Complexity of feature generalizations and learnability

Karthik Durvasula & Adam Liter
Michigan State University

May 8, 2016

Acknowledgements

- For thinking through the issues:
  - Bill Idsardi (UMD), Yen-Hwei Lin (MSU), Tal Linzen (LSCP & IJN), the MSU PhonoGroup, and the audiences at AMP 2015 and LSA 2016.
- For help with recording the stimuli:
  - Mina Hirzel (UMD).
- For help with equipment, participants, and lab facilities (MSU):
  - Curt Anderson, Joe Jalbert, Matt Kanefsky, Mike Kramizeh, Alan Munn, and Cristina Schmitt.

Questions

- How do learners generalize from data that is ambiguous between multiple different generalizations?

Questions

- How do learners generalize from data that is ambiguous between multiple different generalizations?
  - Single *vs.* multiple compatible generalizations?

Questions

- How do learners generalize from data that is ambiguous between multiple different generalizations?
  - Single *vs.* multiple compatible generalizations?
  - General *vs.* most precise generalization?

Questions

- How do learners generalize from data that is ambiguous between multiple different generalizations?
  - Single *vs.* multiple compatible generalizations?
  - General *vs.* most precise generalization?
- The above questions lead to some of the possibilities that have been argued for.

Outline

1. Introduction
   - Background
2. General Experiment Design
   - Training & Testing Phases
3. Experiments
   - Exp. 1
   - Exp. 2
   - Exp. 3
4. Phonotactic Learner (Hayes & Wilson 2008)
5. Conclusion & Future Work
   - Discussion

Background

Possible Viewpoints

- *Simplest Generalization* (**SG**; Chomsky and Halle 1968)

Background

Possible Viewpoints

- *Simplest Generalization* (**SG**; Chomsky and Halle 1968)
  - An evaluation metric that prioritizes the simplest generalization (fewest representational primitives to state the generalization).

Background

Possible Viewpoints

- *Simplest Generalization* (**SG**; Chomsky and Halle 1968)
  - An evaluation metric that prioritizes the simplest generalization (fewest representational primitives to state the generalization).
- *Multiple Simple(st) Generalizations* (**MSG**; Chomsky and Halle 1968; Hayes and Wilson 2008)

Background

Possible Viewpoints

- *Simplest Generalization* (**SG**; Chomsky and Halle 1968)
  - An evaluation metric that prioritizes the simplest generalization (fewest representational primitives to state the generalization).
- *Multiple Simple(st) Generalizations* (**MSG**; Chomsky and Halle 1968; Hayes and Wilson 2008)
  - The Chomsky and Halle view is actually unclear on what happens when there is more than one simplest generalization (defined in terms of number of features).

Background

Possible Viewpoints

- *Simplest Generalization* (**SG**; Chomsky and Halle 1968)
  - An evaluation metric that prioritizes the simplest generalization (fewest representational primitives to state the generalization).
- *Multiple Simple(st) Generalizations* (**MSG**; Chomsky and Halle 1968; Hayes and Wilson 2008)
  - The Chomsky and Halle view is actually unclear on what happens when there is more than one simplest generalization (defined in terms of number of features).
  - It is a reasonable extension of the view (in our opinion) to say that the learner keeps track of all the simple(st) generalizations.

Background

Possible Viewpoints

- *Simplest Generalization* (**SG**; Chomsky and Halle 1968)
  - An evaluation metric that prioritizes the simplest generalization (fewest representational primitives to state the generalization).
- *Multiple Simple(st) Generalizations* (**MSG**; Chomsky and Halle 1968; Hayes and Wilson 2008)
  - The Chomsky and Halle view is actually unclear on what happens when there is more than one simplest generalization (defined in terms of number of features).
  - It is a reasonable extension of the view (in our opinion) to say that the learner keeps track of all the simple(st) generalizations.
  - This is also the extension adopted in Hayes and Wilson (2008).

Background

Possible Viewpoints

- *Subset Principle* (**SP**; Berwick 1985)

[1] However, look at Hale and Reiss (2003) for a representational view.

Background

Possible Viewpoints

- *Subset Principle* (**SP**; Berwick 1985)
  - A more restrictive (*i.e.*, specific) grammar is entertained first, in the face of ambiguity.[1]

[1] However, look at Hale and Reiss (2003) for a representational view.

Background

Possible Viewpoints

- *Subset Principle* (**SP**; Berwick 1985)
  - A more restrictive (*i.e.*, specific) grammar is entertained first, in the face of ambiguity.[1]
  - Infants (age: ∼9 months) more easily learn patterns where there is a specific pattern (Gerken 2006).

[1] However, look at Hale and Reiss (2003) for a representational view.

Background

Possible Viewpoints

- *Subset Principle* (**SP**; Berwick 1985)
  - A more restrictive (*i.e.*, specific) grammar is entertained first, in the face of ambiguity.[1]
  - Infants (age: ∼9 months) more easily learn patterns where there is a specific pattern (Gerken 2006).
    - Looks at ABA and AAB syllable patterns (*e.g.*, *dewede*, *leledi*).

[1] However, look at Hale and Reiss (2003) for a representational view.

Background

Possible Viewpoints

- *Subset Principle* (**SP**; Berwick 1985)
  - A more restrictive (*i.e.*, specific) grammar is entertained first, in the face of ambiguity.[1]
  - Infants (age: ∼9 months) more easily learn patterns where there is a specific pattern (Gerken 2006).
    - Looks at ABA and AAB syllable patterns (*e.g.*, *dewede*, *leledi*).
- Learning proportional to specificity (**PropSpec**; Tenenbaum and Griffiths 2001)

[1] However, look at Hale and Reiss (2003) for a representational view.

Background

Possible Viewpoints

- *Subset Principle* (**SP**; Berwick 1985)
  - A more restrictive (*i.e.*, specific) grammar is entertained first, in the face of ambiguity.[1]
  - Infants (age: ∼9 months) more easily learn patterns where there is a specific pattern (Gerken 2006).
    - Looks at ABA and AAB syllable patterns (*e.g.*, *dewede*, *leledi*).
- Learning proportional to specificity (**PropSpec**; Tenenbaum and Griffiths 2001)
  - This has been termed the *Size Principle*.

[1] However, look at Hale and Reiss (2003) for a representational view.

Background

Possible Viewpoints

- *Subset Principle* (**SP**; Berwick 1985)
    - A more restrictive (*i.e.*, specific) grammar is entertained first, in the face of ambiguity.[1]
    - Infants (age: ∼9 months) more easily learn patterns where there is a specific pattern (Gerken 2006).
        - Looks at ABA and AAB syllable patterns (*e.g.*, *dewede*, *leledi*).
- Learning proportional to specificity (**PropSpec**; Tenenbaum and Griffiths 2001)
    - This has been termed the *Size Principle*.
    - The smaller the extension assigned by the generalization, the higher the probability assigned to it (Xu and Tenenbaum 2007).

[1] However, look at Hale and Reiss (2003) for a representational view.

Background

Possible Viewpoints

- *Subset Principle* (**SP**; Berwick 1985)
  - A more restrictive (*i.e.*, specific) grammar is entertained first, in the face of ambiguity.[1]
  - Infants (age: ∼9 months) more easily learn patterns where there is a specific pattern (Gerken 2006).
    - Looks at ABA and AAB syllable patterns (*e.g.*, *dewede*, *leledi*).
- Learning proportional to specificity (**PropSpec**; Tenenbaum and Griffiths 2001)
  - This has been termed the *Size Principle*.
  - The smaller the extension assigned by the generalization, the higher the probability assigned to it (Xu and Tenenbaum 2007).
  - So, learning should be proportional to the specificity of the hypothesis.

[1] However, look at Hale and Reiss (2003) for a representational view.

Background

Possible Viewpoints

- Learn all compatible generalizations, but simplest preferred (**PropSimple**; Linzen and Gallagher 2014; Linzen and O'Donnell 2015).

Background

Possible Viewpoints

- Learn all compatible generalizations, but simplest preferred (**PropSimple**; Linzen and Gallagher 2014; Linzen and O'Donnell 2015).
  - There are slight differences in the two papers.

Background

Possible Viewpoints

- Learn all compatible generalizations, but simplest preferred (**PropSimple**; Linzen and Gallagher 2014; Linzen and O'Donnell 2015).
  - There are slight differences in the two papers.
  - But, the main point is that there is an initial bias for general patterns, and the more specific pattern is learned more with increasing experience.

Background

Possible Viewpoints

- Learn all compatible generalizations, but simplest preferred (**PropSimple**; Linzen and Gallagher 2014; Linzen and O'Donnell 2015).
  - There are slight differences in the two papers.
  - But, the main point is that there is an initial bias for general patterns, and the more specific pattern is learned more with increasing experience.
  - This is implemented as a bias through Bayesian prior in Linzen and O'Donnell (2015).

Main claims in this talk

- For ambiguous input, learners:

Main claims in this talk

- For ambiguous input, learners:
  - do learn multiple generalizations.

Main claims in this talk

- For ambiguous input, learners:
  - do learn multiple generalizations.
  - do not seem to track generalizations that are featurally more specific (see Subset Principle).

General Experiment Design

- Three experiments in total.
- Each experiment had a *Training* and *Test* Phase.
- Each experiment lasted about 12–15 minutes.
- Participants were run in groups of 6–10.
- The stimuli were presented via PsychoPy (Peirce 2007, 2009).

General Experiment Design

Training Phase

- Participants listened to and silently mouthed 100 CVCV nonce words (2 repetitions each).

General Experiment Design

Training Phase

- Participants listened to and silently mouthed 100 CVCV nonce words (2 repetitions each).
- C=/p,b,t,d,f,v,s,z/ V=/a,i,u/.

General Experiment Design

Training Phase

- Participants listened to and silently mouthed 100 CVCV nonce words (2 repetitions each).
- C=/p,b,t,d,f,v,s,z/ V=/a,i,u/.
- C obeyed both voicing and continuancy harmony simultaneously.

General Experiment Design

Training Phase

- Participants listened to and silently mouthed 100 CVCV nonce words (2 repetitions each).
- C=/p,b,t,d,f,v,s,z/ V=/a,i,u/.
- C obeyed both voicing and continuancy harmony simultaneously.
  - e.g., ✓[tipa, bida, fisa], *[tisa, bipa, fida].

General Experiment Design

Possible Generalizations

Training stimuli: [tapi], [sifa], [sasi] ...

General Experiment Design

Possible Generalizations

Training stimuli: [tapi], [sifa], [sasi] ...

[α voice]

General Experiment Design

Possible Generalizations

Training stimuli: [tapi], [sifa], [sasi] . . .

[α voice]   [α cont]

General Experiment Design

Possible Generalizations

Training stimuli: [tapi], [sifa], [sasi] ...

[α voice]

[α cont]

[α voice, α cont]

General Experiment Design

Testing Phase

- Participants were asked if word was possible in the “language” they learned.

General Experiment Design

Testing Phase

- Participants were asked if word was possible in the "language" they learned.
- CVCV nonce words of the following types:

General Experiment Design

Testing Phase

- Participants were asked if word was possible in the "language" they learned.
- CVCV nonce words of the following types:
  - 12 OldStims (*e.g.*, [tipa, bida, fisa])

General Experiment Design

Testing Phase

- Participants were asked if word was possible in the "language" they learned.
- CVCV nonce words of the following types:
  - 12 OldStims (*e.g.*, [tipa, bida, fisa])
  - 12 NewStims (*e.g.*, [tupi, buda, safi])

General Experiment Design

Testing Phase

- Participants were asked if word was possible in the "language" they learned.
- CVCV nonce words of the following types:
  - 12 OldStims (*e.g.*, [tipa, bida, fisa])
  - 12 NewStims (*e.g.*, [tupi, buda, safi])
  - 12 OnlyVoicing (*e.g.*, [tusi, bazi, vabi])

General Experiment Design

Testing Phase

- Participants were asked if word was possible in the "language" they learned.
- CVCV nonce words of the following types:
  - 12 OldStims (*e.g.*, [tipa, bida, fisa])
  - 12 NewStims (*e.g.*, [tupi, buda, safi])
  - 12 OnlyVoicing (*e.g.*, [tusi, bazi, vabi])
  - 12 OnlyContinuancy (*e.g.*, [tadi, zafu, bupi])

General Experiment Design

Testing Phase

- Participants were asked if word was possible in the "language" they learned.
- CVCV nonce words of the following types:
  - 12 OldStims (*e.g.*, [tipa, bida, fisa])
  - 12 NewStims (*e.g.*, [tupi, buda, safi])
  - 12 OnlyVoicing (*e.g.*, [tusi, bazi, vabi])
  - 12 OnlyContinuancy (*e.g.*, [tadi, zafu, bupi])
  - 12 Disharmony (*e.g.*, [tuvi, zipa, fidu])

Experiment 1

- 25 English-speaking undergraduates
  - 3 were excluded due to non-learning.

Experiment 1

- 25 English-speaking undergraduates
  - 3 were excluded due to non-learning.
- NewStims: C sequences possibly heard in *Training*.

Experiment 1

Predictions

- **SP**: NewStims and OldStims preferred over other three (which are undifferentiated).

Experiment 1

Predictions

- **SP**: NewStims and OldStims preferred over other three (which are undifferentiated).
- **SG**: Some prefer [α voice], some [α cont]; NewStims are as acceptable as either. Thus, all three are equally good.

Experiment 1

Predictions

- **SP**: NewStims and OldStims preferred over other three (which are undifferentiated).
- **SG**: Some prefer [α voice], some [α cont]; NewStims are as acceptable as either. Thus, all three are equally good.
- **MSG**: Both [α voice] and [α cont] are preferred; additive effect on NewStims.

Experiment 1

Predictions

- **SP**: NewStims and OldStims preferred over other three (which are undifferentiated).
- **SG**: Some prefer [α voice], some [α cont]; NewStims are as acceptable as either. Thus, all three are equally good.
- **MSG**: Both [α voice] and [α cont] are preferred; additive effect on NewStims.
- **PropSimple**: [α voice], [α cont], and [α voice, α cont] are all learned; therefore, interactive effect on NewStims. However, interaction effect smaller than either [α voice] or [α cont].

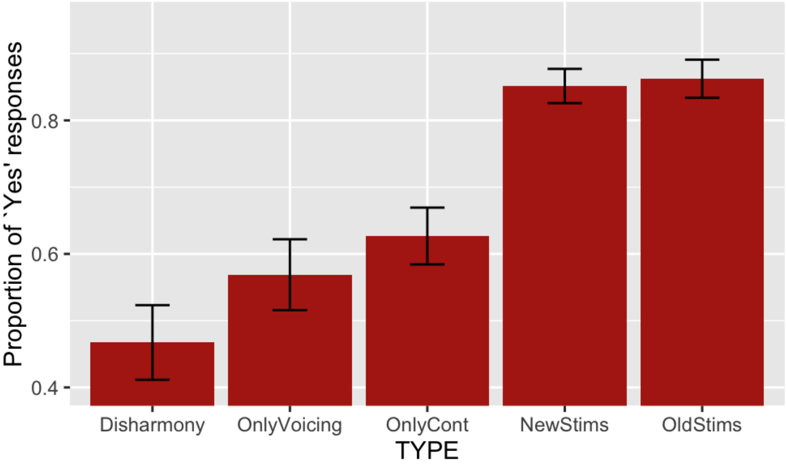
Experiment 1

Predictions

- **SP**: NewStims and OldStims preferred over other three (which are undifferentiated).
- **SG**: Some prefer [α voice], some [α cont]; NewStims are as acceptable as either. Thus, all three are equally good.
- **MSG**: Both [α voice] and [α cont] are preferred; additive effect on NewStims.
- **PropSimple**: [α voice], [α cont], and [α voice, α cont] are all learned; therefore, interactive effect on NewStims. However, interaction effect smaller than either [α voice] or [α cont].
- **PropSpec**: [α voice], [α cont], and [α voice, α cont] are all learned; therefore, interactive effect on NewStims. However, interaction effect larger than either [α voice] or [α cont].

Experiment 1

Results



Experiment 1

Results

Fixed Effect	MeanYes (%)	Estimate	z-value	Pr(>z)
(Intercept)	0.4674	−0.1223	−0.534	0.2968
OnlyVoicing	0.5688	0.4801	2.485	0.0065 **
OnlyCont	0.6268	0.7664	3.897	<0.0001 ***
NewStims	0.8514	2.142	9.05	<0.0001 ***
OldStims	0.8623	2.2292	9.331	<0.0001 ***

Table: Logistic mixed-effects models

Experiment 1

Results

Fixed Effect	Estimate	z-value	Pr(>z)
(Intercept)	-0.1231	-0.544	0.2934
Voicing	0.4758	2.513	0.0059 **
Continuancy	0.7574	3.920	<0.0001 ***
Voicing:Continuancy	0.8881	3.032	0.0012 **

Table: Logistic mixed-effects model—Interaction effects for new test stimuli

Experiment 1

Discussion

- Multiple simple generalizations are learned for ambiguous data.

Experiment 1

Discussion

- Multiple simple generalizations are learned for ambiguous data.
- Furthermore, interaction effect suggests potential support for **PropSimple** and **PropSpec**.

Experiment 1

Discussion

- Multiple simple generalizations are learned for ambiguous data.
- Furthermore, interaction effect suggests potential support for **PropSimple** and **PropSpec**.
- However:
    - Perhaps phonological generalizations can also directly access segmental representations (*i.e.*, segmental primitives) without making reference to the featural content.

Experiment 1

Discussion

- Multiple simple generalizations are learned for ambiguous data.
- Furthermore, interaction effect suggests potential support for **PropSimple** and **PropSpec**.
- However:
  - Perhaps phonological generalizations can also directly access segmental representations (*i.e.*, segmental primitives) without making reference to the featural content.
  - If so, a generalization based on a single segment might be as "simple" as a generalization based on a single feature.

Experiment 1

Discussion

- Multiple simple generalizations are learned for ambiguous data.
- Furthermore, interaction effect suggests potential support for **PropSimple** and **PropSpec**.
- However:
  - Perhaps phonological generalizations can also directly access segmental representations (*i.e.*, segmental primitives) without making reference to the featural content.
  - If so, a generalization based on a single segment might be as "simple" as a generalization based on a single feature.
  - Therefore, **MSG** could also account for the interaction effect.

Experiment 2

- 78 English-speaking undergraduates, and 15 were excluded due to non-learning.

Experiment 2

- 78 English-speaking undergraduates, and 15 were excluded due to non-learning.
- Training stimuli (*e.g.*, [tipa, bida, fisa])
  - One pair of consonant sequences withheld (*e.g.*, [vz, zv])

Experiment 2

- 78 English-speaking undergraduates, and 15 were excluded due to non-learning.
- Training stimuli (*e.g.*, [tipa, bida, fisa])
  - One pair of consonant sequences withheld (*e.g.*, [vz, zv])
  - Identical sequences did occur in training (*e.g.*, [viva, zuzi])

Experiment 2

- 78 English-speaking undergraduates, and 15 were excluded due to non-learning.
- Training stimuli (*e.g.*, [tipa, bida, fisa])
  - One pair of consonant sequences withheld (*e.g.*, [vz, zv])
  - Identical sequences did occur in training (*e.g.*, [viva, zuzi])
  - The consonant pairs that were withheld were randomized per participant.

Experiment 2

- 78 English-speaking undergraduates, and 15 were excluded due to non-learning.
- Training stimuli (*e.g.*, [tipa, bida, fisa])
  - One pair of consonant sequences withheld (*e.g.*, [vz, zv])
  - Identical sequences did occur in training (*e.g.*, [viva, zuzi])
  - The consonant pairs that were withheld were randomized per participant.
- Testing stimuli:
  - Same as Exp. 1: OldStims, OnlyVoicing, OnlyContinuancy, Disharmony

Experiment 2

- 78 English-speaking undergraduates, and 15 were excluded due to non-learning.
- Training stimuli (*e.g.*, [tipa, bida, fisa])
  - One pair of consonant sequences withheld (*e.g.*, [vz, zv])
  - Identical sequences did occur in training (*e.g.*, [viva, zuzi])
  - The consonant pairs that were withheld were randomized per participant.
- Testing stimuli:
  - Same as Exp. 1: OldStims, OnlyVoicing, OnlyContinuancy, Disharmony
  - NewStims had withheld C sequences (*e.g.*, [vuzi, zuva, zavi])

Experiment 2

- 78 English-speaking undergraduates, and 15 were excluded due to non-learning.
- Training stimuli (*e.g.*, [tipa, bida, fisa])
  - One pair of consonant sequences withheld (*e.g.*, [vz, zv])
  - Identical sequences did occur in training (*e.g.*, [viva, zuzi])
  - The consonant pairs that were withheld were randomized per participant.
- Testing stimuli:
  - Same as Exp. 1: OldStims, OnlyVoicing, OnlyContinuancy, Disharmony
  - NewStims had withheld C sequences (*e.g.*, [vuzi, zuva, zavi])
- Therefore, segmental generalizations will not help with the harmony patterns.

Experiment 2

Prediction

- Similar to Exp. 1, but:
    - All three (**MSG**, **PropSimple**, and **PropSpec**) predict a drop in NewStims preference.

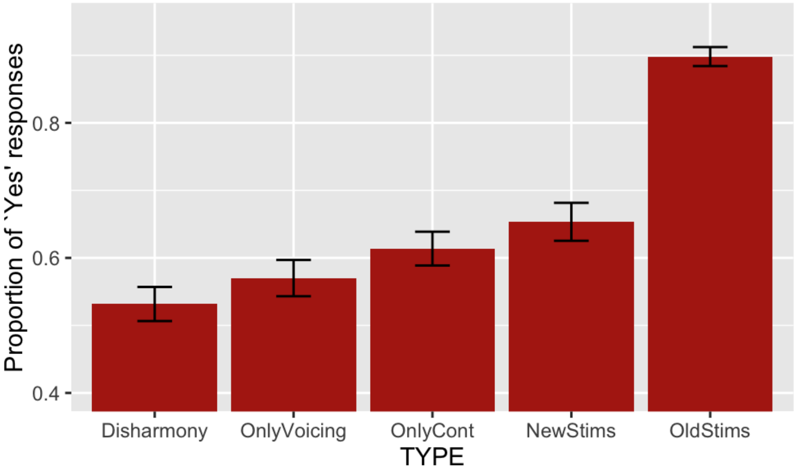
Experiment 2

Prediction

- Similar to Exp. 1, but:
  - All three (**MSG**, **PropSimple**, and **PropSpec**) predict a drop in NewStims preference.
  - However, **PropSimple** and **PropSpec** still predict an interactive effect.

Experiment 2

Results



Experiment 2

Results

Fixed Effect	MeanYes (%)	Estimate	z-value	Pr(>z)
(Intercept)	0.5317	0.1628	1.273	0.1014
OnlyVoicing	0.5701	0.1735	1.387	0.0825 .
OnlyCont	0.6138	0.3889	3.087	0.0010 **
NewStims	0.6534	0.5836	4.433	$<$0.0001 ***
OldStims	0.8981	2.2820	14.274	$<$0.0001 ***

Table: Logistic mixed-effects models

Experiment 2

Results: Was there an interactive effect on NewStims?

- Best model was the one with two main effects for *Voicing* and *Continuancy*.

Experiment 2

Results: Was there an interactive effect on NewStims?

- Best model was the one with two main effects for *Voicing* and *Continuancy*.

Fixed Effect	Estimate	z-value	Pr(>z)
(Intercept)	−0.1841	−1.223	=0.111
Voicing	0.4635	3.100	<0.01 **
Continuancy	0.4972	3.331	<0.001 ***

Table: Logistic mixed-effects models

Experiment 2

Results: Was there an interactive effect on NewStims?

- Best model was the one with two main effects for *Voicing* and *Continuancy*.

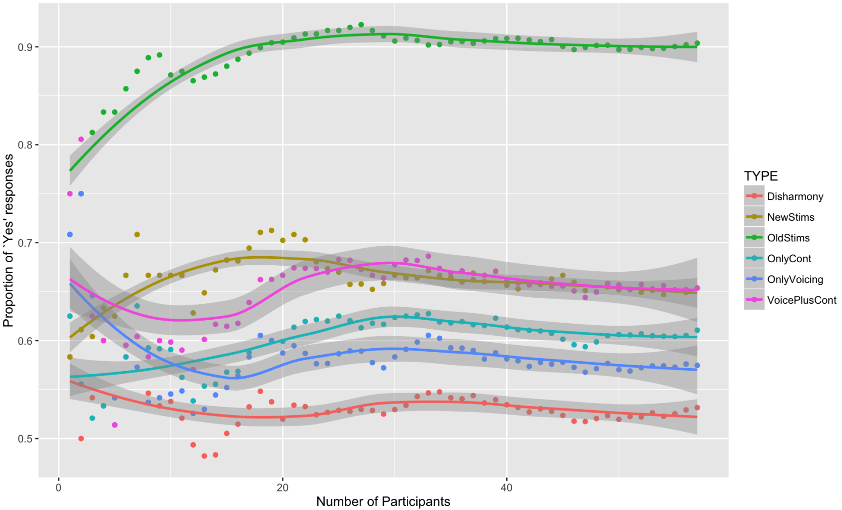
Fixed Effect	Estimate	z-value	Pr(>z)
(Intercept)	−0.1841	−1.223	=0.111
Voicing	0.4635	3.100	<0.01 **
Continuancy	0.4972	3.331	<0.001 ***

Table: Logistic mixed-effects models

- There was no interaction effect.

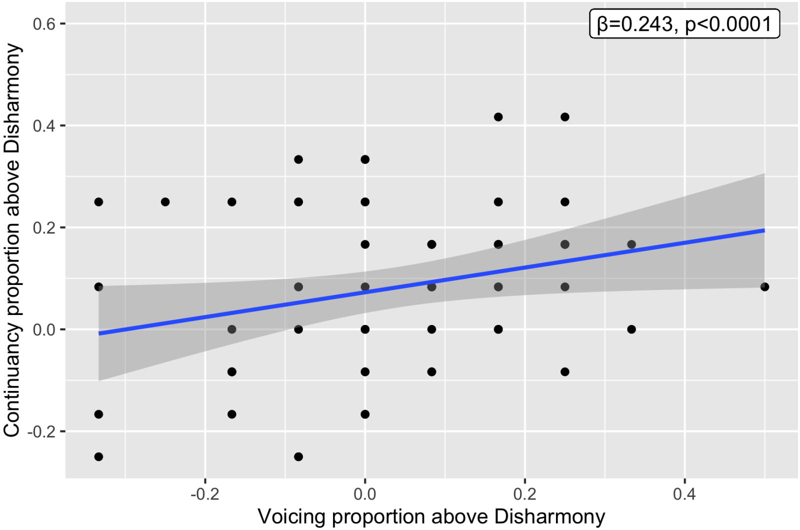
Experiment 2

Results



Experiment 2

Results: Correlation between the two one-feature generalizations



Experiment 2
Discussion

- Noticeable drop in preference for NewStims.

Experiment 2

Discussion

- Noticeable drop in preference for NewStims.
- No evidence of interaction effect.

Experiment 2
Discussion

- Noticeable drop in preference for NewStims.
- No evidence of interaction effect.
- The results are only consistent with participants keeping track of multiple simple generalizations (**MSG**).

Experiment 3

- Putting **MSG** to a further test:
  - Are the NewStims in Exp. 2 really worse than those in Exp. 1?

Experiment 3

- Putting **MSG** to a further test:
  - Are the NewStims in Exp. 2 really worse than those in Exp. 1?
  - Can we replicate the results?

Experiment 3

- Putting **MSG** to a further test:
  - Are the NewStims in Exp. 2 really worse than those in Exp. 1?
  - Can we replicate the results?
- 51 English-speaking undergraduates.

Experiment 3

- Putting **MSG** to a further test:
  - Are the NewStims in Exp. 2 really worse than those in Exp. 1?
  - Can we replicate the results?
- 51 English-speaking undergraduates.
- Both types of NewStims (from Exp. 1 & 2):

Experiment 3

- Putting **MSG** to a further test:
  - Are the NewStims in Exp. 2 really worse than those in Exp. 1?
  - Can we replicate the results?
- 51 English-speaking undergraduates.
- Both types of NewStims (from Exp. 1 & 2):
  - NewWordStims: C sequences heard in *Training*.

Experiment 3

- Putting **MSG** to a further test:
  - Are the NewStims in Exp. 2 really worse than those in Exp. 1?
  - Can we replicate the results?
- 51 English-speaking undergraduates.
- Both types of NewStims (from Exp. 1 & 2):
  - NewWordStims: C sequences heard in *Training*.
    - Therefore, segmental generalizations will help with the harmony patterns.

Experiment 3

- Putting **MSG** to a further test:
  - Are the NewStims in Exp. 2 really worse than those in Exp. 1?
  - Can we replicate the results?
- 51 English-speaking undergraduates.
- Both types of NewStims (from Exp. 1 & 2):
  - NewWordStims: C sequences heard in *Training*.
    - Therefore, segmental generalizations will help with the harmony patterns.
  - NewConsStims: C sequences **not** heard in *Training*.

Experiment 3

- Putting **MSG** to a further test:
  - Are the NewStims in Exp. 2 really worse than those in Exp. 1?
  - Can we replicate the results?
- 51 English-speaking undergraduates.
- Both types of NewStims (from Exp. 1 & 2):
  - NewWordStims: C sequences heard in *Training*.
    - Therefore, segmental generalizations will help with the harmony patterns.
  - NewConsStims: C sequences **not** heard in *Training*.
    - Therefore, segmental generalizations will not help with the harmony patterns.

Experiment 3

- Putting **MSG** to a further test:
  - Are the NewStims in Exp. 2 really worse than those in Exp. 1?
  - Can we replicate the results?
- 51 English-speaking undergraduates.
- Both types of NewStims (from Exp. 1 & 2):
  - NewWordStims: C sequences heard in *Training*.
    - Therefore, segmental generalizations will help with the harmony patterns.
  - NewConsStims: C sequences **not** heard in *Training*.
    - Therefore, segmental generalizations will not help with the harmony patterns.
  - Still 60 test items.

Experiment 3

- Putting **MSG** to a further test:
  - Are the NewStims in Exp. 2 really worse than those in Exp. 1?
  - Can we replicate the results?
- 51 English-speaking undergraduates.
- Both types of NewStims (from Exp. 1 & 2):
  - NewWordStims: C sequences heard in *Training*.
    - Therefore, segmental generalizations will help with the harmony patterns.
  - NewConsStims: C sequences **not** heard in *Training*.
    - Therefore, segmental generalizations will not help with the harmony patterns.
  - Still 60 test items.
    - 10 words for each type (6 types in all).

Experiment 3
Prediction

- **MSG**'s predictions for OnlyVoicing and OnlyContinuancy remain the same.

Experiment 3

Prediction

- **MSG**'s predictions for OnlyVoicing and OnlyContinuancy remain the same.
- **MSG** predicts only an additive effect for NewConsStims

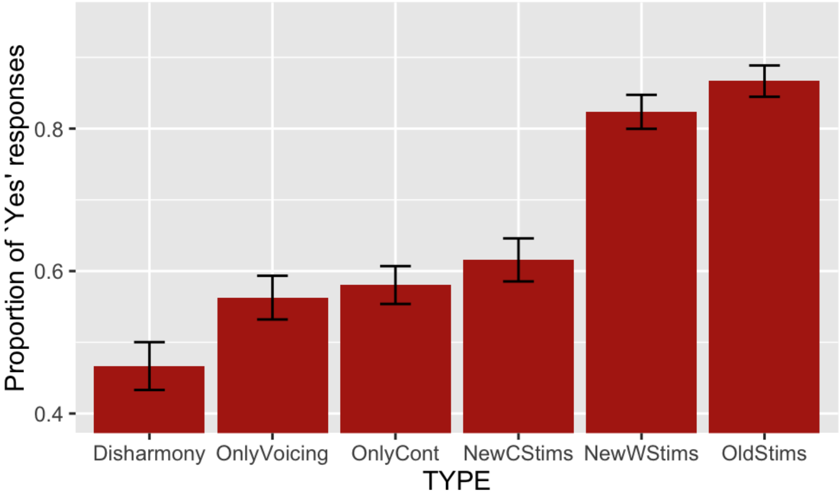
Experiment 3

Prediction

- **MSG**'s predictions for OnlyVoicing and OnlyContinuancy remain the same.
- **MSG** predicts only an additive effect for NewConsStims
- **MSG** predicts an interactive (super-additive) effect for NewWordStims

Experiment 3

Results



Experiment 3

Results

Fixed Effect	MeanYes (%)	Estimate	z-value	Pr(>z)
(Intercept)	0.4667	−0.1354	−0.987	=0.324
OnlyVoicing	0.5627	0.4386	3.122	=0.001 ***
OnlyCont	0.5804	0.5118	3.640	<0.0001 ***
NewConsStims	0.6157	0.6681	4.648	<0.0001 ***
NewWordStims	0.8235	1.8386	11.475	<0.0001 ***
OldStims	0.8667	2.2030	12.771	<0.0001 ***

Table: Logistic mixed-effects models

Experiment 3

Results: Was there an interactive effect on NewConsStims?

- Best model was the one with two main effects for *Voicing* and *Continuancy*.

Experiment 3

Results: Was there an interactive effect on NewConsStims?

- Best model was the one with two main effects for *Voicing* and *Continuancy*.

Fixed Effect	Estimate	z-value	Pr(>z)
(Intercept)	−0.0689	−0.538	=0.295
Voicing	0.3022	2.948	<0.01 **
Continuancy	0.3737	3.646	<0.001 ***

Table: Logistic mixed-effects models

Experiment 3

Results: Was there an interactive effect on NewConsStims?

- Best model was the one with two main effects for *Voicing* and *Continuancy*.

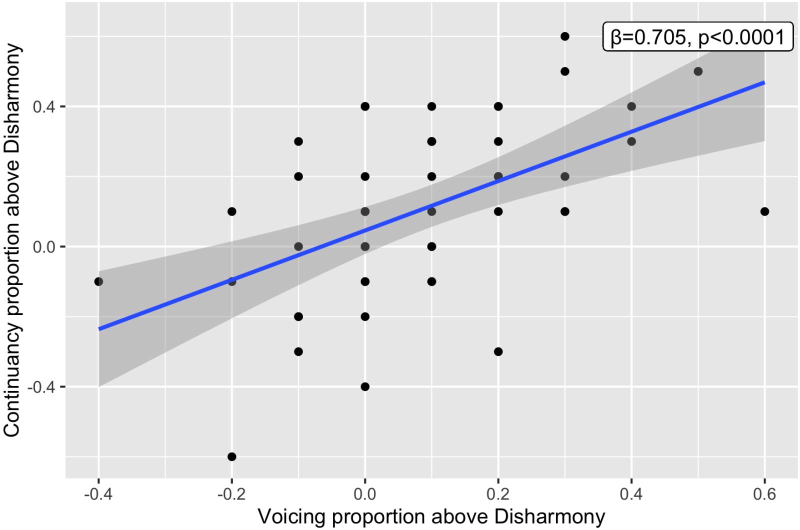
Fixed Effect	Estimate	z-value	Pr(>z)
(Intercept)	−0.0689	−0.538	=0.295
Voicing	0.3022	2.948	<0.01 **
Continuancy	0.3737	3.646	<0.001 ***

Table: Logistic mixed-effects models

- There was no interaction effect.

Experiment 3

Results: Correlation between the two one-feature generalizations



Experiment 3

Discussion

- We replicated the results of Exp. 1 & 2.

Experiment 3

Discussion

- We replicated the results of Exp. 1 & 2.
- The OnlyVoicing and OnlyContinuancy stims were clearly more preferred than Disharmony.

Experiment 3

Discussion

- We replicated the results of Exp. 1 & 2.
- The OnlyVoicing and OnlyContinuancy stims were clearly more preferred than Disharmony.
- There was also clear evidence that both simple generalizations were being learned.

Experiment 3

Discussion

- We replicated the results of Exp. 1 & 2.
- The OnlyVoicing and OnlyContinuancy stims were clearly more preferred than Disharmony.
- There was also clear evidence that both simple generalizations were being learned.
- When the confound of experience with consonantal sequences was controlled for, there was no interactive effect for the NewConsStims.
  - As predicted by **MSG**.

Hayes & Wilson Phonotactic Learner

Using the learner to make predictions for Experiment 3

- This learning model can be seen as an inductive baseline.

Hayes & Wilson Phonotactic Learner

Using the learner to make predictions for Experiment 3

- This learning model can be seen as an inductive baseline.
  - It instantiates the Chomsky & Halle (1968) intuition of learning maximally general patterns/constraints (among constraints that are approximately equal in accuracy).

Hayes & Wilson Phonotactic Learner

Using the learner to make predictions for Experiment 3

- This learning model can be seen as an inductive baseline.
  - It instantiates the Chomsky & Halle (1968) intuition of learning maximally general patterns/constraints (among constraints that are approximately equal in accuracy).
  - Note we do not suggest that this is the actual learner, just that it employs some of the relevant biases.

Hayes & Wilson Phonotactic Learner

Using the learner to make predictions for Experiment 3

- This learning model can be seen as an inductive baseline.
  - It instantiates the Chomsky & Halle (1968) intuition of learning maximally general patterns/constraints (among constraints that are approximately equal in accuracy).
  - Note we do not suggest that this is the actual learner, just that it employs some of the relevant biases.
- We used this learner and

Hayes & Wilson Phonotactic Learner

Using the learner to make predictions for Experiment 3

- This learning model can be seen as an inductive baseline.
  - It instantiates the Chomsky & Halle (1968) intuition of learning maximally general patterns/constraints (among constraints that are approximately equal in accuracy).
  - Note we do not suggest that this is the actual learner, just that it employs some of the relevant biases.
- We used this learner and
  - Trained the learner separately for each participant's *Training* stimuli.

Hayes & Wilson Phonotactic Learner

Using the learner to make predictions for Experiment 3

- This learning model can be seen as an inductive baseline.
  - It instantiates the Chomsky & Halle (1968) intuition of learning maximally general patterns/constraints (among constraints that are approximately equal in accuracy).
  - Note we do not suggest that this is the actual learner, just that it employs some of the relevant biases.
- We used this learner and
  - Trained the learner separately for each participant's *Training* stimuli.
  - Then, used each trained grammar to predict the maxent scores for the corresponding participant's *Test* stimuli.

Hayes & Wilson Phonotactic Learner

Results on Exp. 3 Training & Test Items

- These are the constraints that the learner learns consistently for each participant's training data:

Hayes & Wilson Phonotactic Learner

Results on Exp. 3 Training & Test Items

- These are the constraints that the learner learns consistently for each participant's training data:
  - *[+continuant][-continuant]
  - *[-continuant][+continuant]
  - *[-voice][+voice]
  - *[+voice][-voice]

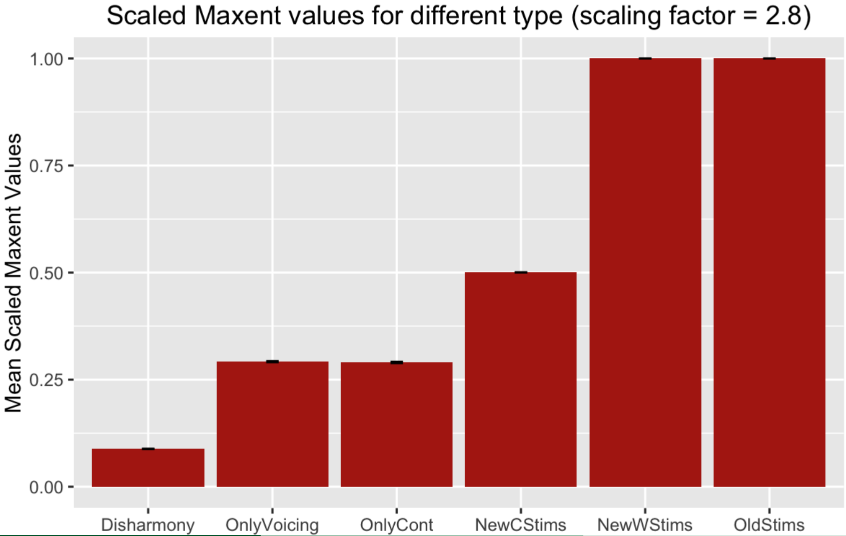
Hayes & Wilson Phonotactic Learner

Results on Exp. 3 Training & Test Items

- These are the constraints that the learner learns consistently for each participant's training data:
  - *[+continuant][-continuant]
  - *[-continuant][+continuant]
  - *[-voice][+voice]
  - *[+voice][-voice]
- It doesn't learn the complex/more specific stop-voicing constraints.

Hayes & Wilson Phonotactic Learner

Results on Exp. 3 Training & Test Items



Hayes & Wilson Phonotactic Learner

Discussion

- The Hayes & Wilson Phonotactic Learner with its bias for general patterns accurately captures the non-interactive effect observed in Exp. 3.

Hayes & Wilson Phonotactic Learner

Discussion

- The Hayes & Wilson Phonotactic Learner with its bias for general patterns accurately captures the non-interactive effect observed in Exp. 3.
- Quite interestingly, the modeling also shows that there is no need to the learner to keep track of segmental patterns per se, to account for the NewWordStims.

Hayes & Wilson Phonotactic Learner

Discussion

- The Hayes & Wilson Phonotactic Learner with its bias for general patterns accurately captures the non-interactive effect observed in Exp. 3.
- Quite interestingly, the modeling also shows that there is no need to the learner to keep track of segmental patterns per se, to account for the NewWordStims.
    - Remember, the Phonotactic Learner keeps track of feature-based phonotactics.

Hayes & Wilson Phonotactic Learner

Discussion

- The Hayes & Wilson Phonotactic Learner with its bias for general patterns accurately captures the non-interactive effect observed in Exp. 3.
- Quite interestingly, the modeling also shows that there is no need to the learner to keep track of segmental patterns per se, to account for the NewWordStims.
  - Remember, the Phonotactic Learner keeps track of feature-based phonotactics.
  - Complex featural generalizations are enough to account for the patterns.

Overall Discussion

- We showed today that, in general, participants' responses to more specific patterns are at most an addictive effect of more general patterns.

Overall Discussion

- We showed today that, in general, participants' responses to more specific patterns are at most an addictive effect of more general patterns.
  - Experiment 1 suggested that there might be an interactive effect.

Overall Discussion

- We showed today that, in general, participants' responses to more specific patterns are at most an addictive effect of more general patterns.
  - Experiment 1 suggested that there might be an interactive effect.
  - However, as Experiments 2 & 3 showed, this interactive effect could be because of experience with relevant segments.

Overall Discussion

- We showed today that, in general, participants' responses to more specific patterns are at most an addictive effect of more general patterns.
  - Experiment 1 suggested that there might be an interactive effect.
  - However, as Experiments 2 & 3 showed, this interactive effect could be because of experience with relevant segments.
  - Once the confound was removed, there was no more interactive effect.

Overall Discussion

- We showed today that, in general, participants' responses to more specific patterns are at most an addictive effect of more general patterns.
  - Experiment 1 suggested that there might be an interactive effect.
  - However, as Experiments 2 & 3 showed, this interactive effect could be because of experience with relevant segments.
  - Once the confound was removed, there was no more interactive effect.
- Our claim is not that the learner doesn't entertain complex generalizations at all.

Overall Discussion

- We showed today that, in general, participants' responses to more specific patterns are at most an addictive effect of more general patterns.
  - Experiment 1 suggested that there might be an interactive effect.
  - However, as Experiments 2 & 3 showed, this interactive effect could be because of experience with relevant segments.
  - Once the confound was removed, there was no more interactive effect.
- Our claim is not that the learner doesn't entertain complex generalizations at all.
- Instead, it is that when there is ambiguity, the learner entertains the more general hypotheses first.

Overall Discussion

- We showed today that, in general, participants' responses to more specific patterns are at most an addictive effect of more general patterns.
  - Experiment 1 suggested that there might be an interactive effect.
  - However, as Experiments 2 & 3 showed, this interactive effect could be because of experience with relevant segments.
  - Once the confound was removed, there was no more interactive effect.
- Our claim is not that the learner doesn't entertain complex generalizations at all.
- Instead, it is that when there is ambiguity, the learner entertains the more general hypotheses first.
- This suggests that the search space is constrained by an evaluation metric of the kind discussed in (Chomsky and Halle 1968).

Overall Discussion

- Furthermore, we also showed using the Hayes & Wilson model

Overall Discussion

- Furthermore, we also showed using the Hayes & Wilson model
  - Lack of interactive effect is expected if the learning model has a preference for more general patterns.

Overall Discussion

- Furthermore, we also showed using the Hayes & Wilson model
  - Lack of interactive effect is expected if the learning model has a preference for more general patterns.
  - The reason there was an interactive effect for the NewWordStims (stims with old consonant sequences, but with new vowel combinations) has to do not with the learner keeping track of segmental patterns, but with the learner tracking other patterns purely in terms of more complex featural patterns.

Conclusion & Future Work

- Bayesian models make an interesting claim.
  - As the amount of experience increases, general patterns incur a penalty compared to more specific patterns.
  - This is due to the likelihood term that is in models.
  - If so, the preference for the general should decrease with increasing experience.

Conclusion & Future Work

- Bayesian models make an interesting claim.
  - As the amount of experience increases, general patterns incur a penalty compared to more specific patterns.
  - This is due to the likelihood term that is in models.
  - If so, the preference for the general should decrease with increasing experience.
- We are hoping to test this out with a series of experiments looking at the effect of experience on the magnitude of the effect with general patterns.
  - So, see how participants respond with increasing amounts of data.

References I

Berwick, Robert C. (1985). *The Acquisition of Syntactic Knowledge.* MIT Press.

Chomsky, Noam and Morris Halle (1968). *The Sound Pattern of English*. Harper and Row.

Gerken, LouAnn (2006). "Decisions, decisions: infant language learning when multiple generalizations are possible". *Cognition* 98, B67–B74.

Hale, Mark and Charles Reiss (2003). "The Subset Principle in phonology: why the tabula can't be rasa". *Journal of Linguistics* 39.02, pp. 219–244.

Hayes, Bruce and Colin Wilson (2008). "A maximum entropy model of phonotactics and phonotactic learning". *Linguistic Inquiry* 39, pp. 379–440.

Linzen, Tal and Gillian Gallagher (2014). "The Timecourse of Generalization in Phonotactic Learning". In: *Proceedings of the 2013 Meeting on Phonology*. Ed. by John Kingston et al. Washington D.C.: Linguistic Society of America, pp. 1–12. doi: 10.3765/amp.v1i1.18.

Linzen, Tal and Timothy J. O'Donnell (2015). "A Model of Rapid Phonotactic Generalization". In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Lisbon, Portugal: Association for Computational Linguistics, pp. 1126–1131.

References II

Peirce, Jonathan W. (2007). "PsychoPy—Psychophysics software in Python". *Journal of Neuroscience Methods* 162.1–2, pp. 8–13. doi: 10.1016/j.jneumeth.2006.11.017.

Peirce, Jonathan W. (2009). "Generating stimuli for neuroscience using PsychoPy". *Frontiers in Neuroinformatics* 2.10. doi: 10.3389/neuro.11.010.2008.

Tenenbaum, Joshua B. and T. L. Griffiths (2001). "Generalization, similarity, and Bayesian inference." *Behavioral and Brain Sciences* 24, pp. 629–640.

Xu, Fei and Joshua B. Tenenbaum (2007). "Word Learning as Bayesian Inference." *Psychological Review* 114 (2), pp. 245–272. doi: 10.1037/0033-295X.114.2.245.