Semantics and pragmatics of numerical approximation expressions

Adèle Mortier

Supervisors:
Paul Égré | Benjamin Spector

June 25, 2019

Introduction: why use *around*?

Numerical approximation?

Approximation expressions: *around n*, *almost n*, *between n and m*...

Introduction: why use *around*?

Numerical approximation?

Approximation expressions: *around n*, *almost n*, *between n and m*...

1. She is between 28 and 32 y.o.
2. She is around 30 y.o.

Introduction: why use *around*?

Numerical approximation?

Approximation expressions: *around n*, *almost n*, *between n and m*...

1. She is between 28 and 32 y.o.
2. She is around 30 y.o.

Around vs between

- 1 provides clear boundaries, and has a disjunctive force (**imprecision**). The speaker isn't sure about the age.

Introduction: why use *around*?

Numerical approximation?

Approximation expressions: *around n*, *almost n*, *between n and m*...

1. She is between 28 and 32 y.o.
2. She is around 30 y.o.

Around vs between

- 1 provides clear boundaries, and has a disjunctive force (**imprecision**). The speaker isn't sure about the age.
- 2 puts focus on one number, without specifying the boundaries (**vagueness**). Either the speaker is not sure, or he knows the age but is too "lazy" to give the exact number.

Introduction: why use *around*?

Numerical approximation?

Approximation expressions: *around n*, *almost n*, *between n and m*...

1. She is between 28 and 32 y.o.
2. She is around 30 y.o.

Around vs between

- 1 provides clear boundaries, and has a disjunctive force (**imprecision**). The speaker isn't sure about the age.
- 2 puts focus on one number, without specifying the boundaries (**vagueness**). Either the speaker is not sure, or he knows the age but is too "lazy" to give the exact number.
- hypothesis: *around* gives "something more" about one's epistemic sate and one's priors (Égré and Verheyen 2018).

Table of contents

Key parameters

Plan

- "The bigger the wider"
- "The coarser the wider"

Order of magnitude (Égré and Verheyen 2018)

Introspective judgments:

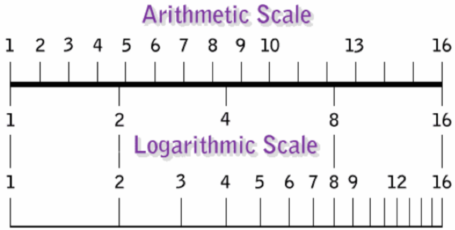
5. He bought this sandwich for about 4€ $\rightsquigarrow$ [3.5; 4.5]
6. He launched the reform for a cost of about 10 bn € $\rightsquigarrow$ [8bn; 12bn]

Order of magnitude (Égré and Verheyen 2018)

Introspective judgments:

5 He bought this sandwich for about 4€$\rightsquigarrow$ [3.5; 4.5]

6 He launched the reform for a cost of about 10 bn €$\rightsquigarrow$ [8bn; 12bn]



Intuition (Dehaene 2003)

- on a **log scale**, perceived spacing between consecutive numbers is "distorted";
- bigger numbers means smaller spacing;
- so, a fixed-size interval contains more big numbers than small numbers

Granularity (Cummins, Sauerland, and Solt 2012; Zhang and Schwarz 2012)

Introspective judgments:

5 She has around 20 marbles in her bag $\rightsquigarrow$ $[16; 24]$

6 She has around 15 marbles in her bag $\rightsquigarrow$ $[11; 19]$

7 ??She has around 17 marbles in her bag $\rightsquigarrow$ $\{17\}$??

8 She drove around 17 kilometers $\rightsquigarrow$ $[16.1; 17.9]$

Remarks

- 5, 6 and 7 are very close on the log scale and yet induce different intervals;
- 7 and 8 feature the same number (same granularity), but 7 seems infelicitous.

Two principles

- **granularity constraint**: coarser granularity, bigger interval;
- **granularity violation**: when the granularity of the number is less than or equal to the minimum granularity of the scale.

Two Bayesian models

Plan

- a model inspired by the Rational Speech Acts framework accounting for granularity effect;
- a model based on probabilistic intervals allowing for speaker uncertainty.

A Rational Speech Acts (RSA) model

Principle (Lassiter and N. D. Goodman 2013; Bergen, Levy, and N. Goodman 2016)

- one set of **utterances** (each with a cost and possible meanings), one set of **observations** (numbers), one set of interval **thresholds** (numbers);

A Rational Speech Acts (RSA) model

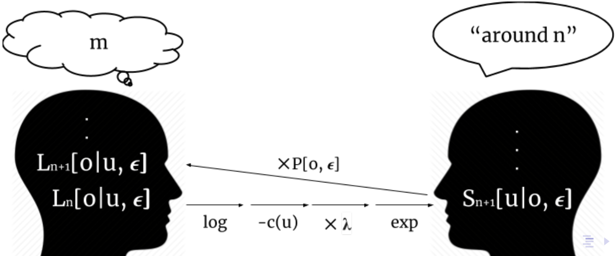
Principle (Lassiter and N. D. Goodman 2013; Bergen, Levy, and N. Goodman 2016)

- one set of **utterances** (each with a cost and possible meanings), one set of **observations** (numbers), one set of interval **thresholds** (numbers);
- mutually recursive Bayesian updates ("I know that you know that I know...");

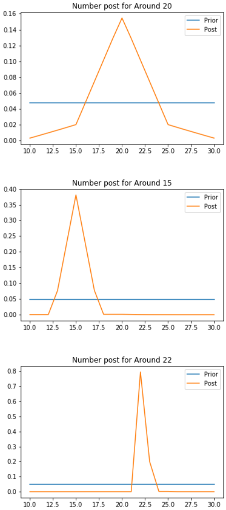
A Rational Speech Acts (RSA) model

Principle (Lassiter and N. D. Goodman 2013; Bergen, Levy, and N. Goodman 2016)

- one set of **utterances** (each with a cost and possible meanings), one set of **observations** (numbers), one set of interval **thresholds** (numbers);
- mutually recursive Bayesian updates ("I know that you know that I know..");
- optimality = tradeoff between **cost** and **informativity**.



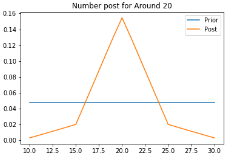
RSA and granularity

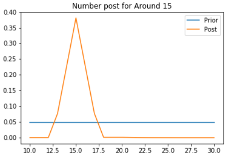


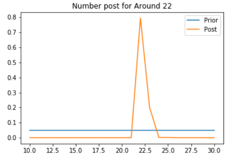
The trick

- granularity encoded in the **cost function**!

RSA and granularity



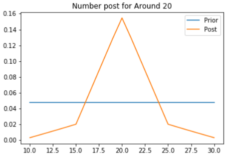


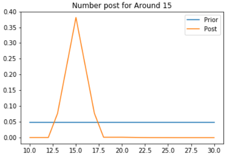


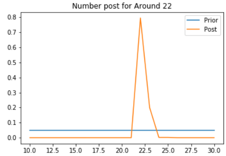
The trick

- granularity encoded in the **cost function**!
- c/i tradeoff ensures that the probability drops at the level of next coarser number (more optimal, and yet not used).

RSA and granularity







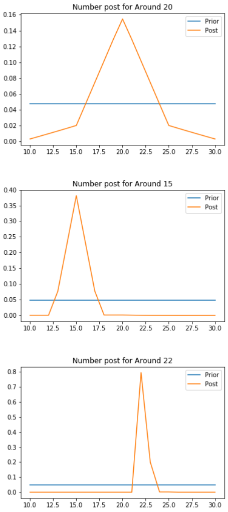
The trick

- granularity encoded in the **cost function**!
- c/i tradeoff ensures that the probability drops at the level of next coarser number (more optimal, and yet not used).

Caveats

- does not come for free... “good” granularity function?

RSA and granularity



The trick

- granularity encoded in the **cost function**!
- c/i tradeoff ensures that the probability drops at the level of next coarser number (more optimal, and yet not used).

Caveats

- does not come for free... "good" granularity function?
- RSA assumes that the speaker knows the exact number... not very realistic!

Probabilistic intervals (Égré and Verheyen 2018)

Principle: 2 levels of uncertainty

- when the speaker utters "around n", he thinks of a certain interval among a **set of possible intervals** (*e.g.* intervals centered around n);

Probabilistic intervals (Égré and Verheyen 2018)

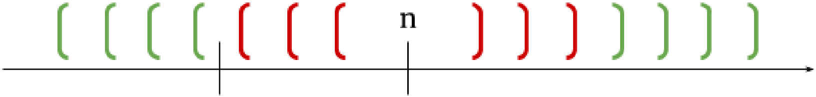
Principle: 2 levels of uncertainty

- when the speaker utters “around n”, he thinks of a certain interval among a **set of possible intervals** (*e.g.* intervals centered around n);
- according to the listener, **each possible interval has a certain probability** (*e.g.* relatively narrow intervals might be more probable);

Probabilistic intervals (Égré and Verheyen 2018)

Principle: 2 levels of uncertainty

- when the speaker utters “around n”, he thinks of a certain interval among a **set of possible intervals** (*e.g.* intervals centered around n);
- according to the listener, **each possible interval has a certain probability** (*e.g.* relatively narrow intervals might be more probable);
- and within a fixed interval, the **“real” number** is guessed according to a certain prior (*e.g.* central numbers might be more probable)
- use Bayes rule!



A simulation with uniform priors

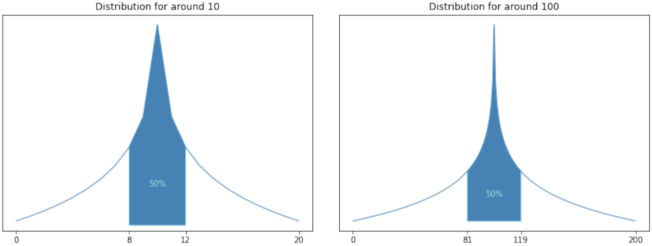


Figure: Curves generated using uniform distributions on intervals and numbers

Properties

- Symmetrical, scales with magnitude, not granularity

A simulation with uniform priors

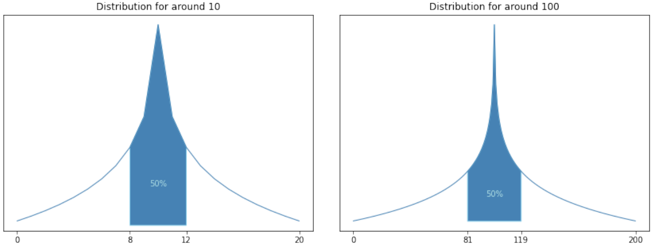


Figure: Curves generated using uniform distributions on intervals and numbers

Properties

- Symmetrical, scales with magnitude, not granularity
- **Increases the probability discrepancy between central and peripheral numbers** (posterior is more peaked than prior)... does *not* depend on priors.

Experiments

Goals

What

- "(Around|between|almost...) *n* people came to the party";

Goals

What

- "(Around|between|almost...) *n* people came to the party";
- elicitation of **intervals of possible values** (Channell 1994...) ;

Goals

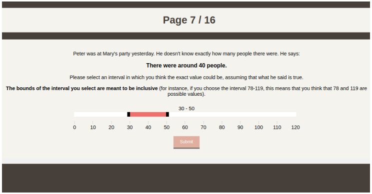
What

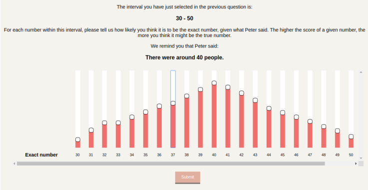
- "(Around|between|almost...) n people came to the party";
- elicitation of **intervals of possible values** (Channell 1994...) ;
- elicitation of **probability distributions** ont these intervals.

Goals

What

- "(Around|between|almost...) n people came to the party";
- elicitation of **intervals of possible values** (Channell 1994...) ;
- elicitation of **probability distributions** ont these intervals.

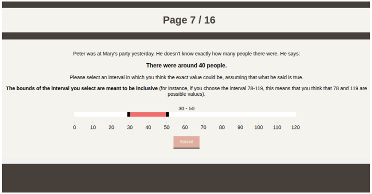


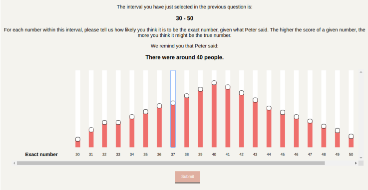


Goals

What

- "(Around|between|almost...) n people came to the party";
- elicitation of **intervals of possible values** (Channell 1994...) ;
- elicitation of **probability distributions** ont these intervals.





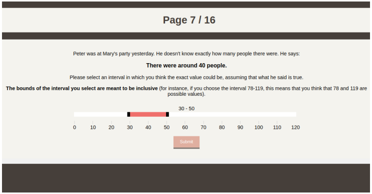
Why

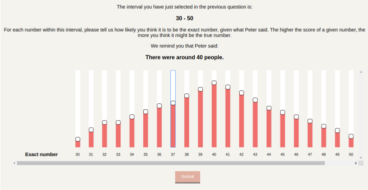
- to check the basic features of *around*;

Goals

What

- "(Around|between|almost...) *n* people came to the party";
- elicitation of **intervals of possible values** (Channell 1994...) ;
- elicitation of **probability distributions** ont these intervals.





Why

- to check the basic features of *around*;
- to compare *around* and *between* and **validate the model**.

Hypotheses and issues

Hypotheses

- with ***between***, the interval is already known, the listener simply uses the **prior** on it;

Hypotheses and issues

Hypotheses

- with ***between***, the interval is already known, the listener simply uses the **prior** on it;
- with ***around***, the interval is not known, the listener conditionalizes over multiple nested intervals, which mechanically increases the peakedness of the distribution (**post**);

Hypotheses and issues

Hypotheses

- with ***between***, the interval is already known, the listener simply uses the **prior** on it;
- with ***around***, the interval is not known, the listener conditionalizes over multiple nested intervals, which mechanically increases the peakedness of the distribution (**post**);
- therefore, **the *around*-distribution must be more peaked than the *between*-distribution**!

Hypotheses and issues

Hypotheses

- with ***between***, the interval is already known, the listener simply uses the **prior** on it;
- with ***around***, the interval is not known, the listener conditionalizes over multiple nested intervals, which mechanically increases the peakedness of the distribution (**post**);
- therefore, **the *around*-distribution must be more peaked than the *between*-distribution**!

Issues

- **target items**: which expressions should we compare?

Hypotheses and issues

Hypotheses

- with ***between***, the interval is already known, the listener simply uses the **prior** on it;
- with ***around***, the interval is not known, the listener conditionalizes over multiple nested intervals, which mechanically increases the peakedness of the distribution (**post**);
- therefore, **the *around*-distribution must be more peaked than the *between*-distribution**!

Issues

- **target items**: which expressions should we compare?
- **prediction**: what should we measure and test?

Hypotheses and issues

Hypotheses

- with ***between***, the interval is already known, the listener simply uses the **prior** on it;
- with ***around***, the interval is not known, the listener conditionalizes over multiple nested intervals, which mechanically increases the peakedness of the distribution (**post**);
- therefore, **the *around*-distribution must be more peaked than the *between*-distribution**!

Issues

- **target items**: which expressions should we compare?
- **prediction**: what should we measure and test?
- **design**: who sees what and when?

Target items

around n vs between x and y

- we should better compare distributions defined on the **same set of possible numbers**...

Target items

around n vs between x and y

- we should better compare distributions defined on the **same set of possible numbers**...
- but the participants are free to choose the domain of their distributions!

Target items

around n vs between x and y

- we should better compare distributions defined on the **same set of possible numbers**...
- but the participants are free to choose the domain of their distributions!
- we can somewhat “force” domain equality by **dynamically determining the target *between* item**;

Target items

around n vs between x and y

- we should better compare distributions defined on the **same set of possible numbers**...
- but the participants are free to choose the domain of their distributions!
- we can somewhat "force" domain equality by **dynamically determining the target *between* item**;
- but then, we cannot control for **order effects**...

Question		Answer
"Around 40?"	$\rightarrow$	[**32**; **48**]
Fillers	$\rightarrow$	...
"Between **32** and **48**?"	$\rightarrow$	[32; 48]

Metric

Model prediction

- the model predicted an increase of peakedness after hearing *around*, no matter the prior distribution;

Metric

Model prediction

- the model predicted an increase of peakedness after hearing *around*, no matter the prior distribution;
- this translates into: $\frac{Post(k)}{Post(k')} \geq \frac{Prior(k)}{Prior(k')}$ when k is more central than k';

Metric

Model prediction

- the model predicted an increase of peakedness after hearing *around*, no matter the prior distribution;
- this translates into: $\frac{Post(k)}{Post(k')} \geq \frac{Prior(k)}{Prior(k')}$ when k is more central than k';
- the *Post* distribution is induced by *around*; while the *Prior* distribution is induced by *between*.

Metric

Model prediction

- the model predicted an increase of peakedness after hearing *around*, no matter the prior distribution;
- this translates into: $\frac{Post(k)}{Post(k')} \geq \frac{Prior(k)}{Prior(k')}$ when k is more central than k';
- the *Post* distribution is induced by *around*; while the *Prior* distribution is induced by *between*.

Caveats

- best is to compare very distant numbers, but we should also avoid division by 0 and the use of salient numbers!

Metric

Model prediction

- the model predicted an increase of peakedness after hearing *around*, no matter the prior distribution;
- this translates into: $\frac{Post(k)}{Post(k')} \geq \frac{Prior(k)}{Prior(k')}$ when k is more central than k';
- the *Post* distribution is induced by *around*; while the *Prior* distribution is induced by *between*.

Caveats

- best is to compare very distant numbers, but we should also avoid division by 0 and the use of salient numbers!
- we should compare probabilities of numbers that are defined in both the *between* and *around* distributions at stake;

Metric

Model prediction

- the model predicted an increase of peakedness after hearing *around*, no matter the prior distribution;
- this translates into: $\frac{Post(k)}{Post(k')} \geq \frac{Prior(k)}{Prior(k')}$ when k is more central than k';
- the *Post* distribution is induced by *around*; while the *Prior* distribution is induced by *between*.

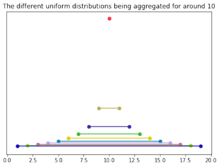
Caveats

- best is to compare very distant numbers, but we should also avoid division by 0 and the use of salient numbers!
- we should compare probabilities of numbers that are defined in both the *between* and *around* distributions at stake;
- we should average multiple ratios to increase robustness.

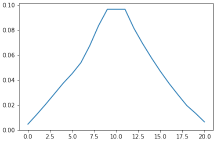
Design

Matched *vs* unmatched

- **matched-pairs**: comparing distributions defined on very similar domains becomes easier;
- **unmatched groups**: using averaged distributions may play in favor of our hypothesis for the wrong reasons
- indeed, when uniform distributions coming from different participants are "piled up", it creates a peaked distribution!



(a) What happens in the model



(b) What could happen with the data

Figure: An unmatched design would emulate what happens at the individual level in the model using inter-individual data: big confusion!

Summary

Pilots

- first pilot (n=20): assess interval width and inter-subject variability, select target numbers;

Summary

Pilots

- first pilot (n=20): assess interval width and inter-subject variability, select target numbers;
- three other pilots (n=15 each): collect subjective probability distributions with varying items, orders, input modes. See what gives the best performances.

Summary

Pilots

- first pilot (n=20): assess interval width and inter-subject variability, select target numbers;
- three other pilots (n=15 each): collect subjective probability distributions with varying items, orders, input modes. See what gives the best performances.

Main experiment

- ask 200 participants about *around*, *between* and *almost*;

Summary

Pilots

- first pilot (n=20): assess interval width and inter-subject variability, select target numbers;
- three other pilots (n=15 each): collect subjective probability distributions with varying items, orders, input modes. See what gives the best performances.

Main experiment

- ask 200 participants about *around*, *between* and *almost*;
- three possible numbers for the target trial: around 40, 50 or 60;

Summary

Pilots

- first pilot (n=20): assess interval width and inter-subject variability, select target numbers;
- three other pilots (n=15 each): collect subjective probability distributions with varying items, orders, input modes. See what gives the best performances.

Main experiment

- ask 200 participants about *around*, *between* and *almost*;
- three possible numbers for the target trial: around 40, 50 or 60;
- target items in fixed positions, randomized fillers;

Summary

Pilots

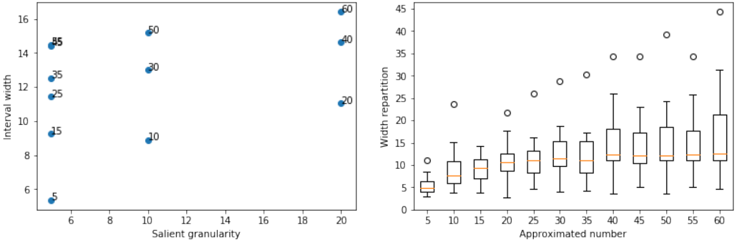
- first pilot (n=20): assess interval width and inter-subject variability, select target numbers;
- three other pilots (n=15 each): collect subjective probability distributions with varying items, orders, input modes. See what gives the best performances.

Main experiment

- ask 200 participants about *around*, *between* and *almost*;
- three possible numbers for the target trial: around 40, 50 or 60;
- target items in fixed positions, randomized fillers;
- target *between* item generated dynamically after *around*.

Results

First pilot



Observations

- effect of **granularity**: close numbers with different intrinsinc salient granularities give rise to different intervals (**coarser = wider**);
- effect of **order of magnitude**: distant numbers with same granularity give rise to different intervals (**bigger = wider**); bigger numbers also give rise to more variable intervals.

Main experiment

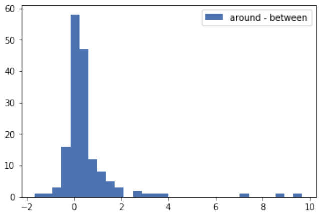


Figure: Repartition of the paired differences between the scores for *around* and the scores for *between*

Main experiment

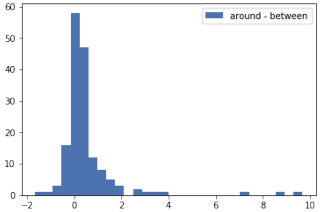


Figure: Repartition of the paired differences between the scores for *around* and the scores for *between*

Hypothesis testing

- Wilcoxon signed-rank test for matched pairs
- Significant difference between the scores for *around* and the scores for *between* (n=162, p=8.504×10^{-13}, effect size=0.56)

Main experiment

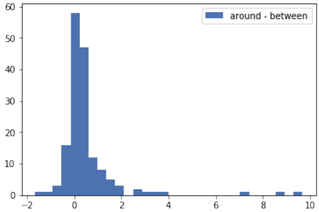


Figure: Repartition of the paired differences between the scores for *around* and the scores for *between*

Hypothesis testing

- Wilcoxon signed-rank test for matched pairs
- Significant difference between the scores for *around* and the scores for *between* (n=162, p=8.504×10^{-13}, effect size=0.56)

Interpretation

- however, we cannot be sure that the effect is really due to a subjective difference between the two approximators...

Main experiment

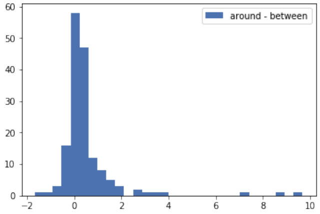


Figure: Repartition of the paired differences between the scores for *around* and the scores for *between*

Hypothesis testing

- Wilcoxon signed-rank test for matched pairs
- Significant difference between the scores for *around* and the scores for *between* (n=162, p=8.504×10^{-13}, effect size=0.56)

Interpretation

- however, we cannot be sure that the effect is really due to a subjective difference between the two approximators...
- indeed, order effects were not controlled...

Main experiment

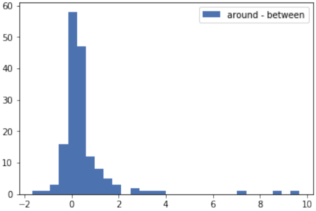


Figure: Repartition of the paired differences between the scores for *around* and the scores for *between*

Hypothesis testing

- Wilcoxon signed-rank test for matched pairs
- Significant difference between the scores for *around* and the scores for *between* (n=162, p=8.504×10^{-13}, effect size=0.56)

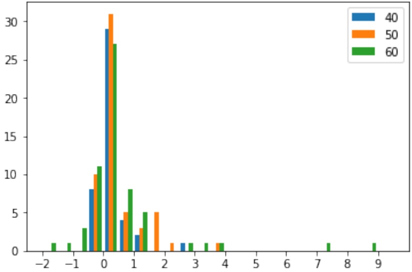
Interpretation

- however, we cannot be sure that the effect is really due to a subjective difference between the two approximators...
- indeed, order effects were not controlled...
- and the variability of the number could have interfered.

Exploratory analysis – Number effect (thanks to Steven Verheyen)

Number-by-number analysis

- participants were assigned randomly to 3 possible target numbers: 40, 50, 60;
- but the distribution of scores may vary with the number, *i.e.* one number "weighs more" in the test;
- we did three separate analyses to check that.

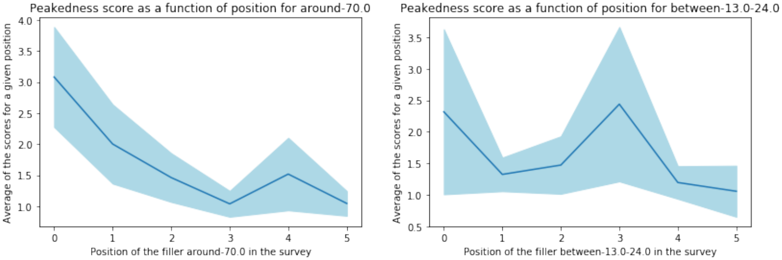


Number	40	50	60
Sample	44	57	61
Median	0.13	0.27	0.30
Std	0.50	1.42	1.60
p-value	4.32×10^{-4}	2.35×10^{-6}	3.15×10^{-4}

Exploratory analysis – Order effects

Assessing order effects

- "quick and dirty" analysis to verify whether order effects may explain a significant part of what has been observed;
- comparison of filler items (*around* and *between*) depending on their position in the survey (randomized over participants);
- participants become more and more uniform with time!



Conclusion

A mixed picture

- our results lack robustness;

Conclusion

A mixed picture

- our results lack robustness;
- our custom score is difficult to interpret, because its properties are largely unknown.

Conclusion

A mixed picture

- our results lack robustness;
- our custom score is difficult to interpret, because its properties are largely unknown.

What's next?

- starting from the PI model, develop a more refined model dealing with *both* speaker uncertainty and granularity;

Conclusion

A mixed picture

- our results lack robustness;
- our custom score is difficult to interpret, because its properties are largely unknown.

What's next?

- starting from the PI model, develop a more refined model dealing with *both* speaker uncertainty and granularity;
- better control for order effects, which leads to go back to a static design and use a more appropriate model (mixed model).

Selected references I

Leon Bergen, Roger Levy, and Noah Goodman. "Pragmatic reasoning through semantic inference". In: Semantics and Pragmatics 9 (May 2016). DOI: 10.3765/sp.9.20. URL: https://doi.org/10.3765/sp.9.20.

Émile Borel. Le Hasard. Paris: Librairie Félix Alcan, 1914.

Joanna Channell. Vague Language. Oxford University Press, 1994.

Chris Cummins, Uli Sauerland, and Stephanie Solt. "Granularity and scalar implicature in numerical expressions". In: Linguistics and Philosophy 35.2 (Apr. 2012), pp. 135–169. DOI: 10.1007/s10988-012-9114-0. URL: https://doi.org/10.1007/s10988-012-9114-0.

Stanislas Dehaene. "The neural basis of the Weber–Fechner law: a logarithmic mental number line". In: Trends in Cognitive Sciences 7.4 (Apr. 2003), pp. 145–147. DOI: 10.1016/s1364-6613(03)00055-x. URL: https://doi.org/10.1016/s1364-6613(03)00055-x.

Paul Égré and Anouk Barberousse. "Borel on the Heap". In: Erkenntnis 79.S5 (Apr. 2014), pp. 1043–1079. DOI: 10.1007/s10670-013-9596-3. URL: https://doi.org/10.1007/s10670-013-9596-3.

Selected references II

Paul Égré. “Vague judgment: a probabilistic account”. In: Synthese 194.10 (May 2016), pp. 3837–3865. DOI: 10.1007/s11229-016-1092-2. URL: https://doi.org/10.1007/s11229-016-1092-2.

Paul Égré. Qu’est-ce que le vague ? Chemins Philosophiques. Vrin, 2018.

Paul Egré and Benjamin Icard. “Lying and Vagueness”. In: The Oxford Handbook of Lying. Ed. by Jörg Meibauer. Oxford: Oxford University Press, 2018.

Paul Egré, David Ripley, and Steven Verheyen. “The sorites paradox in psychology”. In: The Sorites Paradox. Ed. by Sergei Oms and Elia Zardini. Forthcoming. Cambridge University Press.

Paul Égré and Steven Verheyen. “Vagueness, approximation and the Maxim of Quality”. In: Presentation. University of Groningen, Department of Theoretical Philosophy, avril 2018.

H. P. Grice. “Logic and Conversation”. In: ed. by Peter Cole and Jerry L. Morgan. New York: Academic Press, 1975, pp. 41–58.

Selected references III

Noah D. Goodman and Andreas Stuhlmüller. "Knowledge and Implicature: Modeling Language Understanding as Social Cognition". In: Topics in Cognitive Science 5.1 (Jan. 2013), pp. 173–184. DOI: 10.1111/tops.12007. URL: https://doi.org/10.1111/tops.12007.

Laurence Horn. "On the semantic properties of logical operators in English". PhD thesis. University of California, Los Angeles, 1972.

Laurence Horn. In: 990001925760204761. Washington, D.C.: Georgetown University Press, 1984. Chap. Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature, pp. 11–42. ISBN: 0878401199.

Christopher Kennedy. "Vagueness and grammar: the semantics of relative and absolute gradable adjectives". In: Linguistics and Philosophy 30.1 (Mar. 2007), pp. 1–45. DOI: 10.1007/s10988-006-9008-0. URL: https://doi.org/10.1007/s10988-006-9008-0.

Daniel Lassiter. "Vagueness as Probabilistic Linguistic Knowledge". In: Vagueness in Communication. Springer Berlin Heidelberg, 2011, pp. 127–150. DOI: 10.1007/978-3-642-18446-8_8. URL: https://doi.org/10.1007/978-3-642-18446-8_8.

Selected references IV

Daniel Lassiter and Noah D. Goodman. "Context, scale structure, and statistics in the interpretation of positive-form adjectives". In: Semantics and Linguistic Theory 23 (Aug. 2013), p. 587. DOI: 10.3765/salt.v23i0.2658. URL: https://doi.org/10.3765/salt.v23i0.2658.

Paul Marty, Emmanuel Chemla, and Benjamin Spector. "Phantom readings: the case of modified numerals". In: Language, Cognition and Neuroscience 30.4 (June 2014), pp. 462–477. DOI: 10.1080/23273798.2014.931592. URL: https://doi.org/10.1080/23273798.2014.931592.

Ciyang Qing and Michael Franke. "Gradable adjectives, vagueness, and optimal language use: A speaker-oriented model". In: Semantics and Linguistic Theory 24 (Apr. 2015), p. 23. DOI: 10.3765/salt.v24i0.2412. URL: https://doi.org/10.3765/salt.v24i0.2412.

Uli Sauerland. "Scalar Implicatures in Complex Sentences". In: Linguistics and Philosophy 27.3 (June 2004), pp. 367–391. DOI: 10.1023/b:ling.0000023378.71748.db. URL: https://doi.org/10.1023/b:ling.0000023378.71748.db.

Selected references V

Benjamin Spector. "The pragmatics of plural predication: Homogeneity and Non-Maximality within the Rational Speech Act model". In: December 2017. Ed. by Proceedings of the 21st Amsterdam Colloquium. 2017.

Steven Verheyen, Sabrina Dewil, and Paul Égré. "Subjectivity in gradable adjectives: The case of tall and heavy". In: Mind & Language 33.5 (May 2018), pp. 460–479. DOI: 10.1111/mila.12184. URL: https://doi.org/10.1111/mila.12184.

Timothy Williamson. "Precis of Vagueness". In: Philosophy and Phenomenological Research 57.4 (Dec. 1997), p. 921. DOI: 10.2307/2953810. URL: https://doi.org/10.2307/2953810.

Charles Zhang and Norbert Schwarz. "How and Why 1 Year Differs from 365 Days: A Conversational Logic Analysis of Inferences from the Granularity of Quantitative Expressions". In: Journal of Consumer Research 39 (Aug. 2012). DOI: 10.1086/662612.

Additional data (many thanks to Steven Verheyen)

About intervals and social constraints:

1. – I ate around 10 cookies... –Liar, you ate 12 of them!!
2. – I ate around 10 cookies... –*Liar, you ate 8 of them!!
3. – I helped around 10 children –*Liar, you helped 12 of them!!
4. – I helped around 10 children –Liar, you helped 8 of them!!

Role of valence

- negative valence: *around* closer to an upper bound (honesty)
- positive valence: *around* closer to a lower bound (humility)

Additional data (inspired from Zhang and Schwarz 2012)

5. I ran around 1.5 km
6. I ran around 15 hundred meters
7. I ran around 15 hm
8. I ran around 1500 m

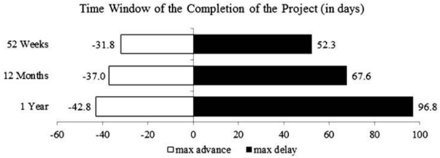


Figure: weeks *vs* months *vs* years, Zhang and Schwarz 2012

Unit conveys granularity

- coarser unit = coarser scale = less precise = less reliable;
- effect disappears when Gricean cooperation no longer holds;
- if unit is too precise, sometimes reverse effect ("too good to be true!").

Proof of $W \propto n$ on log scale

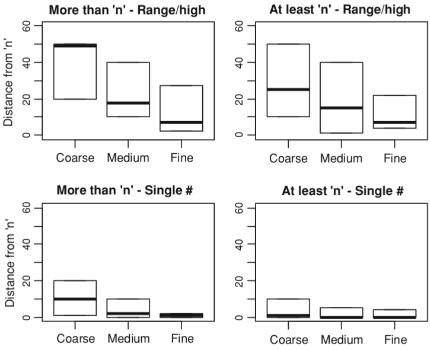
Suppose we want a interval with a fixed size W around n, *on a log scale*. For the size W to remain fixed, what should be the value of the semi-width ϵ, depending on n, on a liner scale?

$$\begin{aligned}
Size([n-\epsilon; n+\epsilon]) = W \quad & \iff \quad \log\left(\frac{n+\epsilon}{n-\epsilon}\right) \propto W \\
& \iff \quad \frac{n+\epsilon}{n-\epsilon} \propto \exp(W) \\
& \iff \quad n+\epsilon \propto \exp(W)(n-\epsilon) \\
& \iff \quad \epsilon(1+\exp(W)) \propto n(\exp(W)-1) \\
& \iff \quad \epsilon \propto \frac{n(\exp(W)-1)}{1+\exp(W)} \\
& \iff \quad \epsilon \propto n
\end{aligned}$$

Granularity effects on quantified numerals (Cummins, Sauerland, and Solt 2012)

Study

- tested *more than n* and *at least n*;
- pointwise estimate or intervals;
- with coarser granularities, average distance to *n* becomes larger;
- effect disappears when number is primed.



Granularity as exhaustification

Why granularity *implicatures*?

- granularity conveys a salient scale, potentially different from unit-induced scale: 2-by-2, 5-by-5, 10-by-10 etc.
- a granularity implicature is an exhaustification process pretty much alike "*exaclty*"-implicatures with bare numerals...
- ... but on the salient scale! (plus convexity assumption)

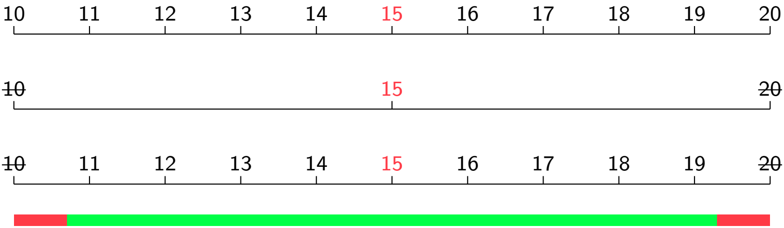
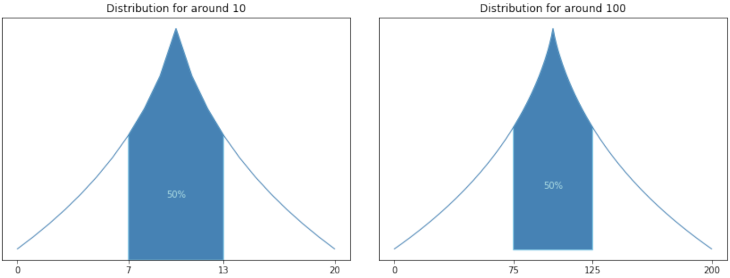


Figure: Granularity implicature for *around 15*: exhaustification on a 5-by-5 scale, then back to 1-by-1 scale (scale assumed for the given unit)

Probabilistic unconstrained intervals



Properties

- intervals containing n, but not necessarily centered around n;
- scales with magnitude;
- less peaked.

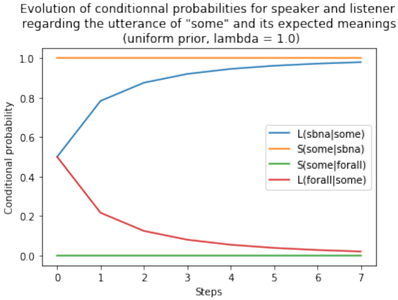
RSA formulae

$$
\begin{array}{rcll}
\mathbb{L}_0[k|u,\epsilon] & \propto & \mathbb{1}_{\{k\star\epsilon\}}\mathbb{L}_0[k] & \text{[base case]} \\
\mathbb{S}_1[u|k,\epsilon] & \propto & \exp\left(\lambda(\log(\mathbb{L}_0[k|u,\epsilon]) - c(u))\right) & \text{[inductive case]} \\
\mathbb{L}_1[k,\epsilon|u] & \propto & \mathbb{S}_N[u|k,\epsilon]\mathbb{L}_0[k,\epsilon] & \text{[inductive case]} \\
\mathbb{L}_1[k|u] & \propto & \sum_\epsilon \mathbb{L}_1[k,\epsilon|u]\mathbb{L}_0[\epsilon] & \text{[``Post on k'']} \\
\mathbb{L}_1[\epsilon|u] & \propto & \sum_k \mathbb{L}_1[k,\epsilon|u]\mathbb{L}_0[k] & \text{[``Post on } \epsilon\text{'']}
\end{array}
$$

Explanations

- first step: for fixed u (e.g. “around x”) and ϵ, just keep the observations that are in [n-ϵ; n+ϵ]; all the other have 0 probability.
- speaker step: the softmax allows to pick some non optimal possibilities with a non-zero (but very small) probability

RSA with quantifiers (replication)



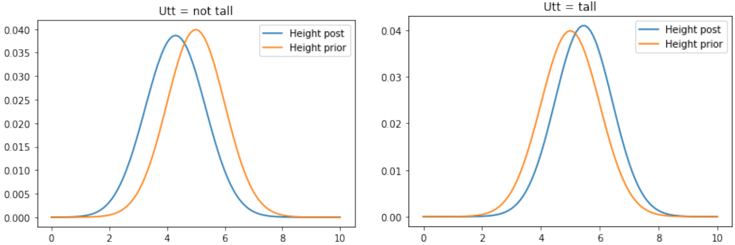
Some *vs* all

- at the beginning, "some" can mean all ($\forall$) are some but not all ($\exists_{\neg\forall}$), and "all" definitely means $\forall$.
- this asymmetry causes the meaning of "some" to converge to $\exists_{\neg\forall}$ after a few iterations.

Caveats

- Sensitive to parameter λ!
- λ is the temperature, higher λ means faster convergence but possibly to a "wrong" optimum.

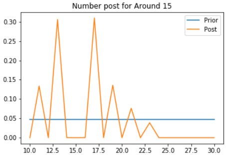
RSA with gradable adjectives (replication)

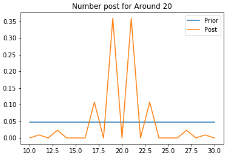


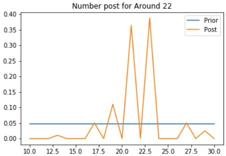
Properties

- A negative utterance ("not tall"), shifts the height prior to the **left**: the listener expects the person to be **smaller**;
- A positive utterance ("tall"), shifts the height prior to the **right**: the listener expects the person to be **taller**;
- "Not tall" has a bigger effect on the prior, because it is more costly. If it has been uttered, then the person is *really* small

RSA with competing *around* and *exactly*







Observations

- "simple" numbers are strongly dispreferred in the interpretation of *around* n, even though they are close to n, or even equal to n;
- indeed, saying an exact version of these numbers appears as strictly optimal, but this has not been done!

RSA with competing *around* and *between*

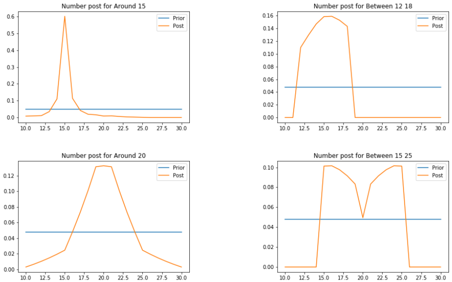
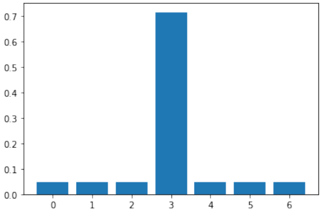


Figure: Prior and posterior distributions for the actual number k, for different target numbers (15, 20) and approximators (*around*, *between*)

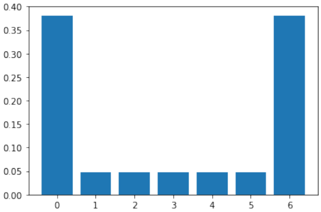
Observations

- *around* is very peaked as in other simulations;
- *between* is "anti-peaked": if the bounds are as they are, their probability *must* be high enough (otherwise, narrower bounds would have been more optimal).

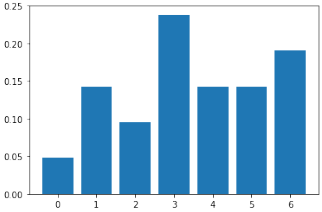
Arguments for a robust score (not a just one ratio)



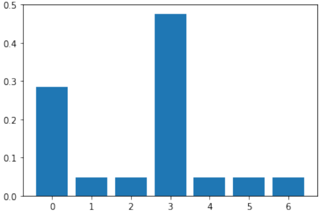
(a) True positive, s = 15.0



(b) True negative, s = 0.125



(c) False positive, s = 5.0



(d) False negative, s = 1.67

PI formula

Bayesian reasoning

- we assume that the speaker chooses an interval in the set of all-positive intervals centered around n;
- for all $i \in [0; n]$, we call $\mathcal{A}_i^n$ the event "speaker chooses interval $[n-i; n+i]$";
- this constitutes a partition of the possible events;
- if we consider event $\mathcal{A}_i^n$, the probability that the speaker chooses number k given $\mathcal{A}_i^n$ is $\mathcal{P}[k|\mathcal{A}_i^n]$.
- note that if $k < n-i$ or $k > n+i$, *i.e.* $i < |n-k|$, $\mathcal{P}[k|\mathcal{A}_i^n] = 0$

Bayes: $\mathbb{P}[k] = \sum_{i=0}^{n} \mathcal{P}[k|\mathcal{A}_i^n]\mathbb{P}[\mathcal{A}_i^n] = \cancel{\sum_{i=0}^{|n-k|-1} \mathcal{P}[k|\mathcal{A}_i^n]\mathbb{P}[\mathcal{A}_i^n]} + \sum_{i=|n-k|}^{n} \mathcal{P}[k|\mathcal{A}_i^n]\mathbb{P}[\mathcal{A}_i^n]$

Ratio inequality

We assume that $|n-k| < |n-k'|$, *i.e.* k closer to n than k'.

$$
\begin{aligned}
\frac{\mathbb{P}[k|\text{'around n'}]}{\mathbb{P}[k'|\text{'around n'}]} &= \frac{\sum_{i=|n-k|}^{n} \mathbb{P}[k|\mathcal{A}_i^n]\mathbb{P}[\mathcal{A}_i^n]}{\sum_{i=|n-k'|}^{n} \mathbb{P}[k'|\mathcal{A}_i^n]\mathbb{P}[\mathcal{A}_i^n]} \\
&= \frac{\sum_{i=|n-k|}^{|n-k'|-1} \mathbb{P}[k|\mathcal{A}_i^n]\mathbb{P}[\mathcal{A}_i^n] + \sum_{i=|n-k'|}^{n} \mathbb{P}[k|\mathcal{A}_i^n]\mathbb{P}[\mathcal{A}_i^n]}{\sum_{i=|n-k'|}^{n} \mathbb{P}[k'|\mathcal{A}_i^n]\mathbb{P}[\mathcal{A}_i^n]} \\
&\geq \frac{\sum_{i=|n-k'|}^{n} \mathbb{P}[k|\mathcal{A}_i^n]\mathbb{P}[\mathcal{A}_i^n]}{\sum_{i=|n-k'|}^{n} \mathbb{P}[k'|\mathcal{A}_i^n]\mathbb{P}[\mathcal{A}_i^n]} \\
&\geq \frac{\mathbb{P}[k|\mathcal{A}_n^n]\mathbb{P}[\mathcal{A}_n^n]}{\mathbb{P}[k'|\mathcal{A}_n^n]\mathbb{P}[\mathcal{A}_n^n]} + \frac{\sum_{i=|n-k'|}^{n-1} \mathbb{P}[k|\mathcal{A}_i^n]\mathbb{P}[\mathcal{A}_i^n]}{\sum_{i=|n-k'|}^{n-1} \mathbb{P}[k'|\mathcal{A}_i^n]\mathbb{P}[\mathcal{A}_i^n]} \\
&\geq \frac{\mathbb{P}[k|\mathcal{A}_n^n]\mathbb{P}[\mathcal{A}_n^n]}{\mathbb{P}[k'|\mathcal{A}_n^n]\mathbb{P}[\mathcal{A}_n^n]} \\
\frac{\mathbb{P}[k|\text{'around n'}]}{\mathbb{P}[k'|\text{'around n'}]} &\geq \frac{\mathbb{P}[k|\mathcal{A}_n^n]}{\mathbb{P}[k'|\mathcal{A}_n^n]} \\
\text{ratio of the posteriors} &\geq \text{ratio of the priors}
\end{aligned}
$$

Score formula

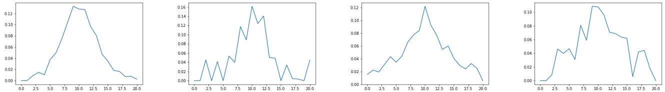
$\mathbb{A}$, $\mathbb{B}$ are respectively the *around* and the *between* distributions. E is the set of salient numbers: bounds of *between* and target value of *around*.

$$
\begin{aligned}
\forall x \in Supp(\mathbb{A} \cap \mathbb{B}) \quad &, \quad d(x) = |n - x| \\
E \quad &= \quad \{n, min(Dom(\mathbb{B})), max(Dom(\mathbb{B}))\} \\
F \quad &= \quad \{(x, y) \in Supp(\mathbb{A} \cap \mathbb{B}) \mid d(x) < d(y) \wedge x, y \notin E\} \\
s(\mathbb{P}) \quad &= \quad \frac{1}{|F|} \sum_{(x,y) \in F} \frac{\mathbb{P}(x)}{\mathbb{P}(y)}
\end{aligned}
$$

Benchmark of the scores (1)

Goal

- compare different versions of our score with more standard metrics (mass ratio, kurtosis);
- use artificial distributions (uniform, Gaussian, Laplace with different std), more or less noisy (4 levels);
- compare how the different metrics sort the distributions.



(a) Gaussian, spread=0.3, noise=0.01

(b) Gaussian, spread=0.3, noise=0.05

(c) Laplace, spread=0.3, noise=0.01

(d) Laplace, spread=0.3, noise=0.05

Figure: Example of noisy distributions generated for benchmarking the peakedness scores

Benchmark of the scores (2)

Score	Transpositions
Simple ratio	17
Averaged ratio	16
Mass ratio	15
Kurtosis	14

Table: Number of transpositions needed to change the ordering induced by a given score into the "gold-standard" ordering.

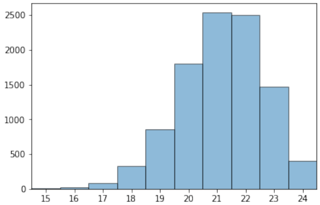


Figure: Distribution of the number of transposition needed to transform a random permutation of size 25 into the identity permutation (10,000 trials)

Alternative

- number of transpositions gives the same importance to "big" swaps and to "small" swaps;
- by using the sum of the distances between each item and its position in the "gold standard" ordering, we take these effects into account...

Designs (pilots)

Block	Numbers	Conveyed g	Critical trials	Control trials	Total
Block 1	{20, 40, 60}	20	3	3	6
Block 2	{20, 40, 60}	20	6	3	3
Block 3	+{10, 30, 50}	10	9	9	18
Block 4	+{10, 30, 50}	10	9	9	18
Block 5	+{5, 15, 25, 35, 45, 55}	5	24	24	48
Block 6	+{5, 15, 25, 35, 45, 55 }	5	24	24	48
Block 7	{(2, 60), (40, 60), (15, 25)}	—	0	3	3
		Total	72	72	147

Table: Pilot 1

Approximator	Number	Lower	Upper bound
almost	20	NA	NA
at least	110	NA	NA
around	target	NA	NA
less than	15	NA	NA
between	NA	0	20
at most and at least	NA	80	100
around	70	NA	NA
between	NA	target	target

Table: Pilot 2

Approximator	Number	Lower	Upper	Roundness
between	NA	86	93	non round
around	70	NA	NA	round
almost	25	NA	NA	round
around	target	NA	NA	round
almost	90	NA	NA	round
between	NA	13	24	non round
almost	60	NA	NA	round
around	20	NA	NA	round
between	NA	target	target	?

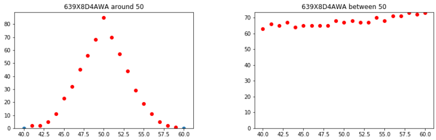
Table: Pilots 3 and 4

Design (main experiment)

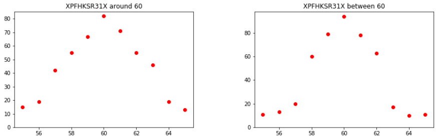
Approximator	Number	Lower	Upper	Roundness	Position
between	NA	80	90	round	randomized
between	NA	13	24	non round	randomized
around	70	NA	NA	round	randomized
around	36	NA	NA	non round	randomized
almost	60	NA	NA	round	randomized
almost	24	NA	NA	non round	randomized
around	target	NA	NA	round	4
between	NA	target	target	?	8

Table: Main experiment

Some participants' densities



(a) Participant with biggest score difference (s(around)-s(between))



(b) Participant with smallest score difference (s(around)-s(between))

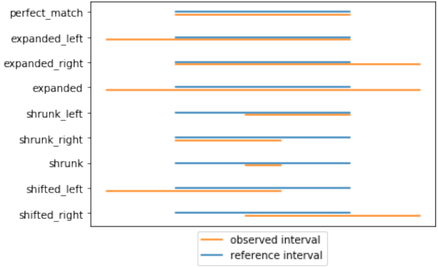
Observation

- the biggest bias toward a peaked *around* appears bigger than the biggest bias toward a peaked *between*...
- if not all the participants comply to our model, at least they do not exactly follow the inverse tendency!

Study of *between*-intervals (1)

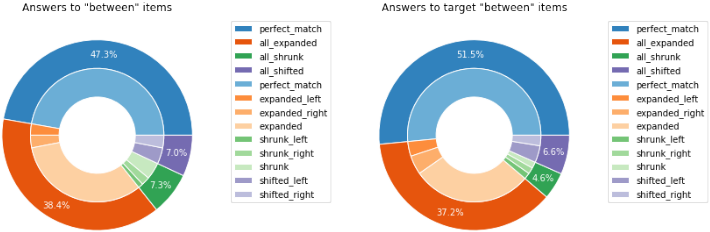
Goals

- evaluate the performance of the participants for the *between*-trials (interval task);
- study the different kinds of “errors” related to the *between*-intervals;
- see whether there is a rationale behind these “errors”.



	Target	All
# of trials	196	588
# of non-overlapping	1	7
% of non-overlapping	0.51 %	1.19 %

Study of *between*-intervals (2)



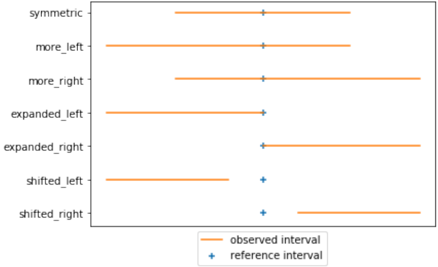
Observations

- perfect match in half of the cases;
- the predominant type of "error" is interval expansion, and especially cases of strict expansion (both sides);
- "phantom readings" ("at least(between(n, m))", Marty, Chemla, and Spector 2014) are not prominent in our case...normal given syntactic structure!

Study of *around*-intervals (1)

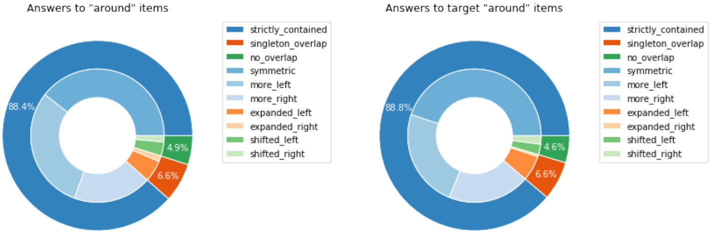
Goals

- evaluate the performance of the participants for the *around*-trials (interval task);
- study the different kinds of "errors" related to the *around*-intervals;
- see whether there is a rationale behind these "errors".



	Target	All
# of trials	196	588
# of non-overlapping	9	29
% of non-overlapping	4.59 %	4.93 %

Study of *around*-intervals (2)



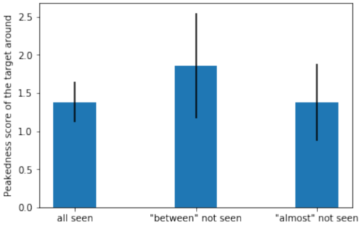
Observations

- *n* “strictly” contained in the interval (not a bound) in almost 90% of the cases;
- bias toward left: cannot be explained by log scale...
- granularity? Not only because targets are skewed;u
- no “phantom readings”.

Exploratory analysis – Effect of available alternatives

Available alternatives and order effects

- some participants may face the first target item (*around*) having seen all kinds of approximators (*around*, *between* and *almost*), or a strict subset of them;
- given what they know about the possible alternatives, they would want to give contrastive distributions;



Test

- we compared the peakedness of target *around* items depending on the alternatives already presented (MW+Bonferroni)
- no huge difference.