $SE = 2.8e{-}7$, $p < .001$, $BF_{01} < .001$
10%	1:1	$t(5886) = -280.89$, $SE = 8.4e{-}7$, $p < .001$, $BF_{01} < .001$
10%	2:3	$t(5888) = -256.57$, $SE = 4.80e{-}7$, $p < .001$, $BF_{01} < .001$
10%	3:7	$t(5853) = -220.15$, $SE = 2.35e{-}7$, $p < .001$, $BF_{01} < .001$
20%	1:1	$t(5849) = 184.08$, $SE = 5.56e{-}7$, $p < .001$, $BF_{01} < .001$
20%	2:3	$t(5908) = 166.14$, $SE = 3.18e{-}7$, $p < .001$, $BF_{01} < .001$
20%	3:7	$t(5827) = 143.30$, $SE = 1.59e{-}7$, $p < .001$, $BF_{01} < .001$
Rhythm 3x vs. 3[Rhythm1x]		
1%	1:1	$t(9598) = 1.74$, $SE = 1.00e{-}6$, $p = .08$, $BF_{01} = 9.52$
1%	2:3	$t(9598) = .59$, $SE = 7.7e{-}7$, $p = .56$, $BF_{01} = 36.53$
1%	3:7	$t(9598) = -.09$, $SE = 3.8e{-}7$, $p = .93$, $BF_{01} = 43.28$
10%	1:1	$t(9598) = 2.44$, $SE = 1.00e{-}6$ $p = .02$, $BF_{01} = 2.23$
10%	2:3	$t(9598) = .78$, $SE = 6.48e{-}7$, $p = .44$, $BF_{01} = 32.11$
10%	3:7	$t(9598) = .01$, $SE = 3.15e{-}7$, $p = .99$, $BF_{01} = 43.45$
20%	1:1	$t(9598) = -.96$, $SE = 7.45e{-}7$, $p = .34$, $BF_{01} = 27.29$
20%	2:3	$t(9598) = .51$, $SE = 4.31e{-}7$, $p = .61$, $BF_{01} = 38.24$
20%	3:7	$t(9598) = 1.63$, $SE = 2.12e{-}7$, $p = .10$, $BF_{01} = 11.50$

==Collectively, these proof of principle simulations demonstrate that a particular isolated feature set (e.g., a song's isolated rhythm) repeated across three separate memory traces will boost the familiarity signal with a cue in which those features are embedded (e.g., a full song clip in which that rhythm occurs) by roughly a factor of three relative to when that particular isolated feature set occurs in only one memory trace. The purpose of Experiment 3 was to examine whether the behavioral data conform to this pattern, and to then simulate that specific behavioral data pattern using MINERVA 2.== Specifically, do the data conform to the formula:

$$Rhythm3x \text{ cues} = 3[Rhythm1x \text{ cues}]$$

The above-mentioned additivity assumption would be violated if, for example, repeating the same feature-type (in this case rhythm) throughout the encoding phase leads to no increase in the perceived familiarity of the song in which it is embedded. That is, if once a feature-type is familiarized, repeating that feature leads to no further increase in the level of perceived familiarity with the song clip in which it is embedded, the predictions of the MINERVA 2 model would not be supported.

The assumption would also be violated if the repeated memory traces for the separate rhythm encodings were to combine multiplicatively instead of additively in the generation of the familiarity signal for the test song clip. For example, if the overall perceived level of familiarity with the test song clip showed an even larger increase when the feature-type was repeated three times than would be expected by merely adding together each iteration of familiarity increase expected by a single encoding instance of that feature, it would suggest that there is added familiarity value to repeatedly familiarizing a single feature-type within the cue. Such a pattern might exist if the whole were greater than the sum of its parts when it came to the computation of the familiarity signal—that is, if the repeating of a single feature-type across episodes adds familiarity value over and above the familiarization level of each instance. For example, if there is a mechanism discovered in Experiment 2 for added value of bundling multiple feature-types over and above that predicted by the mere adding together of their expected individual familiarity levels (i.e., a "whole is greater than the sum of its parts" component to the computation of familiarity), that mechanism might be expected to manifest across repetitions of the same feature-type too. In this way, Experiment 3 provides a different means of examining the same question posed in Experiment 2.

Experiment 1

Experiment 1 sought to first establish that embedding familiarized song features (that had been earlier encoded in isolation) within later song clips at test would increase the perceived familiarity of those test song clips. Toward this end, participants studied either isolated song rhythms (a song clip's rhythm tapped out in a single note on a wood block instrument) or isolated pitch sequences (a song clip's note order extracted from its original rhythm by adhering those notes to an arbitrary rhythm). To ensure that the features were encoded in isolation (without access to other features for that song), the focus was on instances in which the song from which the isolated features came could not be identified at study. This procedure is similar to the reversed recognition without identification procedure first used by Cleary and Greene (2000, Experiments 3a and 3b).

As the present study searched for a novel empirical finding never shown before (i.e., that embedding familiarized musical features that were encoded in isolation within their full song clips at a later time boosts the sense of familiarity with the entire song clip in which they are embedded), we chose to err on the side of a relatively large sample-size for all of the experiments reported here, using the findings of Kostic and Cleary (2009) as our starting point.

In Kostic and Cleary's study, isolated pitch sequences from earlier-studied but unidentified songs led to higher recognition ratings than isolated pitch sequences from unstudied unidentified songs (Cohen's *d* was 1.30), as did isolated rhythms from earlier studied but unidentified songs when compared to isolated rhythms from unstudied unidentified songs (Cohen's *d* was .35). As the present experiments all involved the use of isolated rhythms, which were the smaller effect size in Kostic and Cleary's study, we used that effect size as our starting

point. To achieve a power of 90% with an alpha of .05, an effect size of .35 would require a

sample-size of 88 participants. Thus, if we were merely replicating the isolated rhythm

experiment of Kostic and Cleary, we would need a sample size of 88. However, as the present

paradigm is novel and the potential effect size *a priori* was unclear (and also, we were searching

within-experiment in a mixed design for potential differences in magnitude between isolated

pitch sequences and isolated rhythms in their contributions to the familiarity signal for a whole

song clip), we conservatively aimed for a larger sample size of 120 per experiment.

We targeted 120 total participants per experiment for the present study through our

weekly sign-up process over the course of many weeks (attempting to estimate sign-ups and no

shows week by week), and captured additional participants through that estimation process,

who we included (arriving at 129 for Experiment 1, 123 for Experiment 2, and 128 for

Experiment 3).

Method

Participants. Participants were 129 undergraduates from Colorado State University who

completed the study in exchange for credit toward a course. Three participants were excluded

due to not being native English speakers. An additional three participants were lost due to not

completing the experiment. This left 123 participants.

Materials. The stimuli consisted of the 120 segments of well-known piano songs (e.g., of

children's melodies and pop songs) from Kostic and Cleary (2009) and their separately isolated

rhythms and pitch sequences. To isolate rhythm from tonal information, the rhythm from each

song was played on a wood block instrument in the single note of Middle C (so that there was

no tonal variation). To isolate pitch information, the notes for each song were played in their

original order but according a different, unstudied rhythm. The same arbitrary rhythm used by

Kostic and Cleary was assigned to all song segments with the exception of a few for which there

was some degree of overlap between that rhythm and the one present in the actual song. For

those, a new, non-studied arbitrary rhythm was created by musical composer and professor Dr.

Michael Thaut. This was to ensure no rhythm overlap between pitch feature segments and the

songs from which they came.

Procedure. Each participant completed the experiment, which was created using E-

Prime software (n.d.), in a separate computer booth fitted with speakers which allowed

participants to adjust the volume to a comfortable level. In four counterbalanced versions of

the experiment, every song was either studied as an isolated pitch sequence or as an isolated

rhythm. These conditions were then compared at test to a control condition in which no

features from the test song clip had appeared at study.

The 120 piano song segments were divided into ten study-test blocks. Each study list

consisted of three isolated pitch sequences and three isolated rhythms in a randomized order

for each participant. Participants were initially instructed that they would be completing a

music memory test, and that they should do their best to remember each musical piece

presented at study. After a given isolated feature segment was presented through the speakers

at study, a dialog box appeared on the screen and asked if the participant could identify the

song from which the isolated feature-type came. The participant was to type the name of the

song (or any information about it, such as the artist or lyrics) into the dialog box or simply press

Enter if nothing was known. For example, if what had been played through the speakers was

the tapped-out rhythm from the song "London Bridge", the participant would try to type

"London Bridge," some of song's words, or the artist (if known). A song was considered

identified from its isolated features if the participant named the song, wrote some of the lyrics,

or indicated who had composed the song. Based on these criteria, songs were later separated

into songs identified-at-study and songs unidentified-at-study for analysis.

After the study list of six randomly-ordered isolated feature sets (half of them isolated

rhythms and half of them isolated pitch sequences) was presented, participants then

completed the randomly-ordered test phase for that block, which contained the unaltered

piano version (e.g., the refrain from London Bridge simply played on the treble clef of a piano

instrument, with both is pitch sequence and its rhythm intact) of the six studied song-segments

intermixed with six unaltered piano song segments for which no features (neither rhythm nor

pitch) had been studied.

For each unaltered piano song segment played at test, participants were asked to

indicate if it seemed familiar to them (Y=Yes, N=No), with the added instruction that familiar

would mean that the song segment likely corresponded to one of the studied song feature

segments. Following this, a second dialog box appeared prompting participants to attempt to

identify the song. Although we recorded identification responses for the full song clips at test as

well as for the isolated song features at study, our primary interest was in familiarity judgments

for test song clips that had not been identified from their isolated features at study (songs

unidentified-at-study). Thus, that is the focus of the present study.

Results and Discussion

The data were analyzed using traditional null hypothesis significance testing (NHST) and

Bayes Factors analyses. Using Bayes Factors analyses alongside NHST allowed us to assess

whether the evidence favored the null hypothesis instead of merely failing to reject it (e.g.,

Kruschke, 2013). In the following results sections, we report Bayes Factors (*BFs*), which quantify

the strength of the evidence for the alternative (BF_{10}) and the null (BF_{01}) hypotheses. Using the

recommendations provided by Wagenmakers (2007), we considered a BF_{01} to provide either

anecdotal evidence (BF_{01} = 1-3), substantial evidence (BF_{01} = 3-10), strong evidence (BF_{01} = 10-

30), very strong evidence (BF_{01} = 30-100), or extreme evidence (BF_{01} > 100) for the null

hypothesis. A value of exactly 1.0 constitutes no evidence either way (for either the null or the

alternative), whereas the values in support of the alternative hypothesis are the inverse values

(e.g., a BF_{01} of .33-1.0 would provide anecdotal evidence for the alternative hypothesis; a BF_{01}

of .10-.33 would provide substantial evidence for the alternative hypothesis; a BF_{01} of .03-.10

would provide strong evidence for the alternative hypothesis; a BF_{01} of .01-.03 would provide

very strong evidence for the alternative hypothesis; a BF_{01} of less than .01 would provide

extreme evidence for the alternative hypothesis). All NHST were conducted using IBM SPSS

while all Bayes Factors analyses were calculated with JASP using the JZS prior as it requires the

fewest prior assumptions about the range of the true effect size (Rouder, Speckman, Sun,

Morey, & Iverson, 2009).

Song Identification Rates

Identification of songs from isolated features during the study phase was at floor. The

proportion of songs identified from isolated tonal sequences (*M* = .03, *SD* = .04) was

significantly higher than the proportion of songs identified from isolated rhythms (*M* = .01, *SD* =

.02), *t*(122) = 5.99, *SE* = .003, *p* < .001, BF_{01} = 2.03 x 10^{-6}, 95% CrI [.34, .72]. However, later, there

was no significant difference between identification rates at test for full song clips that had had

their tonal sequences isolated at study (M = .26, SD = .15) and those that had had their rhythms isolated at study (M = .25, SD = .15), $t(122)$ = 1.37, SE = .01, p = .17, BF_{01} = 3.92, 95% CrI [-.05, .30], nor did familiarizing either type of feature at study make the songs more identifiable at test. Identification rates of test song clips that had not had any of their features familiarized at study (M = .24, SD = .13) were not lower than identification rates for test clips that had had their isolated pitch sequences ($t(122)$ = -1.69, SE = .01, p = .09, BF_{01} = 2.36, 95% CrI [-.33, .03]) or rhythms ($t(122)$ = -.21, SE = .01, p = .84, BF_{01} = 9.70, 95% CrI [-.20, .15]) studied. In short, familiarizing a song clip's features of either type (rhythm or pitch) did not make the song itself more identifiable later on.

Perceived Test Song Clip Familiarity

Turning now to the comparisons of interest for the present study, we next report the familiarity judgments given to test song clips whose isolated features were unidentified at study. If, as implied in Figure 1, each test probe is processed as a set of separable features, then whole unaltered piano song clips at test whose rhythms or pitch sequences had been studied should feel more familiar to participants than whole unaltered piano song clips containing no studied features. Indeed, a repeated-measures analysis of variance (ANOVA) performed on the probability of responding yes ("yes, this test song clip is familiar") revealed a significant effect of the feature familiarization manipulation, $F(2, 244)$ = 12.67, MSE = .01, p < .001, η_p^2 = .09, BF_{01} =
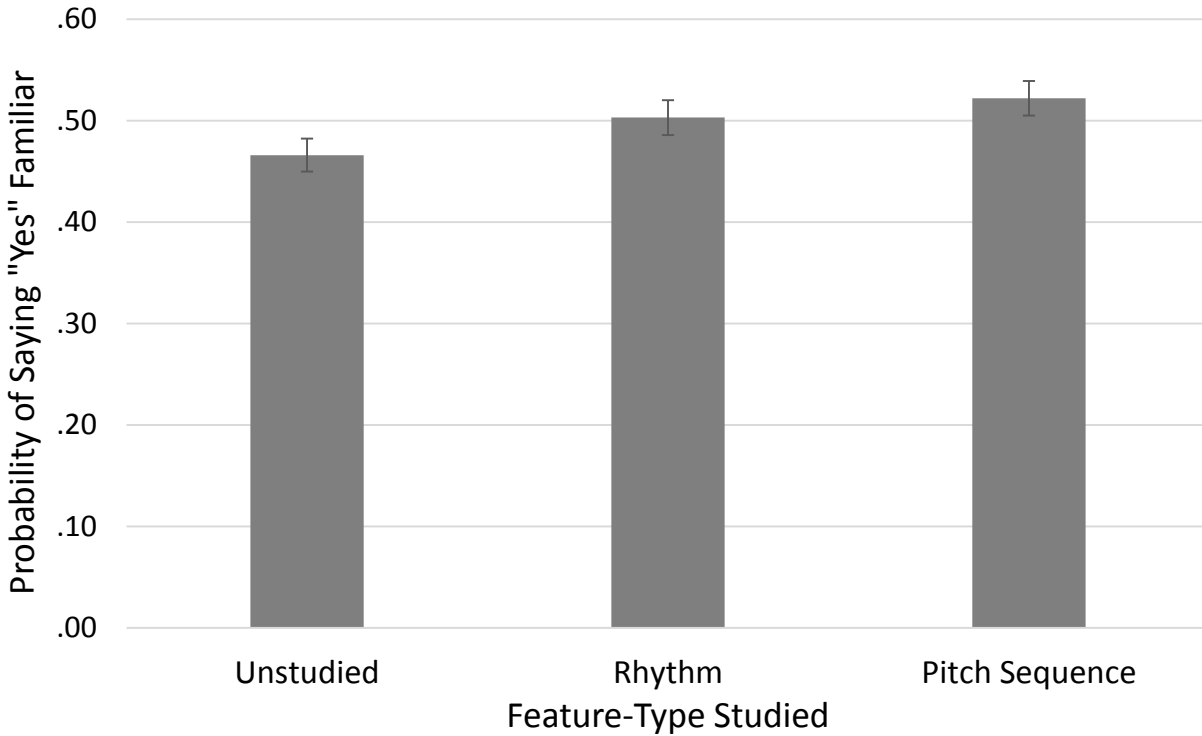
Figure 4. The probability of indicating that a piano song segment played at test felt familiar as a function of Feature Familiarization Condition in Experiment 1. Song clips that had an isolated feature, either the rhythm or tonal sequence, presented at encoding were more likely to prompt feelings of familiarity compared to song clips that did not have any isolated feature presented at encoding.

4.17×10^{-4} (see Figure 4).[1] The proportion of unaltered piano song clips judged as familiar at

test was higher when the song clip's pitch sequences had appeared at study in isolation ($M =$

.52, $SD = .19$) than when none of its features had appeared at study ($M = .47$, $SD = .18$), $t(122) =$

5.16, $SE = .01$, $p < .001$, $d = .31$, $BF_{01} = 7.6 \times 10^{-5}$, 95% CrI [.27, .64].[2] Additionally, the proportion

judged as familiar at test was also higher when the song clip's rhythm had appeared at study in

isolation ($M = .50$, $SD = .19$) than when none of its features had been studied, $t(122) = 3.29$, $SE =$

.01, $p = .001$, $d = .20$, $BF_{01} = 0.06$, 95% CrI [.12, .47].[3]

There was no significant difference in perceived familiarity of test song clips as a function of which particular feature-type (rhythm or pitch) had been studied, $t(122) = 1.61$, $SE = .01$, $p = .11$, $d = .10$, $BF_{01} = 2.85$, 95% CrI [-.31, .03].[4] This is somewhat contrary to what may have been expected, as the isolated pitch sequences could be argued to contain the additional feature of piano timbre whereas the isolated rhythms only contained the timbre of a wooden block. Research done by Poulin-Charronnat et al. (2004) demonstrated that, upon changing the timbre of a song, participants' abilities to identify the song are reduced, suggesting that timbre is an important feature contained in musical representations. In the current study, altering the timbre did not appear to have an effect from study to test, even though it might have been hypothesized that a change in timbre from a wooden block to a piano might affect the perceived familiarity of the song. It may be that any such effects of timbre are very small. Alternatively, it may be that features studied in isolation do not differ substantially from one another in how they are weighted in the later computation of the familiarity signal for a cue presented at test. These findings suggest that a 1:1 rhythm-to-pitch ratio is appropriate for our MINERVA 2 simulations; however, our simulations demonstrated that the predictions for our subsequent experiments did not hinge on the assumption of a 1:1 rhythm-to-pitch ratio.

Taken together, the results of Experiment 1 suggest that familiarizing an isolated subset of a song's component features at encoding increases the likelihood of a subjective sense of familiarity with the whole song at test. This is the first such demonstration with musical stimuli. The fact that this feature familiarization process can be used to increase perceived familiarity with a whole song clip at test means that the general methodology can be used to test more specific predictions of the MINERVA 2 model shown in Figure 1. Along these lines, the present

findings suggest that the different feature-types, namely, familiarized rhythm and familiarized

pitch sequence features, are not weighted differentially in the computation of the familiarity

signal for the whole song clip in which both of those feature-types are embedded. Both feature-

types seem to contribute roughly equally to the computation of the familiarity signal, insofar as

the increase in the likelihood of a reported sense of familiarity was comparable for the

familiarization of either type of feature.

Experiment 2

Having demonstrated in Experiment 1 that embedding a familiarized feature into a

whole song segment at test increases participants' perceived familiarity with the song segment,

Experiment 2 sought to use this basic feature familiarization methodology to test the

hypothesis that presenting two separate feature-types in isolation across different study

episodes (i.e., presenting isolated rhythm in one study episode and the isolated pitch sequence

from the same song but in a different study episode) will lead to an additive increase in the

perceived level of familiarity with the whole song segment at test. As noted earlier from our

proof of principle simulations, the MINERVA 2 model has within it an additivity assumption such

that the level of familiarity increase experienced for a piano test song clip for which the song's

rhythm *and* its pitch sequence were separately familiarized should equal the sum of the

familiarity increase occurring when *only* rhythm was familiarized plus the familiarity increase

occurring when *only* pitch was familiarized. That is, the features of rhythm and pitch should

combine across episodes in the global matching process such that:

(*Rhythm+Pitch* cues) = (*Rhythm-Only* cues) + (*Pitch-Only* cues)

This is because, in the MINERVA 2 model, the intensity of the familiarity signal is determined by the sum of the Activation Values. Thus, in our paradigm, the Activation Value for the memory trace corresponding to a studied rhythm should be added to the Activation Value for the memory trace corresponding to a studied pitch sequence when both are present in the test song clip.

The above pattern will not be obtained if, during the computation of the familiarity signal, features combine multiplicatively rather than additively. For example, the whole might be greater than the sum of its parts when it comes to the computation of familiarity. That is, the presence of more than one familiarized feature-type may lead to a greater overall familiarity intensity level than would be expected by the mere sum of the familiarity increase caused by those familiarized feature-types alone. It is conceivable that the presence of more than one familiar feature-type within the test cue would itself carry diagnostic information about the likelihood that the test cue was previously experienced in some manner. For this reason, it might make evolutionary sense for different feature-types to combine multiplicatively rather than additively in the computation of the familiarity signal across study episodes. This would manifest as a "whole being greater than the sum of its parts" situation with regard to how familiar the test item will seem. In this case, one would expect to find:

(*Rhythm+Pitch* cues) > [(*Rhythm-Only* cues) + (*Pitch-Only* cues)]

In short, if the MINERVA 2 model depicted in Figure 1 provides a good description of how the familiarity signal is computed, then not only should the two different feature-types combine across separate study episodes to increase the level of perceived familiarity with the test cue, but they should combine additively rather than multiplicatively, as demonstrated in

our initial proof of principle simulations—our initial proof of principle simulations show that the

MINERVA 2 model predicts an additive boost to cue familiarity from different feature-types

occurring across different study episodes. The purpose of Experiment 2 was to examine

whether the data conform to those predictions.

Method

      Participants. One hundred and twenty-three Colorado State University undergraduates

participated in exchange for course credit. Three participants were lost due to not finishing the

experiment, leaving 120 participants.

      Materials. Due to time constraints imposed by adding new study episodes relative to in

Experiment 1, the stimuli for Experiment 2 were reduced to 84 of the piano songs and their

corresponding isolated features created by Kostic and Cleary (2009).

      Procedure. The procedure was similar to that used in Experiment 1. Each participant

completed one of four counterbalanced versions of the experiment on a computer in a room

fitted with speakers. The 84 piano song segments were divided into seven study-test blocks,

with each study list consisting of 12 isolated rhythm and pitch sequence presentations, such

that each of the study conditions (Rhythm-only, Pitch-only, and Rhythm+Pitch) corresponded to

three unique songs. Specifically, three isolated rhythms corresponded to the Rhythm-only

condition and therefore to a unique song each. Three isolated pitch sequences corresponded to

the Pitch-only condition and therefore to a unique song each. In contrast, three isolated

rhythms and three isolated pitch sequences corresponded to the Rhythm+Pitch condition, and

therefore, although presented separately, each rhythm shared the same song as a pitch

sequence from this condition (and vice versa). At test, each test list contained 12 test song clips

(three Rhythm-only, three Pitch-only, three Rhythm+Pitch, and three with Unstudied features).

Thus, there were 21 test items per condition overall in the experiment.

As in Experiment 1, at study, participants were instructed to attempt to identify the song from each isolated feature-type if possible (when it was presented at study). Based on the criteria presented in Experiment 1, later test trials were separated into songs identified-at-study and songs unidentified-at-study for analysis. At test, participants were asked to indicate for each test song clip if it seemed familiar (Y=Yes, N=No). They were additionally asked to provide a familiarity rating on a scale of 0 (*definitely not familiar*) to 10 (*strongly familiar*). Our goal in adding the familiarity rating was to determine if participants would further differentiate test song familiarity levels beyond the yes-no familiarity response. After providing a familiarity rating, participants were asked to attempt to identify the song.

Results and Discussion

Song Identification Rates

As in Experiment 1, identification rates for isolated features were at floor (see Table 5 for proportions of correctly identified songs at study and at test). Participants' overall identification rates of songs from their isolated features at study was 1.4% (SD = .02). A repeated-measures ANOVA on Feature-Type indicated that there was a significant difference in identification rates, $F(2, 238) = 18.69$, $MSE = .001$, $p < .001$, $BF_{01} = 1.61$ x 10^{-6}. Participants were significantly more likely to identify songs from isolated pitch features than from isolated rhythms, $t(119) = 4.47$, $SE = .003$, $p < .001$, $BF_{01} = 1.00$ x 10^{-3}, 95% CrI [.22, .59]. Additionally, songs that had their isolated pitch sequences *and* isolated rhythms separately presented at encoding were significantly more likely to be identified at study compared to songs that were

Table 5
Proportion of Song Clips Correctly Identified at Study and Test

| Feature | At study | | At test | |
Condition	*M*	*SD*	*M*	*SD*
Unstudied	-	-	.22	.14
Rhythm	.004	.016	.22	.15
Pitch	.021	.037	.23	.13
Rhythm+Pitch	.029	.045	.23	.15

solely presented as isolated rhythms, $t(119) = 5.83$, $SE = .004$, $p < .001$, $BF_{01} = 4.28 \times 10^{-6}$, 95% CrI [.33, .71]. This is likely due to priming (see Kostic & Cleary, 2009, for a description of similar priming effects in song identification ability). The difference in identification rates between songs that had solely had their pitch sequences studied and songs that had *both* isolated pitch sequences and rhythms studied only approached significance without reaching it, $t(119) = 1.76$, $SE = .004$, $p = .08$, $BF_{01} = 2.23$, 95% CrI [-.33, .02].

At test, participants' overall song identification rate was also at floor, with participants only identifying an average of 22% ($SD = .12$) of song clips. A repeated-measures ANOVA on Feature-Type did not show a significant difference among identification rates, $F(3, 357) = .58$, $MSE = .01$, $p = .63$, $BF_{01} = 56.07$. None of the individual comparisons among identifications rates were significant.

Perceived Test Song Clip Familiarity

Behavioral Familiarity Ratings Data. The familiarity ratings that were added to Experiment 2's methodology did not lead to any additional differentiation beyond the yes-no familiarity responses. As shown in Figure 5, participants' familiarity ratings were largely

differentiated on the basis of their yes-no responses regarding whether the test song clip

seemed familiar or not. A 2 x 4 Yes-No Response (yes vs. no) x Feature Familiarization Condition

(*Unstudied* vs. *Rhythm-Only* vs. *Pitch-Only* vs. *Rhythm+Pitch*) repeated measures ANOVA

performed on participants' mean familiarity ratings revealed a significant main effect of Yes-No

Response, $F(1, 113) = 70.65$, $MSE = 12.35$, $p < .001$, $\eta_p^2 = .39$, $BF_{01} = 2.10 \times 10^{-59}$. As shown in

Figure 5, participants gave higher familiarity ratings to song clips for which they responded

"yes, it is familiar," than to song clips for which they responded "no, it is not familiar." There

was no interaction ($F<1.0$, $BF_{01} = 88.9$). A significant main effect of Feature Familiarization

Condition emerged, $F(1, 113) = 3.00$, $MSE = 0.93$, $p = .03$, $\eta_p^2 = .03$, $BF_{01} = 105.60$.[5] However,

although the NHST suggests a significant effect, the Bayes Factor provides extreme evidence for

the null.
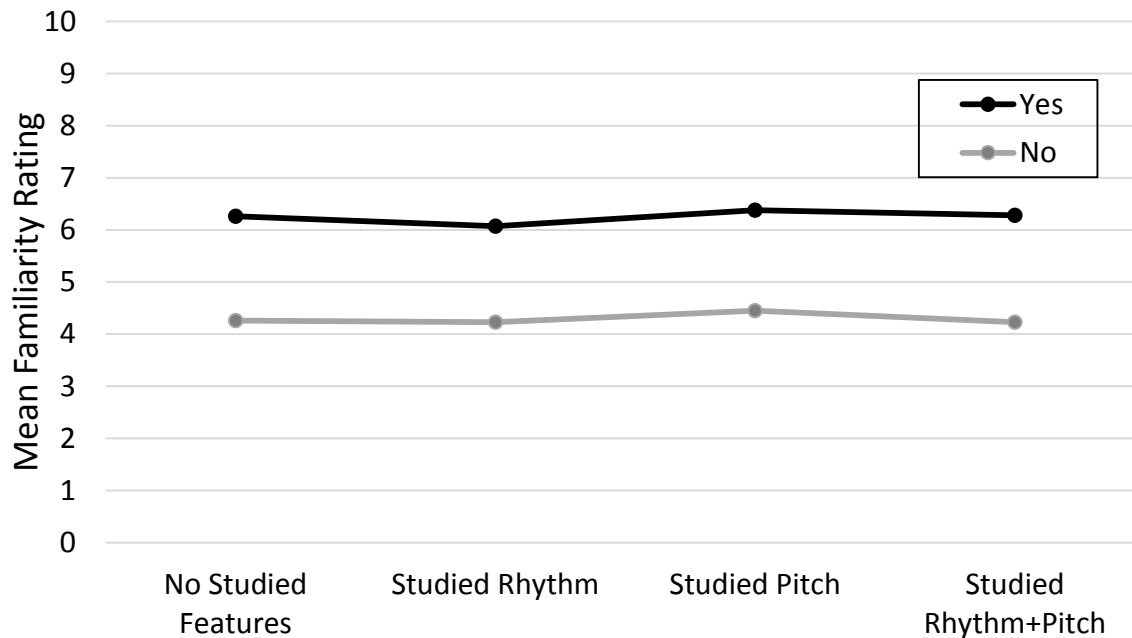
*Figure 5*. Mean familiarity ratings as a function of yes-no responses.

Behavioral Yes-No Data. A repeated-measures ANOVA performed on the probability of

a "Yes" response (Yes, the test song clip is familiar) showed a significant effect of Feature

Familiarization Condition (*Unstudied, Rhythm-Only, Pitch-Only, Rhythm+Pitch*), $F(3, 357) =$

11.92, $MSE = .01$, $p < .001$, $\eta_p^2 = .09$, $BF_{01} = 1.77 \times 10^{-5}$ (see Figure 6).[6] See Table 6 for the

proportions of piano test song clips judged as familiar. As was expected, test song clips that did

not contain any familiarized features were less frequently judged as familiar than test song clips

containing familiarized rhythms ($t(119) = -2.03$, $SE = .01$, $p = .04$, $d = .09$, $BF_{01} = 1.36$, 95% CrI [-

.36, -.005]), test song clips containing familiarized pitch sequences ($t(119) = -2.99$, $SE = .02$, $p =$

.003, $d = .19$, $BF_{01} = 0.15$, 95% CrI [-.44, -.08]), and test song clips containing both familiarized

pitch sequences *and* familiarized rhythms (that were separately familiarized across encoding

episodes), $t(119) = -5.48$, $SE = .01$, $p < .001$, $d = .37$, $BF_{01} = 1.96 \times 10^{-5}$, 95% CrI [-.67, -.31].[7] As in

Experiment 1, there was no difference in the likelihood of a "Yes" response for test songs

containing familiarized rhythms versus test songs containing familiarized pitch sequences,

$t(119) = -1.18$, $SE = .01$, $p = .24$, $BF_{01} = 5.01$, 95% CrI [-.28, .07].[8]

Table 6
Proportion of Songs Judged as Familiar

Feature Condition	M	SD
Unstudied	.46	.22
Rhythm	.48	.21
Pitch	.50	.21
Rhythm+Pitch	.54	.21

The primary theoretical question under investigation in Experiment 2 is whether the

separately isolated features of rhythm and pitch would combine additively across separate

study episodes to increase the overall perceived familiarity with the test song clip. First, test

song clips that contained *both* familiarized pitch sequences and rhythms were indeed more

often found to be familiar than test clips containing only familiarized rhythms (t(119) = 4.05, *SE*

= .01, *p* < .001, *d* = .29, BF_{01} = 5.00 x 10^{-3}, 95% CrI [.18, .55]) or only familiarized pitch sequences,

t(119) = 2.77, *SE* = .01, *p* = .007, *d* = .19, BF_{01} = 0.26, 95% CrI [.06, .43].[9] Thus, the separately

encoded feature-types did combine across study episodes to increase the overall level of

perceived familiarity with the test song over and above that expected by either encoded

feature-type alone.
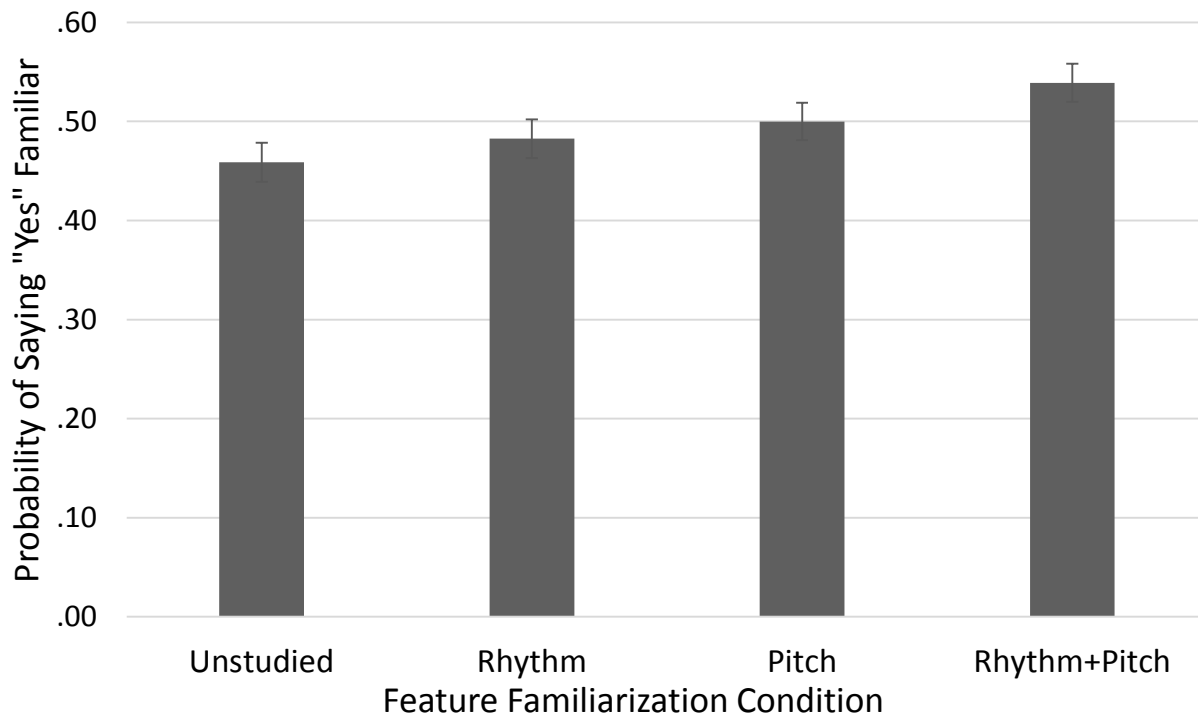
Figure 6. The probability of judging a test song clip as familiar as a function of feature
familiarization condition in Experiment 2. Song clips were more likely to be found familiar if
they had had at least one feature-type (rhythm or pitch) presented in isolation at encoding.
However, song clips were even more likely to be found familiar if both the rhythm *and* the pitch
sequence had each separately been encoded in isolation. This pattern suggests that separately
encoded features combine across memory traces in the computation of the familiarity signal.

A close inspection of the proportions presented in Table 6 suggests that the *amount* of increase in perceived familiarity of the test song was that expected if the separate feature-types had combined additively, as opposed to multiplicatively. That is, the separate features of rhythm and pitch appear to have adhered largely to the formula:

(*Rhythm+Pitch* cues) = (*Rhythm-Only* cues) + (*Pitch-Only* cues)

rather than the formula:

(*Rhythm+Pitch* cues) > [(*Rhythm-Only* cues) + (*Pitch-Only* cues)]

To examine this more closely, for each participant, we computed the amount of increase in the likelihood of reporting "yes, familiar" between test song clips in the control condition (no feature familiarization) and the *Rhythm-Only* familiarization condition (M = .02, SD = .13), the amount of increase between test song clips in the control condition and the Pitch-Only familiarization condition (M = .04, SD = .15), and finally, the amount of increase between the control condition and the Rhythm+Pitch condition (M = .08, SD = .16). For each participant, the sum of the person's rhythm increase (*Rhythm-Only*) plus pitch increase (*Pitch-Only*), relative to the control condition, were also computed to determine the predicted level of increase if the two feature-types' increases were simply added together (M = .07, SD = .23). Paired-samples *t*-tests were then carried out to compare these different values across participants.

Of primary interest is the fact that the likelihood of a "yes, familiar" response for test clips in the *Rhythm+Pitch* familiarization condition (M = .08, SD = .16) did not differ significantly from the predicted level of increase that would be expected from summing participants' *Rhythm-Only* and *Pitch-Only* familiarization increases (M = .07, SD = .23), $t(119)$ = 0.90, SE = .02, p = .37, BF_{01} = 6.66, 95% CrI [-.09, .25].[10] This suggests that the familiarity increase in the

Rhythm+Pitch condition indeed approximated the predicted increase (predicted based on what

Rhythm-Only + *Pitch-Only* would equal).

Another way of going about testing the match of the present pattern to the pattern

expected if the additivity assumption holds is to compare the likelihood of a "yes, familiar"

response in the *Rhythm+Pitch* condition to the value .065 (rounded to .07 above), which, as

mentioned above, was the mean value of (*Rhythm-Only* cues) + (*Pitch-Only* cues) across

participants. A one-sample *t*-test comparing the probability of responding "yes, familiar" in the

Rhythm+Pitch condition with the value .065 revealed no significant difference, $t(119) = 1.05$, *SE*

$= .01$, $p = .30$, $BF_{01} = 5.79$, 95% CrI [-.08, .27].[11]

In short, the participants' behavioral data suggest that separately encoding a song's

isolated rhythm and its isolated pitch (across separate encoding episodes) leads the feature

match assessment from those separate feature-types to combine additively to increase the

likelihood of perceived familiarity with the piano song clip in which both of those features are

embedded, and this additivity can be captured in their yes-no responses. It is also clear that the

two feature-types do not combine multiplicatively. The presence of two different familiarized

features in the test cue conferred no added familiarity advantage over and above the expected

increase of each different feature-type added together. That is, there was not a "whole is

greater than the sum of its parts" effect to having more than one familiarized feature-type

present in the test cue, even though such an idea is evolutionarily plausible.

Model Simulations of The Behavioral Yes-No Data. Our behavioral data exhibited

evidence of adherence to the additivity principle, and this was captured in the yes-no responses

of our participants. As we demonstrated in our earlier proof of principle simulations, the

principle of additivity is present in MINERVA 2's mechanisms underlying mean echo intensity values. Having shown that our behavioral yes-no data exhibited evidence of adherence to additivity, our next step was to simulate our actual data patterns in the form of yes-no responses (our behavioral measure in the present study) using MINERVA 2.

As mentioned when presenting the proof of principle simulations in the introduction, the precedent for simulating unstudied items in a recognition memory paradigm is to assume that test probes in the unstudied condition match zero items in the memory store (e.g., Hintzman, 1988). However, as is clear from our behavioral data above, participants still find probes corresponding to unstudied song features to be quite familiar. This suggests some degree of baseline familiarity present in the stimuli; indeed, the song stimuli used in the present study were all likely to be present in participants' general knowledge base prior to beginning the experiment.

Therefore, in our effort to simulate our actual data patterns using MINERVA 2, we modified the simulation tool used for the proof of principle demonstration in the introduction to incorporate baseline familiarity. Toward this end, we generated preexisting knowledge traces that varied in their baseline frequency levels. Using a methodology similar to that presented in the introduction, these preexisting knowledge traces were randomly generated such that each feature was either 0, +1, or -1 (note that these preexisting memory traces were designed to represent the complete version of the song, such that we did not zero-out part of the trace to represent an isolated feature, as this would more likely resemble song exposures in real-life).

To mimic how different stimuli have varying baseline familiarity levels within memory, we randomly assigned each memory trace to a repetition (i.e., frequency of prior exposure) value, which determined how many representations of that specific memory trace would be added into the preexisting knowledge base. The repetition values were randomly generated using a normal distribution centered at one with a standard deviation of 7, rounding each value to an integer. In order to ensure that there were no negative repetition values, as there cannot be a preexisting memory trace with zero or fewer representations, we then took the absolute value of each repetition value and added 1 (e.g., a generated repetition value of -1 would be transformed into 2). To illustrate how these repetition values were used to represent varying degrees of memory representation, Memory Trace$_1$ might be randomly assigned to a repetition value of 1 while Memory Trace$_2$ might be randomly assigned to a repetition value of 3. When encoding these memory traces into the model, Memory Trace$_1$ would be added once while Memory Trace$_2$ would be added three times, resulting in these traces having low and high degrees of baseline familiarity, respectively. Additionally, when adding the preexisting knowledge memory traces into the model as a function of their repetition value, we also applied noise to the traces, such that there was a 45% probability that each feature within the memory trace would be set to 0.

With the preexisting knowledge traces established, we then randomly selected a subset of the preexisting knowledge memory traces and assigned them to the One Feature-Type Studied condition (note that since our human data did not reveal a difference in feature weightings between the Rhythm and Pitch feature-types, we modeled this condition as just having one feature-type studied, irrespective of the specific musical feature-type; thus, the One

Feature-Type Studied condition was meant to mimic the Rhythm-only and the Pitch-only conditions). Using the methods described in the Introduction, we zeroed out a proportion of these studied trace features (30%) to represent studying only an isolated feature-type. Additionally, when adding these traces back into the model only once, we incorporated noise such that there was a probability of 20% that each feature would be changed to 0. The same methodology was used when creating the Two Feature-Types Studied Condition (meant to mimic the Rhythm+Pitch condition), but with the difference of adding each of the study traces into the model twice. The remaining preexisting knowledge memory traces were assigned to the Unstudied condition, such that there were 2,500 probes in each of the conditions.

Normal distributions were then fit to each of the simulated echo intensity conditions, as can be seen in Figure 7. In order to assess discriminability and further align our simulations with our behavioral data, an echo intensity criterion threshold was set at the mean of the One Feature-Type Studied condition (1.10×10^{-5}). Echo intensities were categorized as a response of "no" ("unfamiliar") if they were lower than the threshold and "yes" ("familiar") if they were greater than the threshold.

With the aforementioned parameters, the proportion of "yes, familiar" responses given by the model was .34 for the Unstudied condition, .42 for the One Feature-Type Studied condition, and .47 for the Two Feature-Types Studied condition. The d' value for the model's performance discriminating cues corresponding to features from the Unstudied condition versus to cues corresponding to features from the One Feature-Type Studied condition was .21 (d' for our behavioral data was .08). The d' value for the model's performance discriminating cues corresponding to features from the Unstudied condition and the Two Feature-Types

Studied condition was .33 (*d'* for our behavioral data was .20). Although we are demonstrating

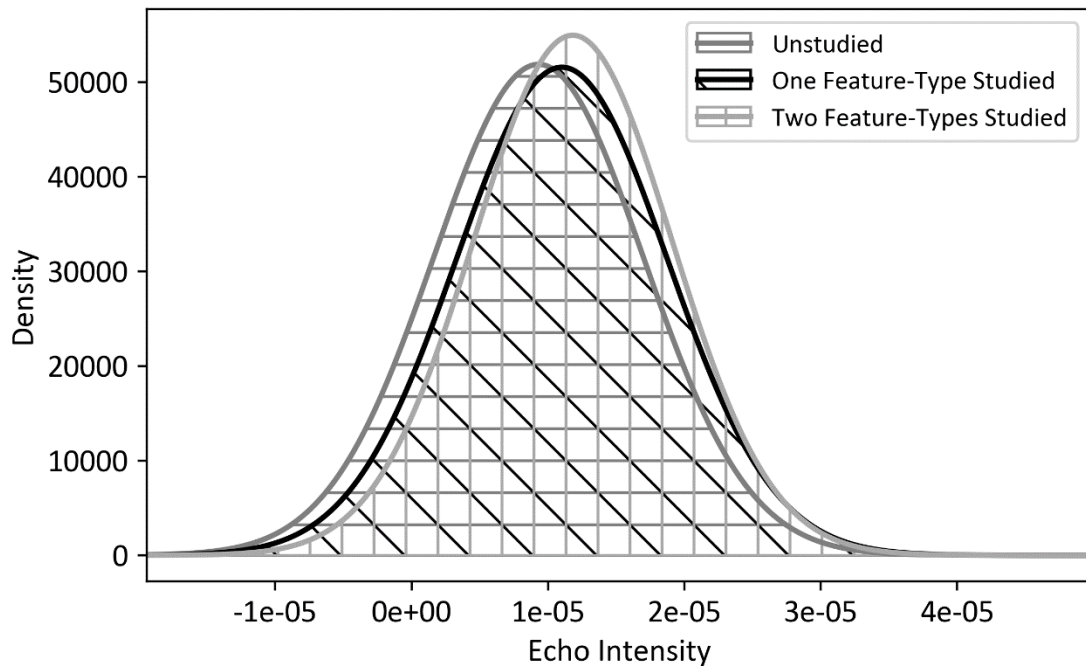higher levels of overall discriminability in our simulations than in our behavioral data, note that

Figure 7. Echo intensity distributions as a function of simulated study condition.

the differences between the *d'* values across the conditions are the same (a *d'* difference of .12

in both the simulated and behavioral data), and as will be discussed in the General Discussion, it

may be advantageous to demonstrate adherence to additivity in the model under

circumstances of higher discrimination than that which was shown in our participants.

As was done with our behavioral yes-no data, the level of increase from baseline was

calculated for each of the studied conditions to determine whether the *predicted* level of

increase for the Two Feature-Types Studied condition would approximately equal that of the

actual level of increase for the Two Feature-Types Studied condition. Our first approach was to

examine the level of *actual* increase from the overall probability of a "yes, familiar" response in

each condition (reported above), out of the 2500 simulation trials. The level of increase from baseline (.34) for the Two Feature-Types Studied condition (.47) was .13. The level of increase from baseline for the One Feature-Type Studied condition (.42) was .08. By multiplying the level of increase from baseline for the One Feature-Type Studied condition (.08) by 2, we arrived at .16 as the *predicted* level of increase.

Although .16 and .13 appear similar, it is impossible to assess whether these two single numerical values are significantly different from one another; therefore, in order to attempt to assess statistical significance, we divided the 2,500 simulated trials into segments of 25 trials each to achieve 100 simulated runs meant to mimic separate participants. As mentioned in the introduction regarding the proof of principle simulations, there is no between-subject variability in model simulations the way that there would be for human participants; however, for the present simulation, an arbitrary segmentation of the data at least allows for multiple data points for the probability of responding "yes, familiar" by the model across the conditions. This, in turn, enables a statistical analysis to be carried out. We chose 100 runs per condition to achieve a relatively high "subject" N for the analyses while having a reasonable number of items (25) per condition (note that there were 21 items per studied condition among actual participants in Experiment 2).

Thus, we segmented the data into 100 separate 25-item runs and computed the probability of a "yes, familiar" response for each 25-item run for each condition. From these, analogously to what was done with our human participants, we computed the amount of increase from baseline for the One Feature-Type Studied condition and for the Two Feature-Types Studied condition for each 25-item run. For each 25-item run, we also computed the

predicted amount of increase from baseline analogously to how it was done for the human

participant data (the amount of increase for the One Feature-Type Studied condition multiplied

by two). This allowed us to carry out a *t*-test analogously to that which was carried out among

the human participant data, comparing the *predicted* amount of increase from baseline, given

the principle of additivity, and the *actual* amount of increase from baseline shown in the

model's yes-no responses. As the two data sets (predicted and actual) were positively

correlated $r(100) = .47$, $p < .001$, we ran a paired-samples *t*-test to assess whether they differed

significantly from one another. The predicted increases ($M = .16$, $SD = .27$) did not differ

significantly from the actual increases ($M = .13$, $SD = .13$), $t(99) = 1.18$, $SE = .02$, $p = .24$, $BF_{01} =$

4.60.

In short, the MINERVA 2 model was able to closely mimic our experimental results in the

form of yes-no judgment patterns. Thus, the additivity principle that is present within the

model's echo intensity mechanisms can be captured with yes-no judgments, and one way in

which it does so closely mimics empirical findings with human participants in our paradigm. In

the General Discussion, we will discuss some of the possible constraints on the model's ability

to demonstrate additivity in yes-no responses, among them being 1) whether preexisting

knowledge is built into the model in the form of a baseline familiarity distribution, 2) whether

there are highly overlapping distributions, and 3) whether there are equal or approximately

equal variances across the distributions.

Experiment 3

The results of Experiment 2 suggest the presence of the additivity principle in our data,

as well as in model simulations that were intended to mimic our experimental data. Encoding

different feature-types in isolation across study episodes led to an additive increase in

perceived familiarity with a test cue containing both of those features. Another prediction

pertaining to how features should combine across study episodes to contribute to the

familiarity signal is that if a feature set of a single feature-type is repeated across different

study episodes, such as if an isolated rhythm were repeated several different times throughout

the encoding phase, each repetition of that feature-type should additively increase the later

familiarity signal elicited by the test cue containing that feature. This threefold increase was

demonstrated in the second set of proof of principle simulations reported in the introduction.

Whether participants' yes-no data exhibit this pattern was examined in Experiment 3, as well as

whether the participants' data patterns could be simulated using MINERVA 2.

Method

Participants. One hundred twenty-eight Colorado State University undergraduate

students completed the experiment for course credit. Six participants were lost due to not

finishing or due to computer crashes, leaving 122 participants.

Materials. The stimuli consisted of 96 of the piano song clips and isolated rhythms used

in Kostic and Cleary (2009).[12] Because of the added repetitions at encoding, for the sake of

time, isolated tonal sequences were not included in the study lists—only rhythms were used.

Procedure. The procedure was similar to that used in Experiment 1, but only isolated

rhythms were used. The participants each completed one of four counterbalanced versions of

the experiment on a computer with speakers. The 96 piano song segments were divided into

eight study-test blocks, with each study list consisting of 12 isolated rhythm presentations.

Three of the isolated rhythm presentations corresponded to the Rhythm1X condition and

therefore to a unique song each. The Rhythm3X condition also corresponded to three unique

songs; however, to incorporate repetition, each unique song has its isolated rhythm presented

three separate times at study, such that the Rhythm3X condition corresponded to nine isolated

rhythm presentations. These repetitions were randomly dispersed throughout the study list. At

test, each test list contained 12 test song clips (three Rhythm1X, three Rhythm3X, and six with

Unstudied features). Overall, there were 24 items in the Rhythm 1X condition, 24 items in the

Rhythm3X condition, and 48 items in the Unstudied condition.

As in Experiments 1 and 2, at study, participants were instructed to attempt to identify

the song from its isolated rhythm if possible when it was presented at study. Retrieval success

status was used to separate test songs into identified-at-study or unidentified-at-study

categories for data analysis. At test, participants were presented with the unaltered piano song

clips and asked to indicate whether each test song clip seemed familiar (Y=Yes, N=No). They

were also asked to attempt to identify the song.

Results and Discussion

Identification Rates

Identification of songs from isolated rhythm during encoding was once again at floor.

However, the proportion of songs identified at study from isolated rhythms presented once (*M*

= .01, *SD* = .02) was significantly lower than the proportion of songs identified from isolated

rhythms presented three times (M = .01, SD = .03), $t(121)$ = -2.07, SE = .003, p = .04, BF_{01} = 1.26,

95% CrI [-.36, -.01].

The proportion of piano song clips at test that had had their rhythms studied once (M =

.19, SD = .13) did not differ significantly from the proportion that had not had their rhythms

studied (M = .19, SD = .11), $t(121)$ = .25, SE = .01, p = .80, BF_{01} = 9.64, 95% CrI [-.20, .16], nor did

the proportion of piano song clips that had had their rhythms studied three times, (M = .19, SD

= .14), $t(121)$ = .41, SE = .01, p = .67, BF_{01} = 9.12, 95% CrI [-.21, .15]. There was also no

difference in the probability of identifying a piano song clip that had had its rhythm studied

once versus three times, $t(121)$ = .16, SE = .01, p = .87, BF_{01} = 9.82, 95% CrI [-.16, .19].

The Effect of Rhythm Exposure on Perceived Test Song Clip Familiarity

Behavioral Yes-No Data. We now turn to the data of primary interest, which are

familiarity judgments for piano song clips at test that could not be identified from their rhythms

at study. A repeated measures ANOVA performed on the probability of responding "Yes,

familiar" across the Rhythm Exposure conditions revealed a significant main effect of Rhythm

Exposure Level (zero vs. one exposure vs. three exposures), $F(2, 242)$ = 10.50, MSE = .01, p <

.001, η_p^2 = .08, BF_{01} = .003 (see Figure 8).[13] Piano test song clips that had had their rhythms

familiarized once at encoding were more likely to be judged as familiar (M = .55, SD = .17) than

test song clips that had not had their rhythms familiarized, (M = .53, SD = .16), $t(121)$ = 2.14, SE

= .01, p = .04, d = .15, BF_{01} = 1.11, 95% CrI [.02, .37].[14] This replicates the results of Experiment

1's rhythm condition.

Of particular interest to the current experiment is the finding that test song clips that

had had their rhythms familiarized three times were judged as familiar significantly more often

(M = .58, SD = .15) than test song clips that had their rhythms familiarized only once, $t(121)$ =

2.36, SE = .01, p = .02, d = .19, BF_{01} = 0.69, 95% CrI [.03, .39]. This suggests that a single feature-

type that is repeated across study episodes leads to a combined increase in familiarity across

the memory traces for those episodes in response to the test cue. Not surprisingly, the

proportion of songs judged as familiar was higher when their rhythms were studied three times

compared to when their rhythms were not studied, $t(121)$ = 4.72, SE = .01, p < .001, d = .35, BF_{01}

= 4.41 x 10^{-4}, 95% CrI [.24, .60].[15] Figure 8 illustrates that increasing exposure to a song's

isolated rhythm during encoding subsequently increases the probability of reporting a whole

song as familiar.
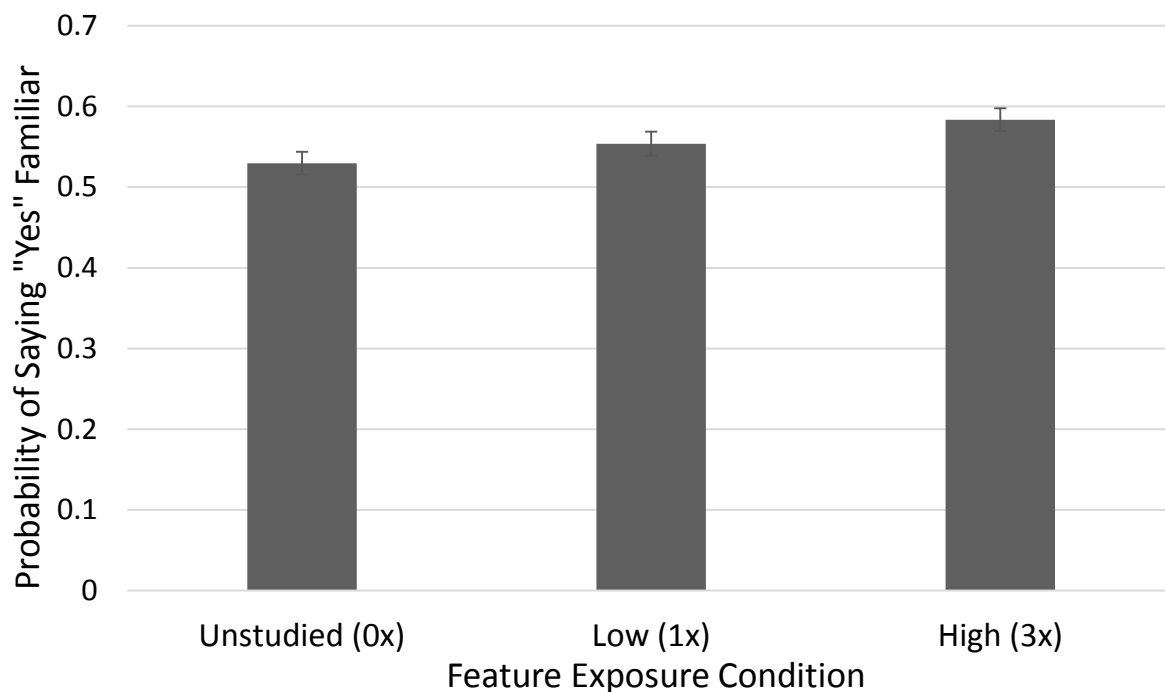
Figure 8. The probability that a piano test song clip was judged as familiar as function of the number of exposures to its isolated rhythm at encoding in Experiment 3. The probability of judging a song clip as familiar increased with an increasing number of exposures to the song's rhythm in the encoding phase.

Our next priority was to examine whether the level of familiarity increase for test clips

whose rhythms were repeated three times was the amount that would be expected if the

Activation Values for each encoded rhythm episode were added together to increase the level

of familiarity for the test probe (see Figure 1). Specifically, does the level of familiarity increase

found for test song clips whose rhythms were presented three times conform to the following

formula?

$$Rhythm3x \text{ cues} = 3[Rhythm1x \text{ cues}]$$

To examine this, the same approach taken in Experiment 2 was applied to the data of

Experiment 3. For each participant, we computed the amount of increase in the likelihood of

reporting "yes, familiar" between test song clips in the control condition (no rhythm

familiarization, or *Rhythm0x*) and in the rhythm-once-familiarized (*Rhythm1x*) condition (M =

.02, SD = .12), and the amount of increase between test song clips in the control condition

(*Rhythm0x*) and the rhythm-three-times-familiarized (*Rhythm3x*) condition (M = .05, SD = .13).

For each participant, we multiplied the person's Rhythm1x increase from the control

(*Rhythm0x*) condition (M = .02, SD = .12) by three (M = .07, SD = .37) to determine if the

familiarity increase caused by repeating the rhythm three times was additive. Paired-samples *t*-

tests were then carried out to compare the expected and the obtained values.

A paired-samples *t*-test revealed that there was no significant difference between the

level of increase in judging a test clip as familiar when its rhythm was presented three times

(*Rhythm3x*) (M = .05, SD = .13) and the number arrived at by multiplying by three participants'

probability of judging a song clip whose rhythm was presented once [3(*Rhythm1x*)] (M = .07, *SD*

= .37), $t(121)$ = 0.59, SE = .03, p = .56, BF_{01} = 8.39, 95% CrI [-.23, .12].[16] In short, there was no

evidence that features combine multiplicatively across study episodes when present in a test

probe. Instead, the assumption present in the MINERVA 2 model depicted in Figure 1 appears

to hold up; the Activation Values for each rhythm instantiation would be expected to be

comparable and to combine additively to increase the familiarity signal. The present results are

consistent with this assumption.

Another way of going about testing the match of the present pattern to the pattern

expected if the additivity assumption holds up is to compare the likelihood of a "yes, familiar"

response in the rhythm-three-times-presented (*Rhythm3x*) condition to the value .07, which, as

mentioned above, was the mean value of multiplying, for each participant, the probability of a

yes response in the rhythm-once-presented (*Rhythm1x*) condition by three (which is equivalent

to adding the level of familiarity increase three times, once for each additional presentation at

study). A one-sample *t*-test comparing the obtained probability of responding "yes, familiar" in

the rhythm-three-times-presented (*Rhythm3x*) condition with the value .07 revealed no

significant difference, $t(121) = 1.41$, $SE = .01$, $p = .16$, $BF_{01} = 3.81$, 95% CrI [-.30, .05].[17] In short,

the results converge on the idea that the multiple presentations of the isolated rhythm in

different encoding episodes combined additively in response to the test song clip to increase

the test clip's level of familiarity.

Model Simulations of the Behavioral Yes-No Data. As in Experiment 2, we conducted

an additional MINERVA 2 simulation aimed at mimicking our yes-no data from Experiment 3. As

was done in Experiment 2's simulations, we incorporated preexisting knowledge in Experiment

3's simulations. We used the same methodology and parameters presented above, with the

exception that in place of the previous Two Feature-Types Studied condition, we now have a

Studied3X condition, as only one feature-type (rhythm) was studied, and it was studied three times in this condition. Thus, test probes presented to the model either corresponded to preexisting knowledge or baseline familiarity (the Unstudied condition), to the rhythm feature-type stored once from the study phase in memory over and above the baseline frequency (the Studied1X condition), or to the rhythm feature-type stored three times from the study phase in memory over and above the baseline frequency (the Studied3X condition). Normal distributions were fit to the echo intensities of the three conditions and are presented below in Figure 9.
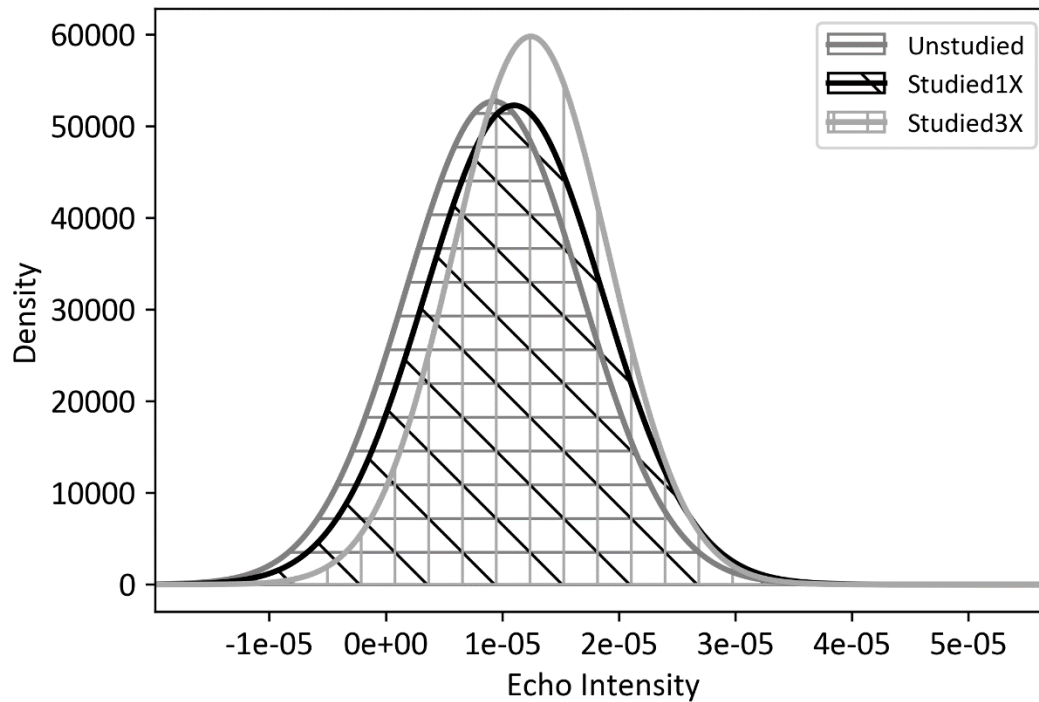
Figure 9. Echo intensities as a function of simulated study condition.

Similar to in Experiment 2, we set the echo intensity criterion threshold at the mean of the Studied1X condition (1.10×10^{-5}). In categorizing echo intensities as either "yes, familiar" or "no, unfamiliar," for the Unstudied condition, the proportion of "yes, familiar" was .35, for the Studied1X condition it was .42, and for the Studied3X condition it was .52. The *d'* value for the model's performance discriminating cues corresponding to features from the Unstudied

condition versus to cues corresponding to features from the One Feature-Type Studied

condition was .20 (*d'* for our behavioral data was .05), and the *d'* for the model's performance

discriminating cues corresponding to the Unstudied condition from cues corresponding to the

Studied3X condition was .45 (our behavioral data had a *d'* of .13 for this condition).

As in Experiment 2 and as was done with our behavioral data, we calculated the level of

increase from baseline for each of the studied conditions (Studied1X and Studied3X) to

determine whether the *predicted* level of increase for the Studied3X condition (i.e., the level of

increase from baseline for the Studied1X condition multiplied by 3) would be approximately

that of the *actual* level of increase for the Studied3X condition. The actual level of increase from

baseline for the Studied3X condition was .17. The level of increase from baseline for the

Studied1X condition (.074) multiplied by 3 is .22. To allow for a statistical analysis of the data, as

was done in our Experiment 2 simulations, we segmented the data from the current

simulations into 100 sets of 25 for each condition. As in the Experiment 2 simulations, the two

data sets under comparison (predicted vs. actual) were significantly correlated, r(100) = .51, *p* <

.001; therefore, we ran a paired-samples *t*-test on the predicted and the actual values for the

simulated participants. The predicted increase (*M* = .22, *SD* = .43) did not differ significantly

from the actual increase (*M* = .17, *SD* = .14), $t(99) = 1.31$, *SE* = .04, *p* = .19, $BF_{01} = 3.96$.

General Discussion

The Role of Features in the Computation of the Familiarity Signal from Music

At the heart of all human memory research is the question of how information is

represented in the human mind then used within memory processes. A long-standing prevailing

assumption in human memory theory, starting at the dawn of cognitive psychology as a field, is

the feature assumption (e.g., Estes, 1950). The feature assumption is the idea that memory

traces are essentially sets of tied-together features from the original experience that they are

meant to represent or record. Though the feature assumption was prevalent in memory theory

from at least the 1950s (e.g., Estes, 1950) onward, substantial research aimed at systematically

identifying the features of memory traces began to take place over the past 20 years or so (e.g.,

Cleary & Greene, 2000, 2001; Cleary, 2004; Cleary, 2014; Cleary et al., 2004; Cleary et al., 2007;

Cleary et al., 2016; Kostic & Cleary, 2009; Ryals & Cleary, 2012) and progress in this area

continues today. The present study contributes to this progress by isolating two musical

feature-types—rhythm and pitch—from their wholes, and determining how these isolated

features in memory traces behave in the computation of the familiarity signal for musical pieces

played on the treble clef of a piano. Specifically, rhythm and pitch were separately isolated

from their original piano song clips to examine how these features, when encoded in isolation,

combine across memory traces for study episodes to produce the familiarity signal experienced

with the later test song clip in which those features are embedded.

Adherence to the Additivity Assumption Present within the MINERVA 2 Model

The results of the present study converge on the idea that the feature-matching process

that enables familiarity-detection during retrieval failure for musical stimuli operates in a

manner similar to that specified by the MINERVA 2 global matching model. First, at the most

basic level, features (e.g., rhythm) that were encoded in isolation but that were later embedded

within a test song segment increased the likelihood of a reported sense of familiarity with that

test song clip. This is consistent with the idea that test items are decomposed into their

component features then matched with features stored in memory as part of the familiarity

computation.

Second, differentially familiarizing different types of features in the encoding phase led

to different levels of overall familiarity with the full test song segment in which they were

embedded. For example, familiarizing only the song's rhythm or only the song's pitch sequence

information led to lower overall familiarity than separately familiarizing (across different study

episodes) both the rhythm and the pitch information (Experiment 2). Also, familiarizing just one

feature-type—the song's rhythm—multiple times throughout the encoding phase led to greater

familiarity with the later whole (the full piano song segment) than familiarizing it only once

during the encoding phase (Experiment 3). These patterns suggest that the feature-matching

process that was giving rise to the perceived familiarity with the test song clips was global in

nature. That is, the features of memory traces from different study episodes were combining to

collectively boost the familiarity signal with the test cue, as the MINERVA 2 model would

predict.

Importantly, the *particular manner* in which the isolated feature-types present in

memory traces combined across encoded episodes to contribute to the familiarity signal is

consistent with the additivity principle that is embedded within the mechanisms of the

MINERVA 2 model. As shown in Figure 1 and in our simulations, in the MINERVA 2 model, once

the degree of feature-match is computed for each memory trace (assessing the degree of

match between the memory trace and the test probe), that index of match (the Activation

Value) for each memory trace combines additively to produce the familiarity signal. This means

that the level of feature-match produced by a memory trace for a particular isolated feature-

type from a particular encoding episode should combine additively with that produced by

another isolated feature-type from a different encoding episode. This is because the Activation

Values are added together to produce the familiarity signal (i.e., echo intensity). This additivity

principle was demonstrated in our proof of principle simulations in the introduction with

respect to mean echo intensity values.

The present behavioral results and corresponding simulations of our participant

behavioral data conformed to this additivity principle. First, as shown in Experiment 2, the

isolated feature-types of rhythm and pitch combined across memory traces for separate study

episodes in an additive fashion as captured in participant yes-no (familiar-unfamiliar)

responses; this additivity was also captured in the model's yes-no responses using parameters

that enabled the model to approximately mimic our participant data. Second, as shown in

Experiment 3 and captured in yes-no (familiar-unfamiliar) judgments, the isolated feature-type

of rhythm, when repeated multiple times across separate study episodes, also combined

additively to increase the likelihood of perceived familiarity with the full test song clip in which

that feature was embedded. This pattern of additivity was also captured in the model's yes-no

responses using parameters that enabled the model to approximately mimic our participant

data.

These patterns in our participant data as captured in yes-no (familiar-unfamiliar)

judgments are informative from a theoretical perspective because other possibilities were

equally plausible beforehand, if not more so. For example, the idea that features might

combine across study episodes in a multiplicative, rather than an additive, fashion seems

evolutionarily plausible. This is because if different feature types that all happen to be

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

familiarized are co-occurring together at once within a cue, that joint co-occurrence of familiar

features could conceivably itself be informative. That is, the mere joint presence of multiple

familiarized features together in one episode could, in and of itself, indicate an increased

likelihood that the episode was experienced before. This might be a reason to expect an

advantage to multiplicative feature combination across episodes as opposed to additive

combination: There may be evolutionary advantages to a memory system in which the "whole"

of familiarity may be greater than the "sum of its parts." However, the present results do not

support multiplicative combination of features across different study episodes. Rather, they

support additive combination across episodes of the type shown in Figure 1.

Some Important Caveats and Considerations for Future Directions

Our primary aims in the present study were 1) to demonstrate, in a set of proof of

principle simulations, that the MINERVA 2 model has an additivity principle within its

mechanisms regarding how the features of different memory traces are combined across traces

for computing familiarity signal intensity (echo intensity), which, to our knowledge, has not

previously been explicitly shown in simulations, 2) to assess whether the embedding of musical

features that had been studied in isolation within a musical test cue would boost the perceived

level of familiarity with that cue in accordance with that additivity principle as measured by

participants' yes-no (familiar-unfamiliar) responses, and 3) to assess whether the MINERVA 2

model could mimic our actual data patterns with yes-no responses. To achieve the first end, we

ran simulations with different parameters (different rhythm-to-pitch ratios and different noise

levels) to assess whether the additivity principle would hold up across all of these hypothetical

situations in the form of mean echo intensities. As shown in these initial proof of principle

simulations, the additivity principle robustly manifested in the mean echo intensities across

these various parameters and simulated experimental situations.

To achieve the second end, we needed a behavioral measure of perceived familiarity

with the cue, and chose to use a simple yes-no response on each trial ("yes, familiar" or "no,

not familiar) to achieve that end. In choosing our index of perceived familiarity, we made the

assumption that differences in the frequency of reporting perceived familiarity across our

conditions would be a reflection of how familiar, on average, test cues in each of those

conditions tended to seem to participants. The additivity principle shown in our MINERVA 2

simulations of echo intensity values was indeed captured in our dependent measure of yes-no

familiarity. Furthermore, with parameter settings that enabled the model to approximately

mimic our data patterns, the model also demonstrated additivity in the form of yes-no

judgments.

That said, it is important to consider some potential limitations on capturing the

additivity principle using this measure and possible reasons why we were able to capture it in

the present study and with the particular parameters used in our MINERVA 2 simulations of our

actual data patterns. A potential limitation on capturing the additivity principle is that whether

the additivity principle translates to simulated yes-no behaviors on the part of the model itself

will necessarily depend on a number of parameter settings within the model. As an example of

how the additivity principle that occurs with the echo intensity values can get masked in some

situations, when the model is set to produce yes-no responses via criterion placement along the

echo intensity continuum rather than simply outputting the echo intensity values themselves,

the distributions (i.e., how far apart they are and how equal or unequal their variances are) and

the criterion placement can matter to whether the additivity principle will manifest in the form of yes-no judgments.

As an extreme example of why this is, one can consider what happens if the criterion is placed too far to the right of all distributions (an overly conservative criterion). The model will always respond "no" and there will be no additivity shown. Conversely, if the criterion is placed too far to the left of all distributions (an overly liberal criterion), the model will always respond "yes," and no additivity will be shown. As another less extreme example, if discriminability is too high (i.e., the distributions in question are too far apart along the echo intensity continuum), additivity will not be exhibited in the dichotomous response pattern (because the model will be too good at discrimination, hitting ceiling by showing perfect discrimination performance in at least one or more of the conditions). For example, as is apparent in Figure 2 and Figure 3, the distributions are far enough apart in those proof of principle simulations that placing a criterion somewhere in the leftward distribution means a near 100% "yes" rate for the rightward distribution (because the distributions are relatively far apart on the familiarity continuum, with not very much overlap).

Our human participants did not perform with nearly such high levels of discriminability between the distributions as occurred in our proof of principle simulations. The boosts to familiarity from the feature familiarization manipulations in our study are subtle and small, suggesting highly overlapping distributions, whereby the distributions move ever so slightly to the right with increasing feature familiarization through our manipulations. In addition, the false alarm rates in our study were quite high, suggesting that our stimuli were high in preexisting or baseline familiarity; indeed, the song clips used at test (e.g., London Bridge) were

highly familiar children's songs, nursery rhymes, and pop songs that would have existed in participants' knowledge-bases prior to the experiment. Specifically, our data suggest that moving from the *Rhythm-Only* or *Pitch-Only* conditions to the *Rhythm+Pitch* condition, or from the *Rhythm1x* to the *Rhythm3x* condition, involves a slight shift to the right along the familiarity continuum from a preexisting baseline familiarity distribution on that continuum, with all three distributions in question for each experiment being highly overlapping. Indeed, we were able to simulate our basic findings with MINERVA 2 when creating such conditions in the model, even with higher discrimination levels than those shown by our participants. This latter point is important given that extremely low levels of discrimination can potentially force a linear pattern, thereby forcing adherence to additivity. Therefore, demonstrating that the pattern can be shown in the model with yes-no responses under circumstances of higher discrimination than that shown among participants in the present study is important. Future research should aim to investigate whether manipulations that push discrimination levels even higher (such as increased amounts of repetition) continue to lead to additivity in the data patterns among human participants.

Another key factor that may be at work in determining when the MINERVA 2 model shows additivity in yes-no responses is the degree to which the variances are equal versus unequal. The additivity pattern that is present among the mean echo intensities within the model will be better able to be captured in the form of yes-no judgments when the variances of the distributions are equal. Although increasing the frequency of memory traces has been shown to also increase the variances of the underlying distributions in MINERVA 2 (Hintzman, 1988), the differences in variances across the distributions are amplified when the starting

point for the differing distributions is zero, or no existing traces in the memory store. Under these circumstances, as the frequency of a given memory trace is increased (from 0x to 1x to 2x to 3x to 4x to 5x), the distribution widens with each incremental increase in frequency (see Hintzman, 1988, p. 531, Figure 4). However, the increases in variance are attenuated in situations in which the starting point is not zero, but rather, is a pre-experimental store of existing knowledge traces having varying preexisting frequencies. The presence of preexisting traces act as a sort of buffer against dramatically changing variances with study list presentation. This principle is demonstrated in the present study's simulations (see Figures 7 and 9). Thus, a key factor in determining whether the MINERVA 2 model will show additivity in behavioral responses or not may be the extent to which preexisting knowledge is built into the model to enable a closer approximation to equal variances across the distributions. Future research should aim to simulate different levels of preexisting knowledge (i.e., different baseline frequencies of memory traces) and the effects of this on the incremental rightward shifts of distributions with different frequencies of study list presentation as well as on the variances of those shifted distributions and how this effects adherence to additivity.

General Theoretical Implications

The fact that our paradigm captured the additivity principle in different ways (i.e., in both the *Rhythm+Pitch* condition and in the *Rhythm3x* condition) represents an interesting empirical demonstration with theoretical implications – irrespective of the particular theoretical model – regarding how features combine across memory traces to produce a familiarity signal with a cue in which those features are embedded. One reason why the additivity effects inform theory even without reference to MINERVA 2 is that multiplicative,

rather than additive, combination across traces seemed more evolutionarily plausible *a priori*,

as additional importance can be carried by the fact that features in the cue had occurred on

separate occasions rather than only once, and this form of additional information could be

captured in a multiplicative model. A multiplicative model would be able to capture a "whole is

greater than the sum of its parts" notion in its mechanisms. Thus, the fact that we consistently

found additivity, rather than multiplicative combination, is somewhat surprising outside of the

context of MINERVA 2. Our proposed reason, *a priori*, for hypothesizing additivity was our

proposal that what gives rise to familiarity-detection with a musical cue in which familiarized

features are embedded is a feature-matching process akin to that taking place in the MINERVA

2 model, which has within it this counterintuitive principle of additivity. Ultimately, this enabled

our discovery of the additivity principle in our paradigm.

Future research should explore the precise mechanistic reasons why our paradigm is

able to capture additivity using yes-no judgments of familiarity, and should do so in ways that

include other theoretical possibilities besides MINERVA 2's mechanisms. The MINERVA 2 global

matching model is likely not the only model that might accommodate the manner in which

isolated features appear to combine across memory traces for study episodes to produce

familiarity-detection during retrieval failure. An assessment of fit of different particular models

goes beyond the scope of the present study, as the primary focus of the present study was on

the global matching assumption, and whether features combine additively or multiplicatively

across traces of episodes in the computation of the familiarity signal. Future research might

focus instead on the specific manner in which different global matching models compute the

global match of the test probe to the memory traces when it comes to familiarity-detection

during retrieval failure.

Also, newer dynamic models, such as that proposed by Cox and Shiffrin (2017), are

worth considering in future research on familiarity-detection during retrieval failure. Future

research might even examine the temporal unfolding of familiarity-detection as a dynamic

process within the context of this type of model. For example, musical stimuli are temporally

dynamic in nature and thus lend themselves to investigation of their temporal unfolding over

time. Future research might therefore examine the temporal accumulation of feature

information in the computation of familiarity by having participants press a button as soon as a

test stimulus becomes familiar. This approach might allow for a testing of predictions made by

dynamic models of the familiarity computation.

Another important consideration is that other models not typically intended to account

for the computation of the familiarity signal might also be able to account for the manner in

which features are combined across memory representations to produce familiarity. For

example, the notion of features in memory representations is present within neural network

models that have distributed representations (e.g., McClelland & Rumelhart, 1985; Plaut, 1995;

Seidenberg & McClelland, 1989), and feature overlap can occur between a current situation (a

test cue) and representations within the distributed network in memory. Although such neural

network models of linguistic or other distributed featural representations do not tend to

incorporate familiarity signals within their mechanisms, a distributed neural network model

could somewhat easily incorporate a signal mechanism that specifically indexes subthreshold

but "nearing threshold" activation as a potential index of familiarity intensity. This could be

done, for instance, by having the activation level itself for a representation (or representations)

within the network, which is not usually considered the output in such models but is there

nonetheless, used as a form of output that would index the feature-overlap-based familiarity

signal in response to a cue. The signal would then index the amount of feature overlap between

the test cue and the representations in memory and could potentially accommodate global

feature-matching effects in familiarity-detection that occurs during retrieval failure. This is

another potential avenue for future research.

That said, continuing to investigate these feature-based familiarity-detection effects

within the context of MINERVA 2 is likely to also be worthwhile, given that mathematical work

has suggested that the principles underlying the MINERVA 2 model scale up in a manner

applicable to implementation in a distributed neural network model (Kelly, Mewhort, & West,

2017). Thus, MINERVA 2's principles are neurally plausible regarding what may underlie

familiarity-detection from feature-matching, making it worth continued consideration.

At a very broad level, research on how exactly features of memory representations

behave to produce familiarity-detection has been somewhat stymied by the overarching

debates over the past few decades regarding dual- versus single-process accounts of

recognition memory. Because researchers have not been able to come to consensus regarding

whether more than one process contributes to old-new discrimination on standard recognition

memory tasks (e.g., Diana et al., 2006; Hintzman, 2011; Mandler, 2008; Mickes et al., 2009;

Onyper, Zhang, & Howard, 2010; Wais et al., 2008; Yonelinas, 2002), specific questions about

the nature of the familiarity signal that presumably gives rise to familiarity-detection have been

difficult to approach. It is important to recognize that although researchers do not agree on the

nature of the specific processes that give rise to discrimination on old-new recognition memory

tasks, people generally do not deny that familiarity-detection is a memory process that occurs

in the world (e.g., Mandler, 1991). As Hintzman (2011) points out, researchers often become so

focused on debating the processes that contribute to particular laboratory tasks, that they lose

sight of the bigger picture of examining memory processes that occur in the world. The task

used in the present study is not a standard old-new recognition memory task, and thus may not

be relevant to the ongoing debates about whether dual- versus single-process models are

better accounts of old-new discrimination in standard recognition tasks. The task used in the

present study was aimed at isolating and examining familiarity-detection when conscious

retrieval is absent, with the goal of better understanding this specific process (not whether it

contributes to old-new discrimination on standard recognition memory tasks).

In terms of better understanding the familiarity signal that presumably gives rise to

familiarity-detection during retrieval failure, other future directions include continued

exploration regarding how different feature-types are weighted in the computation of the

familiarity signal. Just because we found no evidence for differential weightings of different

feature-types in the present study (specifically, rhythm and pitch) does not mean all features of

all types are weighted equally in the computation of familiarity. There may even be *a priori*

reason for predicting some features to be more important than others in some situations. For

example, the language literature presents reason for predicting that first and last letters of

words may be more important than middle letters (e.g., Grainger & Whitney, 2004); thus, there

is *a priori* reason for predicting that first and last letters may carry more weight in the

computation of the familiarity signal for visual word stimuli.

Future research should also investigate whether other feature-types combine additively

or multiplicatively in the computation of familiarity. As mentioned in the present study, there

are *a priori* reasons for predicting that a multiplicative combinatorial feature-matching process

might be an adaptive feature of memory. Although the current findings do not clearly support a

differential role of rhythm and tone in the computation of familiarity, it may be that as

exposure to these feature-types increases, a non-linear relationship would emerge, such that

tone, when experienced multiple times, would carry more weight than rhythm. As

demonstrated in our simulations, differing feature-type ratios within a stimulus should not alter

the basic MINERVA 2 predictions regarding additivity, but discovering which feature-types

might carry more weight in the computation of the familiarity signal would increase

understanding of the role of various types of features in familiarity. Additionally, future

research should examine how features of other stimulus-types might combine differently, such

as semantics, phonemes, or geons.

Finally, another important question for future research concerns whether all stimuli

actually segment into features in the computation of familiarity. The literature suggests that

whereas some stimuli may be primarily segmented into component features for processing,

others may be processed more holistically (e.g., Meltzer & Bartlett, 2019; Moscovitch, Winocur,

& Behrmann, 1997). It is as yet unclear how or if this notion applies to the computation of the

familiarity signal. Future research should explore this question.

Summary

In summary, the present results constitute a first step toward understanding, at a

mechanistic level, how features combine across memory traces to produce a familiarity signal

that allows for a subjective sensation of familiarity with a stimulus to occur. Though the present

study provides a noteworthy beginning to understanding the specifics of how isolable features

in memory traces operate in the computation of the familiarity signal that allows familiarity-

detection during retrieval failure, it also raises important questions and directions for future

research.

References

Anisfeld, M., & Knapp, M. (1968). Association, synonymity, and directionality in false

recognition. *Journal of Experimental Psychology, 77*(2), 171-179. doi:10.1037/h0025782

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding.

Psychological Review, 94(2), 115-147. doi:10.1037/0033-295X.94.2.115

Bregman, A. S., & Chambers, D. W. (1966). All-or-none learning of attributes. *Journal of*

Experimental Psychology, 71, 785-793. doi:10.1037/h0023210

Brockdorff, N., & Lamberts, K. (2000). A feature-sampling account of the time course of old-new

recognition judgments. *Journal of Experimental Psychology: Learning, Memory, and*

Cognition, 26(1), 77-102. doi:10.1037/0278-7393.26.1.77

Brown, A. S. (2004). *The déjà vu experience*. Psychology Press, New York.

Brown, A. S. (2012). *The tip of the tongue state*. Psychology Press, Hove, UK.

Bower, G. A. (1967). A multicomponent theory of the memory trace. In K. W. Spence & J. T.

Spence (Eds.), *The Psychology of Learning and Motivation* (*Vol. 1*). New York: Academic

Press.

Clark. S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the

models match the data. *Psychonomic Bulletin & Review, 3*(1), 37-60.

doi:10.3758/BF03210740

Cleary, A. M. (2004). Orthography, phonology, and meaning: Word features that give rise to

feelings of familiarity. *Psychonomic Bulletin & Review, 3*(11)*,* 37-60.

doi:10.3758/BF03196593

Cleary, A. M. (2014). The sense of recognition during retrieval failure: Implications for the

nature of memory traces. *Psychology of Learning and Motivation, 60,* 77-112.

doi:10.1016/B978-0-12-800090-8.00003-2

Cleary, A. M., Brown, A. S., Sawyer, B. D., Nomi, J. S., Ajoku, A. C., & Ryals, A. J. (2012).

Familiarity from the configuration of objects in 3-dimensional space and its relation to

déjà vu: A virtual reality investigation. *Consciousness and cognition*, *21*, 969-975.

doi:10.1016/j.concog.2011.12.010

Cleary, A. M., & Claxton, A. B. (2018). Déjà vu: An illusion of prediction. *Psychological Science,*

29, 635-644. doi:10.1177/0956797617743018

Cleary, A. M., & Greene, R. L. (2000). Recognition without identification. *Journal of Experimental*

Psychology: Learning, Memory, and Cognition, 26, 1063-1069. doi:10.1037/0278-

7393.26.4.1063

Cleary, A. M., & Greene, R. L. (2001). Memory for unidentified items: Evidence for the use of

letter information in familiarity processes. *Memory & Cognition, 29*, 540-545.

doi:10.3758/BF03196405

Cleary, A. M., Huebert, A. M., McNeely-White, K. L., & Spahr, K. S. (2019). A postdictive bias

associated with déjà vu. *Psychonomic Bulletin & Review*, *26*, 1433-1439.

doi:10.3758/s13423-019-01578-w

Cleary, A. M., Langley, M. M., & Seiler, K. R. (2004). Recognition without picture identification:

Geons as components of the pictorial memory trace. *Psychonomic Bulletin & Review*, *11*,

903-908. doi:10.3758/BF03196719

Cleary, A. M., McNeely-White, K. L., Huebert, A. M., & Claxton, A. B. (2018). Déjà vu and the

feeling of prediction: An association with familiarity strength. *Memory (Special Issue on*

Déjà vu). doi:10.1080/09658211.2018.1503686

Cleary, A. M., Ryals, A. J., & Nomi, J. S. (2009). Can déjà vu result from similarity to a prior

experience? Support for the similarity hypothesis of déjà vu. *Psychonomic Bulletin &*

Review, 16, 1082-1088. doi:10.3758/PBR.16.6.1082

Cleary, A. M., Ryals, A. J., & Wagner, S. R. (2016). Recognition during recall failure: Semantic

feature matching as a mechanism for recognition of semantic cues when recall fails.

Memory & Cognition, 44(1), 50-62. doi:10.3758/s13421-015-0545-z

Cleary, A. M., Winfield, M. M., & Kostic, B. (2007). Auditory recognition without

identification. *Memory & Cognition*, *35*, 1869-1877. doi:10.3758/BF03192921

Chang, K. K., Mitchell, T., & Just, M. A. (2011). Quantitative modeling of the neural

representation of objects: how semantic feature norms can account for fMRI activation.

Neuroimage, 56, 716-727. doi:10.1016/j.neuroimage.2010.04.271

Cox, G. E., & Shiffrin, R. M. (2017). A dynamic approach to recognition memory. *Psychological*

Review, 124, 795-860. doi:10.1037/rev0000076

Diana, R. A., Reder, L. M., Arndt, J., & Park, H. (2006). Models of recognition: A review of

arguments in favor of a dual-process account. *Psychonomic Bulletin & Review, 13*, 1-21.

doi:10.3758/BF03193807

E-Prime (Version 2.0) [Computer software]. Pittsburgh, PA: Psychology Software Tools.

Estes, W. K. (1950). Toward a statistical theory of learning. *Psychological Review*, *57*(2), 94-07.

doi:10.1037/h0058559

Grainger, J., & Whitney, C. (2004). Does the huamn mnid raed wrods as a whole? *TRENDS in Cognitive Sciences, 8*(2)*,* 58-59. doi:10.1016/j.tics.2003.11.006

Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review, 114*(2), 211-244. doi:10.1037/0033-295X.114.2.211

Gronlund, S. D. & Ratcliff, R. (1989). Time course of item and associative information: Implications for global memory models. *Journal of Experimental Psychology*: *Learning*, *Memory*, *and Cognition*, *15*, 846-858. doi:10.1037/0278-7393.15.5.846

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93,* 411-428. doi:10.1037/0033-295X.93.4.411

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95*, 528-551. doi:10.1037/0033-295X.95.4.528

Hintzman, D. L. (2011). Research strategy in the study of memory: Fads, fallacies, and the search for the "coordinates of truth." *Perspectives on Psychological Science, 6*(3), 253-271. doi:10.1177/1745691611406924

Kelly, M. A., Dewhort, D. J. K., & West, R. L. (2017). The memory tesseract: Mathematical equivalence between composite and separate storage memory models. *Journal of Mathematical Psychology, 77,* 142-155. doi:10.1016/j.jmp.2016.10.006

Kluyver, T., Ragan-Kelley, B., Pérez, P., Granger, B., Bussonnier, M., Frederic, J., … Willing, C. (2016). Jupyter Notebooks – a publishing format for reproducible computational workflows. In F. Loizides & B. Schmidt (Eds.), *Positioning and Power in Academic Publishing: Players, Agents and Agendas* (pp. 87–90).

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of

knowing. *Psychological Review, 100*, 609-639. doi:10.1037/0033-295x.100.4.609

Kostic, B., & Cleary, A. M. (2009). Song recognition without identification: When people cannot

"name that tune" but can recognize it as familiar. *Journal of Experimental Psychology:

General*, *138*(1), 146-159. doi:10.1037/a0014584

Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental

Psychology: General*, *142*, 573-603. doi:10.1037/a0029146

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic

analysis theory of acquisition, induction, and representation of knowledge.

Psychological Review, 104(2), 211-240.

Malmberg, K. J. (2008). Investigating metacognitive control in a global memory framework. In J.

Dunlosky & R. E. Bjork (Eds.), *Handbook of Metamemory and memory* (pp. 265-283).

Psychology Press.

Mandler, G. (1991). Your face looks familiar but I can't remember your name: A review of dual-

process theory. In W.E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data:

Essays on human memory in honor of Bennet B. Murdock* (pp. 207-225). Hillsdale, NJ:

Erlbaum.

Mandler, G. (2008). Familiarity breeds attempts: A critical review of dual-process theories of

recognition. *Perspectives on Psychological Science, 3*, 390-400. doi:10.1111/j.1745-

6924.2008.00087.x

McClelland, J. L., & Chappell, M. (1998). Familiarity breeds differentiation : A subjective-

likelihood approach to the effects of experience in recognition memory. *Psychological*

Review, 105, 724-760. doi:10.1037/0033-295X.105.4.734-760

McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of

general and specific information. *Journal of Experimental Psychology: General, 114*(2),

159-188. doi:10.1037//0096-3445.114.2.159

McCrae, K., Cree, G. S., Seidenberg, M. S., & Mcnorgan, C. (2005). Semantic feature production

norms for a large set of living and nonliving things. *Behavior Research Methods, 37*, 547-

559. doi:10.3758/BF03192726

Meltzer, M. A., & Bartlett, J. C. (2019). Holistic processing and unitization in face recognition

memory. *Journal of Experimental Psychology: General, 148,* 1386-1406.

doi:10.1037/xge0000640

Metcalfe, J. (1993). Novelty monitoring, metacognition, and control in a composite holographic

associative recall model: Implication for Korsakoff Amnesia. *Psychological Review*,

100(1), 3-22. doi:10.1037/0033-295X.100.1.3

Metcalfe, J., Schwartz, B. L., & Joaquim, S. G. (1993). The cue-familiarity heuristic in

metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition,*

19, 851-861. doi:10.1037/0278-7393.19.4.851

Mickes, L., Wais, P. E., & Wixted, J. T. (2009). Recollection is a continuous process: Implications

for dual-process theories of recognition memory. *Psychological Science, 20*, 509-515.

doi:10.1111/j.1467-9280.2009.02324.x

Moscovitch, M., Winocur, G., & Behrmann, M. (1997). What is special about face recognition?

Nineteen experiments on a person with visual object agnosia and dyslexia but normal

face recognition. *Journal of Cognitive Neuroscience, 9*, 555-604.

doi:10.1162/jocn.1997.9.5.555

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition, 18,* 251-269.

doi:10.3758/bf03213879

Oliphant, T. E. (2006). *A guide to NumPy* (Vol. 1). Trelgol Publishing, USA.

Onyper, S. V., Zhang, Y. X., & Howard, M. (2010). Some-or-none recollection: Evidence from

item and source memory. *Journal of Experimental Psychology: General, 139*, 341-364.

doi:10.1037/a0018926

Peynircioğlu, Z. F. (1990). A feeling-of-recognition without identification. *Journal of Memory

and Language, 29*, 493-500. doi:10.1016/0749-596X(90)90068-B

Plaut, D. C. (1995). Semantic and associative priming in a distributed attractor network.

Proceedings of the 17th Annual Conference of the Cognitive Science Society (pp. 37-42).

Hillsdale, NJ: Lawrence Erlbaum Associates.

Poulin-Charronnat, B., Bigand, E., Lalitte, P., Madurell, F., Vieillard, S., & McAdams, S. (2004).

Effects of a change in instrumentation on the recognition of musical materials. *Music

Perception, 22*(2), 239-263. doi:10.1525/mp.2004.22.2.239

Reder, L. M. (1987). Strategy selection in question answering. *Cognitive Psychology, 19*(1), 90-

138. doi:10.1016/0010-0285(87)90005-3

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for

accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225-

237. doi:10.3758/PBR.16.2.225

Ryals, A. J., & Cleary, A. M. (2012). The recognition without cued recall phenomenon: Support

for a feature-matching theory over a partial recollection account. *Journal of Memory*

and Language, 66, 747-762. doi:10.1016/j.jml.2012.01.002

Ryals, A. J., Cleary, A. M., & Seger, C. A. (2013). Recall versus familiarity when recall fails for

words and scenes: The differential roles of the hippocampus, perirhinal cortex, and

category-specific cortical regions. *Brain Research, 1492,* 72-91.

doi:10.1016/j.brainres.2012.10.068

Seidenberg, M. S. (2007). Connectionist models of reading. In G. Gaskell (Ed.), *Oxford Handbook*

of Psycholinguistics. (pp. 235-250). Oxford University Press.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word

recognition and naming. *Psychological Review, 96*, 523-568. doi:10.1037/0033-

295x.96.4.52

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM – retrieving

effectively from memory. *Psychonomic Bulletin & Review, 4*(2), 145-166.

doi:10.3758/BF03209391

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A

feature model for semantic decisions. *Psychological Review, 81*(3), 214-241.

doi:10.1037/h0036351

Schwartz, B. L. (2002). *Tip-of-the-tongue states: Phenomenology, mechanism, and lexical retrieval.* Lawrence Erlbaum, New Jersey.

Schwartz, B. L., & Metcalfe, J. (1992). Cue familiarity but not target retrievability enhances feeling-of-knowing judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 1074-1083. doi:10.1037/0278-7393.18.5.1074

Tulving, E., & Bower, G. H. (1974). The logic of memory representations. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8). New York: Academic Press.

Tulving, E., & Watkins, M. J. (1975). Structure of memory traces. *Psychological Review, 82*(4), 261-275. doi:10.1037/h0076782

Underwood, B.J. (1969). Attributes of memory. *Psychological Review, 76*, 559-573. doi:10.1037/h0028143

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review*, *14*, 779-804. doi:10.3758/BF03194105

Wais, P. E., Mickes, L., & Wixted, J. T. (2008). Remember/know judgments probe degrees of recollection. *Journal of Cognitive Neuroscience, 20,* 400-405. doi:10.1162/jocn.2008.20041

Wickens, D. D. (1970). Encoding categories of words: An empirical approach to meaning. *Psychological Review, 77*(1), 1-15. doi:10.1037/h0028569

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*, 441-517. doi:10.1006/jmla.2002.2864

Footnotes

1. Using the full dataset, regardless of whether participants failed to identify the song from the isolated features, the same pattern emerged, such that the repeated-measures ANOVA performed on the probability of responding "yes, familiar" revealed a significant effect of feature familiarization, $F(2, 244) = 17.36$, $MSE = .008$, $p < .001$, $\eta_p^2 = .13$, $BF_{01} = 7.78 \times 10^{-6}$.

2. The same pattern of results emerged when using the full dataset of the probabilities of judging a song as familiar for songs containing no familiarized features or familiarized tonal features, $t(122) = 6.01$, $SE = .01$, $p < .001$, $d = .36$, $BF_{01} = 1.88 \times 10^{-6}$, 95% CrI [-.72, -.34].

3. When comparing the probabilities of saying "yes, familiar" for songs containing familiarized rhythms and songs containing no familiarized features using the full dataset, a similar pattern emerged, $t(122) = 3.51$, $SE = .01$, $p = .001$, $d = .21$, $BF_{01} = 0.03$, 95% CrI [-.49, -.13].

4. Note that when using the full dataset, rather than focusing on instances of song identification failure from the features, a significant difference did emerge. Participants were more likely to judge a song as familiar if it contained a familiarized rhythm than a familiarized tonal sequence, $t(122) = 2.32$, $SE = .01$, $p = .02$, $d = .15$, $BF_{01} = 0.76$, 95% CrI [-.38, -.03]. Note that encoded episodes in which the song was identified from the features likely had more features encoded than just the isolated features (e.g., the other non-presented features as well as the name).

5. The 2x4 repeated-measures ANOVA conducted on the full dataset produced similar results, such that a significant main effect of Yes-No response ($F(1, 113) = 76.29$, $MSE = 12.31$, $p < .001$, $\eta_p^2 = .40$, $BF_{01} = 1.73 \times 10^{-63}$) emerged. Although the main effect of Feature Familiarization condition was significant using NHST, Bayes Factors analysis suggested very strong evidence for the null ($F(3, 339) = 4.20$, $MSE = .91$, $p = .004$, $\eta_p^2 = .04$, $BF_{01} = 61.21$). Additionally, the interaction was not significant, $F < 1.0$, $BF_{01} = 63.08$.
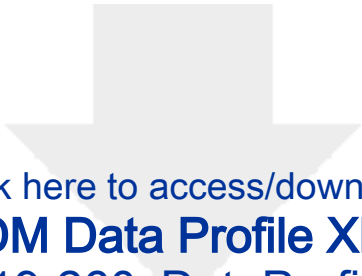
6. Upon analyzing the full dataset, regardless of whether participants failed to identify the songs from the isolated features at encoding, the repeated-measures ANOVA on the probability of saying "yes, familiar" as a function of feature type revealed a significant effect, $F(3, 357) = 16.38$, $MSE = .01$, $p < .001$, $\eta_p^2 = .12$, $BF_{01} = 6.21 \times 10^{-8}$.

7. The same pattern of results held up when using the full dataset. Songs containing no familiarized features were less frequently judged as familiar than test songs containing familiarized rhythms ($t(119) = -2.18$, $SE = .01$, $p = .03$, $d = .12$, $BF_{01} = 1.01$, 95% CrI [-.37, -.02]), familiarized pitch sequences ($t(119) = -3.82$, $SE = .01$, $p < .001$, $d = .25$, $BF_{01} = 0.01$, 95% CrI [-.53, -.16]), and familiarized tones *and* rhythms, $t(119) = -6.29$, $SE = .01$, $p < .001$, $d = .43$, $BF_{01} = 5.29 \times 10^{-7}$, 95% CrI [-.76, -.37].

8. This difference approached significance but did not reach it when using the full dataset, and the evidence was anecdotally in favor of the null, $t(119) = -1.84$, $SE = .01$, $p = .07$, $d = -.15$, $BF_{01} = 1.92$, 95% CrI [-.35, .01].

9. Similar patterns emerged when using the full dataset, such that songs containing familiarized rhythms ($t(119) = 4.68$, $SE = .01$, $p < .001$, $d = .33$, $BF_{01} = 5.36 \times 10^{-4}$, 95% CrI [.23, .61]) or familiarized pitch sequences ($t(119) = 2.96$, $SE = .01$, $p = .004$, $d = .19$, $BF_{01} = 0.16$, 95% CrI [.09, .44]) were less likely to be judged as familiar than songs containing both familiarized rhythms *and* pitch sequences.

10. When using the full dataset, a similar pattern emerged, such that the likelihood of saying "yes, familiar" did not differ between the *Rhythm+Pitch* familiarization probability and the sum of the *Rhythm-Only* and *Pitch-Only* probability, $t(119) = .82$, $SE = .02$, $p = .42$, $BF_{01} = 7.12$, 95% CrI [-.10, .25].

11. This one-sample *t*-test approached significance when using the full data set, but the evidence was anecdotally in favor of the null, $t(119) = 1.79$, $p = .08$, $BF_{01} = 2.09$, 95% CrI [-.02, .34].

12. A coding error in E-Prime resulted in two songs ("Do Wah Ditty" and "Love Me Do") not being studied in their designated exposure condition. Therefore, the trials that corresponded to these songs were excluded from all analyses.
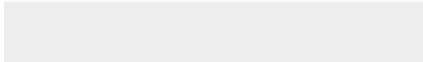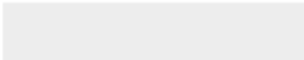
13. Upon analyzing the full dataset, regardless of whether song identification from the isolated feature sets succeeded or failed, a similar effect was found, such that increasing exposure to a song's isolated rhythm at study subsequently increased the probability of saying "yes, familiar" at test, $F(2, 242) = 12.72$, $MSE = .01$, $p < .001$, $BF_{01} = 3.97$ x 10^{-4}.

14. A similar pattern was found when using the full dataset, with participants being more likely to say "yes, familiar" if the test song clip had had its rhythm familiarized once as opposed to not at all during study, $t(121) = 2.43$, $SE = .01$, $p = .017$, $d = .17$, $BF_{01} = .59$, 95% CrI [.04, .39].

15. These effects were replicated when using the full dataset. Test song clips that had had their rhythms familiarized three times at study were significantly more likely to be judged as familiar compared to test song clips containing a rhythm familiarized only once at study ($t(121) = 2.53$, $SE = .01$, $p = .013$, $d = .20$, $BF_{01} = .48$, 95% CrI [.05, .40]) and also those that did not contain any familiarized rhythms, $t(121) = 5.23$, $SE = .01$, $p < .001$, $d = .39$, $BF_{01} = 5.70$ x 10^{-5}, 95% CrI [.27, .65].

16. When using the full dataset, there was no significant difference found, $t(121) = -.72$, $SE = .03$, $p = .47$, $BF_{01} = 7.71$, 95% CrI [-.24, .11]

17. This one-sample t-test produced similar results when using the full dataset, $t(121) = -.90$, $SE = .01$, $p = .37$, $BF_{01} = 6.70$, 95% CrI [-.25, .10].

Author Note

The data and stimuli for our experiments can be found at the following link: https://osf.io/nt3mv/

Do not remove this file (contains research data)

Click here to access/download
RDM Data Profile XML
JML-19-260_DataProfile.xml

Katherine L. McNeely-White: Conceptualization, Methodology, Software, Validation, Formal analysis, Writing - Original Draft, Writing - Review & Editing, Visualization, Simulations, Project administration **David G. McNeely-White**: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Resources, Writing – Review & Editing, Visualization **Anne M. Cleary**: Conceptualization, Methodology, Software, Validation, Formal analysis, Writing - Original Draft, Writing - Review & Editing, Visualization, Supervision, Project administration