In search of pragmatic discounting: Lessons from direct discourse and causal adjuncts John Duff, Pranav Anand, Adrian Brasoveanu, and Amanda Rysling UC Santa Cruz Linguistics

Processing Meets Semantics (PROMS) @ Utrecht & 3 September 2022 jduff@ucsc.edu 1

Linguistic representations and comprehension processes

- Syntactic representations are constructed during parsing, constrain analysis of incoming material (e.g. Traxler & Pickering, 1996; Dillon et al., 2013)
- What about pragmatic representations?
 - Guide expectations for online input (e.g. Clifton & Frazier, 2012)
 - Affect the interpretation of:
 - Syntactic ambiguity (e.g. Rohde et al. 2011)
 - Pronouns (e.g. Kehler & Rohde, 2013, 2017)
 - Ellipsis (e.g. Kroll, 2020)





A larger role for pragmatics?

- Dillon and colleagues (2014, 2017): Pragmatic structure organizes processing.
 - **Pragmatic Discounting:** Some pragmatic constituents are encapsulated so that they exert less influence on later operations.

Jack took a Dutch train,

- But evidence comes only from one construction, appositive relative clauses.
 - What kind of constituent conditions discounting?
 - Can we find evidence for discounting with other constructions?





which was yellow.



Discounting isn't about pragmatic units.

Review and replicate discounting effects for appositives.

Consider whether discounting is tied to speech act representations.
 Examine direct discourse reports. (...said "Dutch trains are yellow.")
 No discounting effects.

Consider whether discounting is tied to discourse unit representations. ➡ Examine causal adjuncts. (...because Dutch trains are yellow.)

➡ No discounting effects.



Roadmap

- **1.** Introduction
- 2. Appositives are discounted
- 3. Direct discourse reports are not discounted
- 4. Causal adjuncts are not discounted
- 5. Discussion and conclusions



Restrictive and appositive relative clauses

- **Restrictive RCs** (RRCs) restrict the entities picked out by their host.
- Appositive RCs (ARCs) do not restrict, they provide a parallel comment.

RRC: Do the trains which leave from Amsterdam arrive on time? **ARC**: Do the trains, which leave from Amsterdam, arrive on time?

Modern consensus in formal semantics: ARCs somehow constitute a separate pragmatic unit from their matrix. (e.g. Potts, 2005; Arnold, 2007; AnderBois et al, 2015; Jasinskaja, 2016)



Pragmatic Discounting, in full

Pragmatic Discounting: (in the spirit of Dillon et al, 2017)

 (i) The online interpretation of natural language depends on the construction of pragmatic representations roughly the size of a sentence.
 (e.g. Grosz & Sidner, 1986)

 (ii) After the complete interpretation of one such unit, the material associated with it is discounted in later parsing and decision making.
 (e.g. Potter & Lombardi, 1990)

Evidence: ARC content interferes less in (a) filler-gap resolution and (b) subject dependencies, and (c) is less influential in judgments.



Dillon et al. (2017): Discounting and filler-gap resolution

was cooking dinner for ____.

Filler-gap resolution is harder across an RRC than across an ARC.

- ARCs show reduced penalty in ratings.
- ARCs show reduced slowdown (in eyetracking) at the gap site (____).

The butcher asked who the lady (,) who bought Italian ham (,)



- Attractors in ARCs show reduced attraction effects in SPR.
- But note: Verbs in both RCs are equally susceptible to matrix attractors.

Matrix Attractor, RC V: The musicians (,) who the reviewer praise (,) were...

Recent work: Discounting and agreement attraction CUNY/HSP talks: McInnerney & Atkinson (2020); Kim & Xiang (2022)

RC Attractor, Matrix V: The waitress (,) who sat near the girls (,) were...

Agreement is susceptible to more interference from an RRC than an ARC.





Dillon et al. (2014): Discounting and judgments

That evil man...

RRC: who was on the cruise (Mary took to the Pacific Islands) ARC: the one who was on the cruise (Mary took to the Pacific Islands), ...tried to intimidate the waitress.

Complexity penalties in ratings are reduced for ARCs.

- Note: Consistent across medial and final RCs.

 Assumption here: Ratings require integration of relevant evidence from across the rated stimulus, potentially weighting some evidence more than others.



E1: Replicating Dillon et al. (2014)

That man...

Short

who was on the cruise Mary took to the who was on the RRC Pacific Islands cruise

the one who was on the one who was on the cruise Mary took ARC the cruise, to the Pacific Islands,

Original design: 24 items in 8 conditions (3 observations each per participant)

Long

...tried to throw a waitress overboard.

Our designs: 32 items in 4 conditions (8 observations each per participant)

E1: Participants & Procedure

- n = 48 native English speakers recruited and paid \$12/hr through Prolific
- Procedure followed Dillon et al. (2014):
 - 7-point "naturalness" judgments
 - Same guided practice prompts



Randomized with same 72 grammatical and ungrammatical fillers



E1: Results

- Penalty for longer RCs: $\hat{\beta} = -0.24, \, _{95\%}(-0.36, \, -0.13)$
- ARCs rated more natural: $\hat{\beta} = 0.27, _{95\%}(0.11, 0.42)$
- Qualified by an interaction, smaller length penalty for ARCs: $\hat{\beta} = 0.08, _{95\%}(0.01, 0.16)$
- $BF_{10} = 5.75$, indicating moderate evidence for a discounting interaction of the expected size





Models fit for BF analysis used empirical priors derived from Dillon et al. (2014).

E1: Discussion

- The critical Dillon et al. (2014) discounting interaction replicates with our population and power.
- We take differences in complexity effects with naturalness judgments to provide a simple, flexible, and reliable measure for the presence of discounting.
- main goal here.

• We can then look for discounting effects in other constructions to resolve our

What type of pragmatic representation could Pragmatic Discounting be targeting?



Roadmap

- **1.** Introduction
- 2. Appositives are discounted
- 3. Direct discourse reports are not discounted
- 4. Causal adjuncts are not discounted
- 5. Discussion and conclusions



Speech acts

ARCs, and not RRCs, contribute their own speech acts.

Speech acts:

The minimal linguistic unit which may carry communicative purpose. (Austin, 1962; Ross, 1970; Stalnaker, 1978; Farkas & Bruce, 2010...)

E.g., only ARCs have a distinct illocutionary force from their matrix. (Arnold, 2007; Murray, 2014; AnderBois et al. 2015)

RRC: Do the trains which leave from Amsterdam arrive on time? **ARC:** Do the trains, which leave from Amsterdam, arrive on time?







Indirect and direct discourse speech reports

- Indirect discourse (ID) reports are integrated descriptions of speech meaning.
- **Direct discourse** (DD) reports are mimetic re-creations of speech form.

Only DD contributes its own speech act (one from another conversation).

ID: Did Morwenna say that Dutch trains are yellow?

DD: Did Morwenna say "Dutch trains are yellow"?

We know the difference is relevant in on-line reading:

Comprehenders modulate reading speed to simulate talker speed for only DD. (Yao & Scheepers, 2011; see also Alexander & Nygaard, 2008)









E2: Predictions

Speech Act Discounting:

content and the communicative context.

with it is discounted in later parsing and decision making.

Prediction: DD should also show discounting effects.

- The online interpretation of natural language depends on the construction of speech act representations in order to compute relationships between the
- After the complete interpretation of a speech act, the material associated

 - In particular, adding complexity within a DD report should have a smaller effect on ratings than adding complexity within an ID report.



E2: Stimuli & Procedure

Evan said...

Short

that the cruise departed ID three hours behind schedule.

"The cruise departed three DD hours behind schedule."



Long

that the cruise Mary took to the Pacific Islands departed three hours behind schedule.

"The cruise Mary took to the Pacific Islands departed three hours behind schedule."

Same procedure as E1: naturalness ratings, 32 items, n = 48 Prolific participants



E2: Results

- Penalty for longer reports: $\hat{\beta} = -0.51, _{95\%}(-0.68, -0.36)$
- No credible interaction: consistent length penalty: $\hat{\beta} = 0.03, _{95\%}(-0.05, 0.11)$
- $BF_{10} = 0.12$, indicating moderate evidence for the <u>absence</u> of a discounting interaction of the expected size







E2: Discussion

- The critical discounting interaction is absent, counter the predictions of Speech Act Discounting.
- Direct discourse is not discounted.

Speech acts aren't the relevant representation for discounting.

• Two replications and an attempt to test the Dillon et al. (2017) effect concur:



Roadmap

- **1.** Introduction
- 2. Appositives are discounted
- 3. Direct discourse reports are not discounted
- 4. Causal adjuncts are not discounted
- 5. Discussion and conclusions



Discourse units

ARCs, and not RRCs, contribute their own discourse units.

Discourse units:

The units of meaning which may participate in coherence relations. (Hobbs, 1979; Asher & Lascarides, 2003; Kehler & Rohde, 2013...)

E.g., only ARCs may head coherence relations like Background. (Burton-Roberts, 1999; Jasinskaja, 2016; though cf. Hoek et al, 2021)

RRC: Today, Lisa talked to the man she saw_{t1} downtown yesterday. She was_{t2} shopping for notebooks.
t1 ⊆ t2

ARC: Today, Lisa talked to Manuel, who she saw_{t1} downtown yesterday. She was_{t2} shopping for notebooks.
OK t1 ⊆ t2



Adjunct clauses with when and because

When: **Because:** Morwenna is always early because she takes the train.

- **Because** adjuncts mark the tails of Explanation relations. \bullet
- When adjuncts restrict the temporal index described by their matrix. (e.g. Johnston, 1994; Larson & Sawada, 2012; De la Fuente, 2015)

quantification only ever happens internal to a single discourse unit.)

- Morwenna is always early when she takes the train.

(e.g. Kehler, 2002; Asher & Lascarides, 2003)

(Critical linking assumption: Language doesn't have quantifiers over discourse units, so



E3: Predictions

Discourse Unit Discounting:

The online interpretation of natural language depends on the construction of discourse units in order to compute implicit relations between internally-coherent segments of information.

After the complete interpretation of a discourse unit, the material associated with it is discounted in later parsing and decision making.

Prediction: Because adjuncts should also show discounting effects.

In particular, adding complexity within a because adjunct should have a smaller effect on ratings than adding complexity within a when adjunct.



E3: Stimuli & Procedure

Evan often complains to the travel agent...

Short

when storms delay When the cruises.

because storms delay Because the cruises.

Temporal quantifiers (often, always, usually, rarely) throughout Same procedure as E1: naturalness ratings, 32 items, n = 48 Prolific participants



Long

when storms delay the cruises Mary takes to the Pacific Islands.

because storms delay the cruises Mary takes to the Pacific Islands.





E3: Results

- Penalty for longer adjuncts: $\hat{\beta} = -0.53, _{95\%}(-0.63, -0.43)$
- No credible interaction: consistent length penalty: $\hat{\beta} = 0.02, _{95\%}(-0.05, 0.10)$
- $BF_{10} = 0.17$, indicating moderate evidence for the <u>absence</u> of a discounting interaction of the expected size





Models fit for BF analysis used empirical priors derived from Dillon et al. (2014).

E3: Discussion

- The critical discounting interaction is absent, counter the predictions of Discourse Unit Discounting.
- Follow-up comparisons guard against the worry that attachment height

Discourse units aren't the relevant representation for discounting.

preferences may be concealing the interaction. (More details available.)



Roadmap

- **1.** Introduction
- 2. Appositives are discounted
- 3. Direct discourse reports are not discounted
- 4. Causal adjuncts are not discounted
- 5. Discussion and conclusions



Discounting isn't about pragmatic units

$\bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet$

Speech Acts? • ARCs vs. RRCs ✓ • DD vs. ID X

$\bullet \bullet \bullet$

Discourse Units? ● ARCs vs. RRCs ◆ Because vs. When X



Discounting Not-At-Issue Content?

Not-at-issue content:

- Potts (2005): A formal category of non-truth-conditional meaning. • Since then: A gradient pragmatic category of "backgrounded" information. (AnderBois et al, 2015; Syrett & Koev, 2015; but cf. Potts, 2012)

Prediction: Discounting should correlate with "backgroundedness".

• Final ARCs pass fewer backgrounding ARCs that answer a QUD are diagnostics than medial ARCs. notionally less backgrounded. Are they less discounted? Are they less discounted?

 \rightarrow No. (Dillon et al, 2014)

- ➡ No. (Kroll & Wagers, 2019)
- Any at-issueness based approach would have to be strictly Pottsian.



Prosodic representations and discounting

Implicit Prosody Hypothesis: In silent reading, a default prosodic contour is projected onto the stimulus [which can then affect online comprehension].

RRC: Do the trains which leave from Amsterdam arrive on time? **ARC:** Do the trains, [which leave from Amsterdam,], arrive on time?

(Dehé, 2009; Astruc-Aguilera & Nolan, 2007; but cf. Hirschberg & Avesani, 1997; Watson & Gibson, 2004)

But what prosodic representations are projected for DD and because?

(Fodor, 2002; see Frazier et al. 2006)







In sum...

Pragmatic Discounting, as an encapsulation phenomenon, could explain several reliable patterns of reduced downstream influence for appositives.

But the two most straightforward ways to specify Pragmatic Discounting so that it can be tested beyond appositives make predictions we find incorrect.

An expanded empirical understanding of discounting effects gives us a better sense of what other kinds of theories could still make the right cut.



Happy to share more information and answer questions about:

- Items, procedure and modeling details for these three experiments
- Diagnostics and assumptions re: speech acts and discourse units
- Results of an experiment using the filler-gap manipulations from Dillon et al. (2017)
- Work probing the status of appositives in memory (w/ Lalitha Balachandran)





Thanks! Special thanks to Lalitha Balachandran, Margaret Kroll, Brian Dillon, Matt Wagers, Sandy Chung, audiences at CUNY 2020, and various anonymous reviewers.







Appendix A: Items



More sample items from E1

- (i) That girl, the one at the big party John attended at the Fine Arts Center, ended up dating Phil's brother.
- (ii) That chemist, the one in the local lab Trisha worked in last semester, decided to hire an English major.
- (iii) That butcher, the one in the busy shop Amy visited on Third Avenue, bought his meat from local farmers.
- (iv) That pilot, the one from the budget airline Liz flew to Atlanta
 on, quit to become an organic dairy farmer.
- (v) That client, the one beside the hairdresser Maria frequented last year, now lives near the Liberty Bell.
- (vi) That supervisor, the one from the construction company Jason hired at the last minute, pays a fair wage.
- (vii) That teacher, the one at the library Holly criticized in the newspaper, was always kind to children.



More sample items from E2

- (i) Doreen said, "The party John attended at the Fine Arts Center was for my sister."
- (ii) Steve said, "The lab Trisha worked at in 2017 found exciting new results."
- (iii) Larry said, "The shop Amy visited on Third Avenue sold me thirty jars of coconut oil."
- (iv) Theo said, "The airport Allie departed from last week lowered its long-term parking costs."
- (v) Andrew said, "The river Kelsey skated on in January runs beside
 my parents' home."
- (vi) Lucy said, "The construction company Anthony hired for the new development pays its employees a fair wage."
- (vii) Lydia said, "The library Hans criticized in the newspaper means the world to me."

37

More sample items from E3

- (i) Doreen always buys the local arts paper because clever columnists review the concerts John attends at the Fine Arts Center.
- (ii) Steve usually praises the university because politicians visit the lab Trisha works at on the weekends.
- (iii) Larry rarely buys in bulk because health inspectors shut down the budget grocery store Amy visits on Third Avenue.
- (iv) Theo often considers spontaneous vacations because the train goes
 by the airport Allie departs from every week.
- (v) Andrew always worries about climate change because cracks form in the pond Kelsey skates on with her mother.
- (vi) Lucy usually records her podcast on Sundays because the contractor gives a day off to the crew Anthony hired for the remodel.
- (vii) Lydia rarely brings her kids downtown because protesters picket the library Hans criticized in the newspaper.



Appendix B: Analysis Details





Dillon et al. (2014) meta-analysis





Analysis: Model specifications

- Regularizing ("weakly informative") \bullet priors set in brms
 - Thresholds: $\mathcal{N}(0,5)$
 - Fixed effects: $\mathcal{N}(0, 1)$
- STAN meta-parameters:
 - 6 chains of 10,000 iterations each, including 2,000 iterations of warmup
 - Parameters initiate at 0



Expected response distributions given regularizing priors.





Analysis: Empirical priors for BF calculations

Effect	Posterior $\hat{\beta}$	Posterior σ_{β}
Threshold 1 2	-3.15	0.32
Threshold 2 3	-2.52	0.21
Threshold 3 4	-1.95	0.19
Threshold 4 5	-1.36	0.16
Threshold 5 6	-0.44	0.18
Threshold 6 7	0.82	0.36
Structure (ARC)	0.22	0.26
Complexity (Long)	-0.36	0.08
Structure \times Complexity	0.15	0.06

Posteriors from Dillon et al. (2014) meta-analysis used as empirical priors.



Expected response distributions given empirical priors.



Analysis: E1 Model

Effect	Posterior $\hat{\beta}$	Posterior σ_{β}	95% CRI Lower	95% CRI Upper
Threshold 1 2	-3.19	0.18	-3.54	-2.84
Threshold 2 3	-2.38	0.15	-2.68	-2.08
Threshold 3 4	-1.85	0.15	-2.14	-1.57
Threshold 4 5	-1.20	0.14	-1.49	-0.93
Threshold 5 6	-0.24	0.14	-0.52	0.03
Threshold 6 7	0.93	0.14	0.65	1.21
Structure (ARC)	0.27	0.08	0.11	0.42
Complexity (Long)	-0.24	0.06	-0.36	-0.13
Structure \times Complexity	0.08	0.04	0.01	0.16



Analysis: E2 Model

Effect	Posterior $\hat{\beta}$	Posterior σ_{β}	95% CRI Lower	95% CRI Upper
Threshold 1 2	-4.06	0.27	-4.60	-3.55
Threshold 23	-3.24	0.21	-3.65	-2.83
Threshold 34	-2.70	0.20	-3.09	-2.31
Threshold 4 5	-2.21	0.19	-2.59	-1.83
Threshold 56	-1.49	0.19	-1.86	-1.12
Threshold 6 7	-0.31	0.18	-0.67	0.06
Structure (DD)	0.10	0.06	-0.01	0.22
Complexity (Long)	-0.51	0.08	-0.68	-0.36
Structure \times Complexity	0.03	0.04	-0.05	0.11



Analysis: E3 Model

Effect	Posterior $\hat{\beta}$	Posterior σ_{β}	95% CRI Lower	95% CRI Upper
Threshold 1 2	-2.67	0.15	-2.97	-2.38
Threshold 2 3	-1.72	0.14	-1.98	-1.45
Threshold 3 4	-1.18	0.13	-1.44	-0.92
Threshold 4 5	-0.58	0.13	-0.83	-0.33
Threshold 5 6	0.13	0.13	-0.12	0.39
Threshold 6 7	1.12	0.13	0.87	1.38
Structure (Because)	-0.06	0.05	-0.16	0.03
Complexity (Long)	-0.53	0.05	-0.63	-0.43
Structure \times Complexity	0.02	0.04	-0.05	0.10



Appendix C: Effects of adjunct scope in E3?



Ambiguous attachment for because

Evan often complains to the traveletion behind schedule.

- Target interpretation: Evan's complaints have a high frequency because the cruises are systematically late. (*Because > Often*; "high-scope")
- Alternative interpretation: Many of Evan's complaints are because of cruise lateness. (Often > Because; "low-scope")
- Possible confound: If *because* adjuncts are only discourse units on a highscope reading, low-scope readings could be obscuring discounting in E3.

Evan often complains to the travel agent because the cruises depart



Reply #1: Low scope is dispreferred

behind schedule.

- - General bias for high attachment.
 - An exception to Late Closure!
 - See also Hemforth & Konieczny (2004), Koizumi (2009)

Evan often complains to the travel agent because the cruises depart

• Frazier & Clifton (1996): Participants in reading time experiments had trouble with because adjuncts with content that best fit with a low-scope reading.



Reply #2: We made low scope implausible

Elizabeth usually feels melancholy in Florida because teenagers litter on the beaches.

- teenagers are habitually messy.
- by teenagers being messy.
- •

• High-scope: Elizabeth's Floridian melancholy has a high frequency because

• Low-scope: Many of Elizabeth's moments of Floridian melancholy are onset

Intuition: The habitual predicates in the adjuncts are going to be hard to discretize into individual causing events to support a low-scope reading.



Reply #3: More high-scope ≠ more discounting

Linda rarely purchases European makeup because domestic products are available at the fancy shop.

0.6 0.4 0.2 0.0 Ŷ of **Proportion** 0.4 · 0.2 ·

0.0

Caroline always eats dinner in the kitchen because the others turn on the cop show.

Besponses 0.4 0.2 0.0 of **Proportion** 0.4 -0.2 -0.0 -

Naturalness Rating Distributions by Complexity and Structure



















Appendix D: Discounting and filler-gap resolution





The butcher asked...

-Filler

if the lady who said that she who the lady who said that she would like a nice big ham would like a nice big ham ID was cooking for a party. was cooking for ____.

who the lady who said, "I if the lady who said, "I would like a nice big ham," would like a nice big ham," DD was cooking for a party. was cooking for ____.

+Filler

Same procedure as E1: naturalness ratings, 32 items, n = 48 Prolific participants





E2': Results

- Penalty for +Filler sentences: $\hat{\beta} = -0.41, _{95\%}(-0.55, -0.26)$
- DD rated more natural: $\hat{\beta} = 0.19, _{95\%}(0.13, 0.26)$
- Qualified by an interaction, smaller length penalty for ARCs: $\hat{\beta} = -0.05, \, _{95\%}(-0.12, \, 0.01)$
- $BF_{10} = 0.12$, indicating moderate evidence for a absence of a discounting interaction of the expected size

Models fit for BF analysis used empirical priors derived from Dillon et al. (2017).





Naturalness Rating Distributions by Dependency Presence and Structure













Appendix E: Appositives in memory (Balachandran, Duff, Anand, & Rysling @ AMLaP 2022 next week)











Discounting and memory

- Two ways of caching out the mechanism behind discounting:
 - Compression: ARCs are rendered globally inaccessible at right edge.
 - Partition: ARCs induce new structure to memory representations in a way that allows both ARC and matrix content to be accessed to the exclusion of the other.
- We provide evidence that Compression can't be true, and Partition is probably true. \bullet
 - Recognition memory (E1) reveals that the syntactic form of ARCs is no worse remembered than the form of RRCs.
 - Nominal ellipsis resolution (E2) reveals that online retrieval of ARC content is not harder than RRC content, and online retrieval of matrix content is easier in sentences with an ARC.





E1: Recognition memory items, procedure

ARC

Same	The father, who cooked the kids a meal after the orchestra performance, was grateful for instant noodles.	The father that cooked the kids a meal after the orchestra performance was grateful for instant noodles.
Different	The father, who cooked a meal for the kids after the orchestra performance, was grateful for instant noodles.	The father that cooked a meal for the kids after the orchestra performance was grateful for instant noodles.

- 48 participants on Prolific, 48 items with following RM prompts
- Mixed dative vs. double object stimuli

RRC



E1: Recognition memory results

	d_a	AUC	2.5%	97.5%
ARC Same	0.84	0.67	0.64	0.7
RRC Same	0.64	0.63	0.6	0.66
	$D_{boot} = 1.76$		p = 0.08	

- **X** Compression hypothesis: Sensitivity to ARCs is not lower than sensitivity to RRCs.
- **? Partition** hypothesis: Numerically higher sensitivity to ARCs.







E2: Ellipsis resolution items

Control	The struggling author that pul that published forty over th
ARC-1	The struggling author, who pu that published forty over th
ARC-2	The struggling author that pub who published forty over th

- 72 participants on Prolific, 36 items presented in an A-Maze task
- Regions of interest: Immediately following ellipsis site, two word spillover

- blished **two novels** resented the successful hack he past three decades.
- Iblished **two novels**, resented the successful hack he past three decades.
- blished **two novels** resented the successful hack, he past three decades.



E2: Ellipsis resolution results

Linear m/e model for $\ln(RT)$:

Critical (over)	\hat{eta}	95% Crl
C vs. ARC-1	-0.01	(-0.05,0.02)
C, ARC-1 vs. ARC-2	-0.03	(-0.08,0.02)
Spillover (the past)	\hat{eta}	95% Crl
C vs. ARC-1	-0.01	(-0.02, 0.01)
*C, ARC-1 vs. ARC-2	-0.02	(-0.05,-0.000

- ✓ Partition: Retrieval in ARC-2 is faster than in Control.
- **? Compression**: Numerically, retrieval in ARC-1 is faster than in the Control but also slower than in ARC-2.

