Article:

Bannard, C., Rosner, M. and Matthews, D. orcid.org/0000-0003-3562-9549 (2017) What's worth talking about? Information theory reveals how children balance informativeness and ease of production. Psychological Science, 28 (7). pp. 954-966. ISSN 0956-7976

https://doi.org/10.1177/0956797617699848

In press: *Psychological Science.*

What's worth talking about? Information theory reveals how children balance informativeness and ease of production.

Colin Bannard[1], Marla Rosner[2] & Danielle Matthews[3]

[1] Department of Psychological Sciences, University of Liverpool

[2] Department of Linguistics, University of Texas at Austin

[3] Department of Psychology, University of Sheffield

Author note:

This research was supported in part by British Academy grant number SG131952. The authors would like to thank Derya Kadipasaoglu for drawing the stimuli, Cate Schrim, Paige Hartman, Juliette Seive, Amanda Meeks, Amanda Lucio, Havilland Voss and Gabriella Sepulveda for help with data collection and coding and Kirsten Abbot-Smith, Ben Ambridge, Dan Carroll, Franklin Chang, Cat Davies, Dagmar Divjak, Julian Pine, and Gemma Stephens for helpful comments on the manuscript.

The authors declare no conflicts of interest with respect to the authorship or the publication of this article.

Correspondence concerning this article should be addressed to Colin Bannard, Department of Psychological Sciences, Eleanor Rathbone Building, University of Liverpool, Liverpool, L69 7ZA

Contact: colin.bannard@liverpool.ac.uk

Abstract

Of all the things we could say, what determines what is worth saying? Greenfield's principle of informativeness states that, right from the onset of language, humans selectively comment on whatever they find unexpected. We quantify this tendency using information theoretic measures, and test the counterintuitive prediction that children will produce words that are low frequency given the context because these will be most informative. Using corpora of child directed speech, we identified adjectives that varied in how informative (i.e., unexpected) they were given the noun they modified. Three-year-olds (N=31, replication N=13) heard an experimenter use these adjectives to describe pictures. The children's task was then to describe the pictures to another person. As the information content of the experimenter's adjective increased, so did children's tendency to comment on the feature that adjective had encoded. Furthermore, our analyses suggest that children balance this informativeness with a competing drive to ease production.

Keywords: Information theory, pragmatics, child language, language production.

What's worth talking about? Information theory reveals how children balance informativeness and ease of production

How do we decide what to talk about? It is generally accepted that when we speak, we try to be informative (Grice, 1975) but pinning down how we achieve this has proved a challenge. What empirical work there is tends to test cases where information is needed for disambiguation (e.g., a speaker comments on a feature of an object in order to identify which of two alternatives is intended). Yet, in everyday conversation, our goals are not always so constrained. Often we have free range of what to comment on, if anything. Of all the things we could say in a given moment, then, what determines what is worth saying?

Greenfield's principle of informativeness (Greenfield, 1979; Greenfield & Smith, 1976), proposes that, right from the onset of language, infants choose to comment on things they find unexpected or uncertain and leave unmentioned whatever is constant or can be assumed. Greenfield suggested that this behaviour might be captured by the concept of information provided by the mathematical theory of communication (Shannon, 1948). In this case, a message provides information to the extent that it is unpredictable given what is already known. However, this early sketch of how to quantify informativeness was abandoned following a critique by Pea (1979). In this study, we demonstrate that an information theoretic approach is viable. Moreover, adopting it brings to light a trade-off between informativeness (which requires using unlikely forms) and ease of production (which favours using likely – i.e., frequent - forms). We show that, even from 3 years of age, while children find it easier to produce frequent forms, they nonetheless make the effort to talk about the unexpected.

Greenfield's Principle of Informativeness

Greenfield & Smith (1976) analysed naturalistic recordings of children at the one-word stage and observed cases where a child had the choice of saying one of two words to talk about an event (e.g., the words *skate* and *on* to talk about putting skates on). They explained why one or other word was said at any given moment using the concept of uncertainty. For example, if the object was out of the child's possession, then it became uncertain and the word referring to it (*skate*) was likely to be produced. If the child had the object, it became certain and they would express something else, such as a desired change of state (*on*). On the basis of these observations, Greenfield and Smith argued for 'certainty-uncertainty as the perceptual-cognitive basis for the distinction between presupposition and assertion in language' (pp. 186).

To test this claim experimentally, Greenfield and Zukow (1978) had parents perform sets of actions and describe them as they did so. For example, a parent might hand their child Mommy's shoe, then Cathy's shoe, then Alice's shoe such that the object was constant but the possessor varied. For each case like this, the authors derived rules (e.g., 'when the object is given but the possessor changes, comment on the possessor') and children followed these at above-chance rates. Other studies using this contrastive method found similar results with children able to produce multi-word speech (e.g., Baker & Greenfield, 1988; Greenfield & Dent, 1982; O'Neill & Happé, 2000; Salomo, Lieven, & Tomasello, 2009).

Greenfield & Smith (1976) originally proposed that these findings could all be explained by appeal to information theory, which connects uncertainty and communication via probability theory (Shannon, 1948). However, this approach, where a message provides information to the extent that it is unpredictable, did not bear fruit at the time as it was subject to harsh critique.

Pea's Critique

Pea (1979) objected to using information theory to explain word choice for two main reasons. First, he questioned whether equating unexpectedness (low probability) with informativeness would mean that we should talk in nonsensical utterances all the time. Were unexpected speech per se our goal, this would be the case. However, Greenfield's claim was that we aim to comment on the unexpected rather than simply make unexpected comments (Greenfield, 1980). Moreover, this drive to be informative can only ever be one of a number of pressures on speech production, including the need to be conventional (Clark, 2007), to reduce effort (Zipf, 1949) and, critically, to ground what is being said in a topic of conversation (E. Bates, 1976).

Pea's second concern was that the informativeness observed by Greenfield was context-specific while, in its standard formulation, information theory is concerned with the probability of events independent of the situation in which they occur (e.g., the probability of a symbol being sent over a communication channel without reference to the on-going discourse). In Greenfield's studies, the likelihood of a word being used is defined with respect to specific situations (e.g., saying *shoe* in the above scenario becomes less likely as the contrast set is established). It is well recognised that such contrastive language forms a special case - often marked with distinctive prosody (Chafe, 1974 pp. 117-118). Nonetheless, we argue that the principle of informativeness should extend to language production in general and that it should be quantifiable. Context-specificity should not be a barrier to quantification for two reasons.

First, the probabilities used in information theory need not be context free - one can assign conditional likelihoods to events in context. Recent studies suggest children make use of such situation-specific conditional probabilities in language comprehension and information seeking (Frank & Goodman, 2014; Nelson, Divjak, Gudmundsdottir, Martignon, & Meder, 2014) and it is plausible that such probabilities would affect language production too.

Second there are many communicative situations in which, while the informativeness of an utterance might in principle depend on a particular context, it can nonetheless be reliably estimated without direct reference to it. For example, if one turns on the radio without any expectations about what will be on, then the probability of hearing any particular word could be well estimated by looking at the relative frequency of all English words across all contexts in corpora of transcribed speech (see Jurafsky, 2003 for evidence that people are sensitive to such language statistics and, by extension, the real world events that generate them). By looking at corpora of child directed speech, we can obtain such estimates of children's expectations.

In the current study, we took adjective+noun combinations as a test case, and quantified how informative children should find different adjectives given the nouns they modified. We estimated the frequency of each adjective+noun phrase and of the noun alone using a corpus of child-directed speech, then calculated the information content of the adjective as follows:

$$IC = -\log_2 f(\text{ADJECTIVE} + \text{NOUN})/f(\text{NOUN}) \quad (1)$$

The question of interest was whether children would find relatively more informative adjectives more worthy of mention. The way we approached this experimentally was to have children hear an adult describe pictures using adjective-noun combinations of varying information content (e.g., 'pretty dress'). We then invited children to describe the pictures to another person. The pictures depicted a single object such that any adjectives served a descriptive function rather than a contrastive one (Karmiloff-Smith, 1979) and were thus optional. That is, producing a noun alone would have been perfectly felicitous. The question was whether hearing an adult's informative adjective would make children attend to the feature it encoded and later deem it more worthy of comment.

Information content and ease of production

On the above definition of information content, children should want to say infrequent things. This is counterintuitive given that, generally speaking, the more frequent a word or phrase is, the easier children find it to say (Ambridge, Kidd, Rowland, & Theakston, 2015). Thus, if speakers try to minimise the effort of producing utterances, they should say frequent (i.e., uninformative) things. This tension was captured theoretically by Zipf (1949) in his *Principle of Human Least Effort*. Zipf proposed that human languages are subject to contradictory forces of *speaker economy* – the need to minimise the effort of production - and *auditor economy* – the need to make the message useful to the listener. In testing children's tendency to be informative, then, we also assessed the impact of the known tendency to produce easier forms.

We predicted that, where the experimenter produced a highly informative adjective (i.e., where the adjective was infrequent given the noun), children would be more likely to themselves produce an adjective. If the child did so, we also considered whether they used the same one as the experimenter or another form (e.g., '*muddy shoes'* instead of '*dirty shoes'*). Using the same adjective should be easier if it is frequent given the noun (Bannard & Matthews, 2008) and this should act as a counter pressure in production.

Finally, since children's descriptions can be affected by their addressee's knowledge state from around three years of age (Menig-Peterson, 1975; Perner & Leekham, 1986; Saylor, Baird, & Gallerani, 2006), we explored whether this might affect children's production. Thus, in Experiment 1, half the children talked to an addressee who could see the pictures, and half to an addressee who could not. Since this distinction did not affect children's performance, this variable was removed in replication Experiment 2.

Experiment 1

Method

This experiment and the following replication were approved by the Institutional Review Board of the University of Texas at Austin and were carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki.

Participants. Thirty-one typically developing, English-speaking 3-year-olds (mean age: 41 months; Range: 36-47 months) were included. Their families were residents of Austin, Texas, primarily middle-class and European-American. Nine additional children were tested but excluded (1 due to experimenter error, 8 due to producing codable responses on fewer than 50% of trials). A sample size of 32 was determined in advance according to 1) the number of participants available within an academic year and 2) the aim of testing in multiples of 4 for counterbalancing.

Materials. Eight English nouns were paired with two English adjectives each to form phrases found in all mothers' and/or fathers' speech in the English language transcripts in the CHILDES database (MacWhinney, 2000). Frequencies were taken from the combined set of all transcripts. Phrases were chosen with the objective of a) spanning a range of noun, adjective and phrase frequencies, and b) minimizing any correlation between the three frequency types. This was important in order to be able decouple any effect of word frequency from that of conditional probability and by extension conditional information content. The phrases used are listed in table 1. There was no significant correlation between any of the three frequencies. Nor was there any correlation between adjective information content and log adjective frequency ($r(14) = -0.06$, $p = .819$), adjective length in syllables ($r(14) = .28$, $p = .287$) or with syllable frequency based on syllabification from CELEX (Baayen, Piepenbrock, & Gulikers, 1995) and frequencies from the CHILDES corpus described above (log frequency of first syllable: $r(14) = .06$, $p = .833$; log mean syllable frequency: $r(14) = .25$, $p = .344$). The stimuli were split into two sets so that each child encountered each noun only once. For each phrase, a different picture was created to visually represent it (see supplementary materials).

Stimuli were checked after data collection for the experimenter's prosody to check if it differed as a function of adjective information content. Following the procedure reported by Kaland, Krahmer & Swerts (2014), we extracted the maximum F0 (Hz) from all adjectives and nouns for all participants using the autocorrelation method implemented in Praat (Boersma & Weenink, 2016) and converted these to Equivalent Rectangular Bandwidth (ERB) values. We subtracted the pitch value (ERB) of the noun from the pitch value of the adjective (ERB). We then built a multi-level logistic regression model predicting this relative measure, with adjective information content included as a fixed effect, and considering all possible random effects structures. Including adjective information content did not give a significant improvement over a null model, regardless of the random effects structure employed.

Table 1.

Stimuli and associated statistics for Experiment 1 (corpus contains 9.1M words)

Phrase	Set	Adjective information content (bits)	Phrase frequency	Adjective frequency	Noun frequency	Frequency of first (and second) syllable
stinky baby	1	11.99	3	191	12234	1090 (11472)
silly baby	2	9.49	17	4242	12234	15096 (87552)
pretty dress	1	4.57	61	4144	1446	7290 (26044)
little dress	2	7.18	10	27845	1446	87552 (35656)
straight line	1	4.06	52	964	868	970
long line	2	5.24	23	3996	868	5462
young man	1	5.86	86	379	5000	459
mean man	2	11.29	2	5538	5000	5595
funny pajamas	1	9.64	1	4519	797	4962 (38395)
cuddly pajamas	2	8.05	3	66	797	66 (87552)
tiny road	1	8.66	3	1040	1217	6913 (38395)
bumpy road	2	5.29	31	169	1217	536 (8898)
huge tower	1	10.16	1	276	1147	277
tall tower	2	4.58	48	773	1147	774
kind woman	1	8.32	1	3583	320	3603
old woman	2	3.46	29	2739	320	2797

Procedure. Upon arrival, each child played with two experimenters. Once they were at ease, the child was asked to sit down at a small table in the centre of the room, with E1 sitting down across from them. E2 announced that she had work to do and that she would have to come back later and left the room. E1 then pulled out some cards and informed the child that, for this game, she had some pictures for them to look at and that they would "say what we see." Following this, E1 would go through one of the sets of eight images (either set 1 or set 2, counterbalanced), holding each image up in turn and labelling it with the associated adjective-noun phrase. If the child attempted to repeat the phrases at this point or to talk over E1, they would be asked to "look and listen" until E1's turn was over. When this was finished, E2 would re-enter the room, at which point E1 would invite E2 to join the game and then prompt the child to "tell E2 about the pictures." This occurred under one of two conditions (16 participants in the first condition, 15 in the second, randomly assigned):

Picture-visible-to-addressee condition. E2 would respond to the invitation to join in by taking a seat at the table next to the child, from which they could clearly see each image along with the child.

Picture-hidden-from-addressee condition. E2 would respond to the invitation to join in by walking over to a laptop previously hidden under a blanket behind E1, suggesting that she could use it to write down the child's descriptions of each image. She would then state that the laptop seemed to be too heavy to move and that she would have to sit down where she was—a location from which she was clearly unable to see the images herself.

In both conditions E1 would then hold up each picture in turn for the child to label. If the child did not offer a response for any given picture, E1 would prompt them again to "tell E2 what you see in the picture". If after an interval of time the child was still unable to respond, E1 would move on to the next picture.

The order in which the pictures were presented was identical for the experimenter's and the children's production. The order of presentation was randomized for each child. Due to experimenter error, two children heard the same order.

Transcription and Coding. The recordings were coded by the second author. Each utterance was coded for a) whether the child had produced an adjective, and b) whether that adjective was identical to the adjective used by E1. Rare cases where children produced alternative or erroneous grammatical forms of modification (e.g., saying 'bumps road' instead of 'bumpy road') were included since the question of interest was whether the feature that the experimenter's adjective encoded was commented on (not grammatical form). Data from 30 participants was reliability coded by a research assistant who was blind to the purpose of the study and agreed with the first coder on 94% of responses for the first coding decision (i.e., whether an adjective was produced, $\kappa = 0.852$) and 93% of responses for the second coding decision (i.e., whether it was the same adjective as the experimenter, $\kappa = 0.803$).

Results

Of the 222 responses children made in total, 21% of these contained no adjective, 60% contained the same adjective as the experimenter had used, and 19% contained a different adjective. For each of the 42 responses in which a different adjective was used, all but 6 responses unambiguously referred to the same semantic feature. Given the low number referring to a different feature, this distinction was not included in the following models (i.e., all cases were collapsed together into the category 'different adjective').

We analysed the data using a sequential logistic procedure (also known as a nested dichotomies model, Fox, 1997; Tutz, 1991). We fitted binary logit models to 1) whether or not the child used an adjective, and then 2) whether or not the adjective (if produced) was the same as the experimenter's (see figure 1 for the structure of this procedure). Information content was centered so that its lowest value became zero in order to allow meaningful interpretation of intercepts, but no scaling was performed so as to retain generalizability to other items. As each child participated in multiple trials, and to take account of additional differences between items, we used a multilevel version. Participant, noun and adjective were considered for inclusion as random effects on all appropriate model terms but excluded if they did not improve fit, using the iterative procedure of Bates, Kliegl, Vasishth & Baayen (2015). Based on this selection procedure, a random effect of participant on the intercept for decision one (did the child produce an adjective) and a random effect of noun on the slope for decision 2 (did the child use the same adjective) were included. The fit of the model was not improved by including visual access condition (whether E2 could see the pictures or not; $\chi^2(1)=0.076$, $p = 0.782$), or an interaction between visual access condition and information content ($\chi^2(2)=0.75$, $p = 0.686$) and thus these terms were not included in reported models.

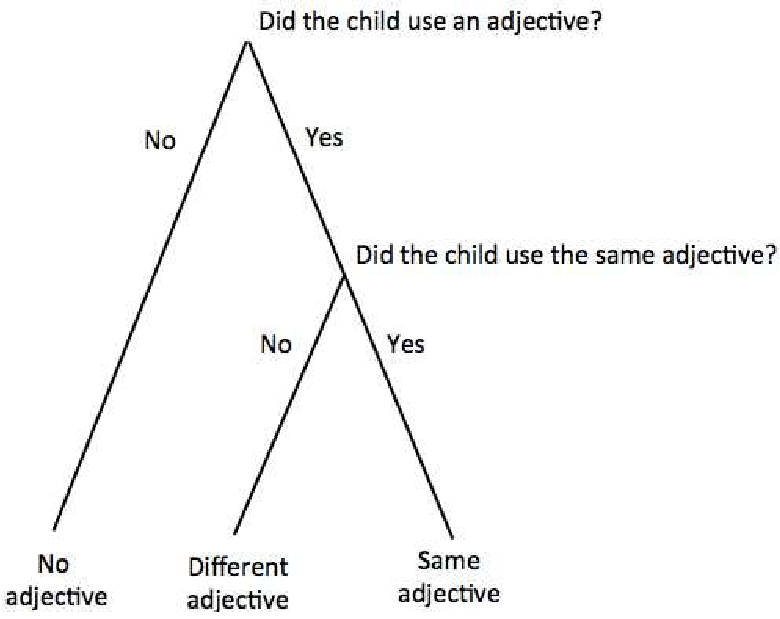


Figure 1: Sequential logistic model with two stages.

The final models were fitted with Bayesian Markov Chain Monte Carlo methods using the JAGS software (Plummer, 2003). Along with estimates, 90% credible intervals, standard deviations and one-tailed p-values, order-restricted Bayes factors are reported for the parameters of interest (the effect of adjective informativeness on both decisions). These tell us the ratio of the likelihood of the hypothesis (that the true value of the slope parameter differs from zero) and the null hypothesis (in all cases here that the true value of the slope parameter is zero) given the data. We calculated these using the Savage-Dickey method (Wagenmakers, Lodewyckx, Kuriyal, & Grasman, 2010). This requires setting an informative range for the prior on a parameter of interest, while allowing diffuse priors for all other terms. We assume that for any single decision, the movement in odds as a function of informativeness will most likely fall below 99 which is equivalent to a change between from 1% of participants and half the participants (or half the participants and 99% of participants) producing the adjective, and thus use a normal distribution with a mean of zero and a standard deviation of (log 99) 4.6, scaled to the range of the information content predictor.

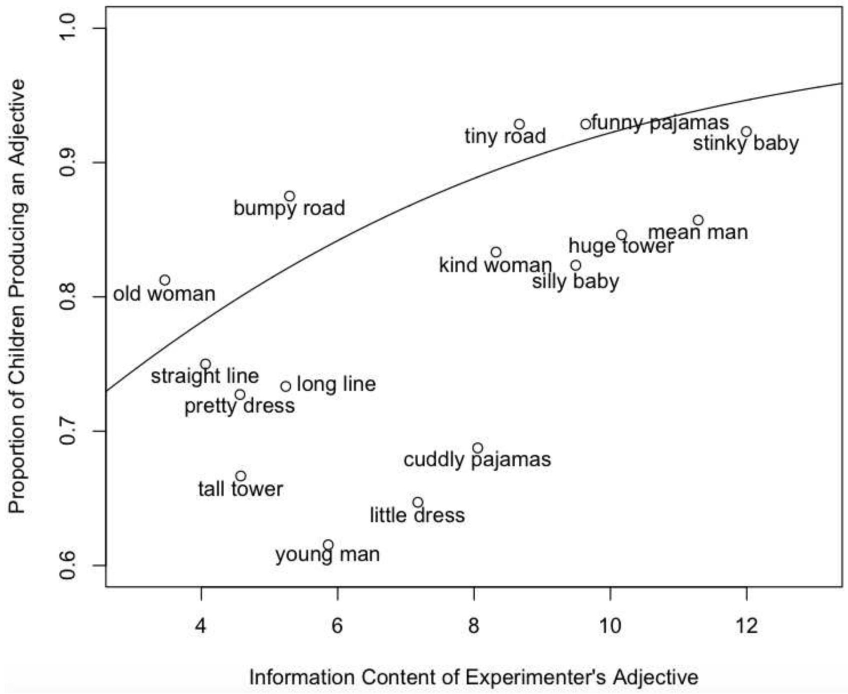
The inferred parameters for the sequential logistic model are shown in table 2. The effects are visualized in figure 2. The lines represent the fitted models with intercepts adjusted to the participant harmonic mean where relevant random terms were included. Figure 2a plots the proportion of children who produced an adjective as a function of information. For those cases in which an adjective was produced, figure 2b shows whether or not it was the same adjective as the experimenter's. The likelihood of a child producing an adjective increases by 22% for each additional bit of information given by the experimenter's adjective. That is, as the information content of the experimenter's adjective increased, so did the tendency for the children to produce an adjective. The likelihood of a child who produces an adjective producing the same adjective as the experimenter, in contrast, *decreases* by 31% for each bit of information it provides. That is, as the information content of the experimenter's adjective increased (i.e., the frequency of the adjective-noun combination decreased) children were less likely to use the same adjective as the experimenter and more likely to use a different adjective that meant the same thing. The Bayes factors tell us that, according to the standard interpretative scale of Kass and Rafferty (1995), there is "positive" evidence in favour of both of these effects.

Table 2.

Summary of fixed effects for a) the proportion of participants who produced an adjective (adjective coded as 1, no adjective coded as 0); b) the proportion of the participants who produced the same adjective as the experimenter (for cases in which an adjective was produced; same adjective coded as 1, different adjective coded as 0).

	Did the child use an adjective?				Was it the same adjective as E1?			
	β	SD	*P*	Bayes Factor	*β*	SD	*p*	Bayes Factor
Intercept	1.79 (0.67 - 3.10)	0.75	.003		2.35 (1.64 - 3.12)	0.45	<.001	
Information content	0.20 (0.05 - 0.36)	0.09	.013	3.64	-0.28 (-0.54 - -0.04)	0.16	.038	3.19

a)



b)

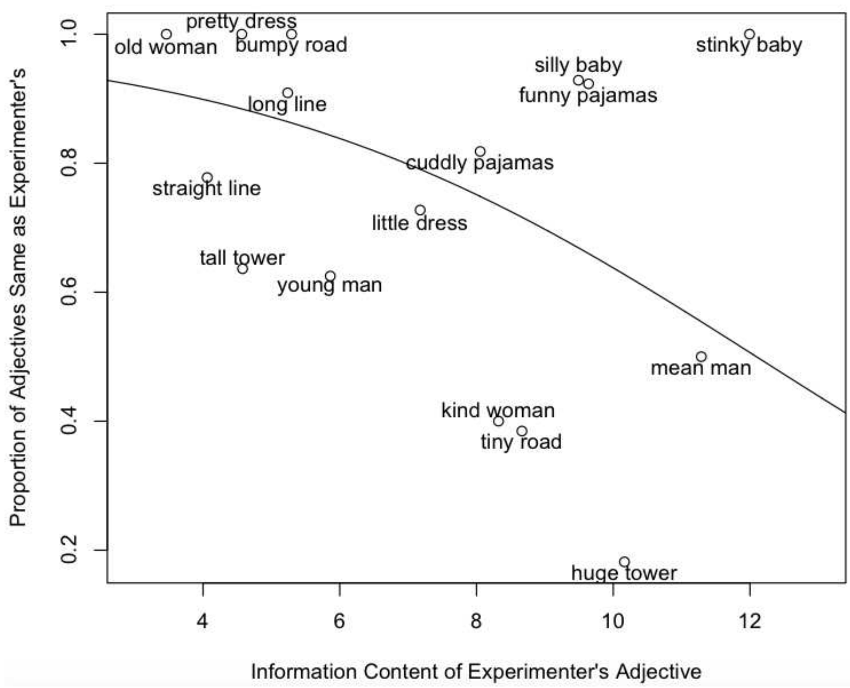


Figure 2. The relation (in Experiment 1) between adjective informativeness and a) the proportion of participants who produced an adjective; b) the proportion of the participants who produced the same adjective as the experimenter (for cases in which an adjective was produced).

The statistics (*p*-value, Bayes Factor) in table 2 concern whether the slopes for each decision differ from zero i.e., whether 1) the adjective information content slope for the decision to produce an adjective is significantly above zero and 2) whether the adjective information content slope for the decision to produce the same adjective as the experimenter is significantly below zero. It is also possible to ask whether the slopes differ from each other – i.e., whether the effect of adjective information content is different for the two decisions. While it follows from the above analysis that this is the case, for completeness – and to directly quantify the weight of evidence for the hypothesis that informativeness is in tension with ease of production - we built a model including a shared slope term and a same-adjective decision specific term (with doubled standard deviation on the prior to allow the slopes to vary in direction) and tested whether the value of this second term differed from zero. We observe a positive shared slope (mean estimate = .20, SD =.09, p = .016) with a negative same-adjective decision specific term (mean estimate = -.49, SD = .19, p = .008, Bayes Factor = 9.78). This indicates positive evidence that adjective information content has a different relationship to deciding whether to produce an adjective than it does to deciding whether to produce the same adjective as the experimenter.

Discussion

Experiment 1 suggests that information theory can explain what children choose to say. As the information content of the experimenter's adjective increased, so did the tendency for the children to comment on the feature it encoded. However, the likelihood that they would use the exact same adjective to do so decreased, suggesting a counter-pressure from ease of production. In Experiment 2, we tested the robustness of the primary finding via replication with more items, new participants and an improved method (stimulus descriptions played on a computer).

Experiment 2: Replication

Method

Participants. A bootstrap power analysis (Efron & Tibshirani, 1993) using the data and model from Experiment 1 suggested that the primary effect of informativeness (i.e., its impact on whether or not to produce an adjective) seen in Experiment 1 should be apparent with 13 participants. In this replication thirteen typically-developing, English-speaking 3-year-olds (mean age: 39 months; Range: 36-44 months) were included. Their families were residents of Austin, Texas, primarily middle-class and European-American.

Materials. There were 20 items (two adjectives paired with each of 10 nouns, with each child encountering each noun only once). These items were an expanded set of the phrases from Experiment 1. Two item pairs from the original study were excluded because they were either distracting (*stinky baby*) or potentially gender biased in experience (*pretty/little dress*). Items are presented in table 3. There was no significant correlation between any of the noun, adjective and phrase frequencies. Nor was there any correlation between adjective information content and log adjective frequency ($r(18) = .07$, $p = .768$), adjective length in syllables ($r(18) = .32$, $p = .168$) or with syllable frequency (log frequency of first syllable: $r(18) = .21$, $p = .384$; log mean syllable frequency: $r(18) = .23$, $p = .329$).

To control delivery of the items, a research assistant pre-recorded all adjective noun combinations so they could be played over a computer. The prosodic properties of the recorded stimuli were checked following the procedure for study 1. A multi-level logistic regression model predicting the difference between the maximum F0s of the noun and adjective was built. Including adjective information content (with random effects of noun on the intercept) did not give a significant improvement over a null model (NB no random slopes were considered for model identifiability reasons due to there being 20 data points).

Table 3: *Stimuli and associated statistics for Experiment 2 (corpus contains 9.1M words)*

Phrase	Set	Adjective information content (bits)	Phrase frequency	Adjective frequency	Noun frequency	Frequency of first (and second) syllable
messy hair	1	9.39	7	457	4686	6077(15096)
curly hair	2	7.24	31	158	4686	600(87552)
big hat	1	6.6	40	18106	3887	18106
magic hat	2	10.92	2	544	3887	5726(270)
straight line	1	4.06	52	964	868	970
long line	2	5.24	23	3996	868	5462
young man	1	5.86	86	379	5000	459
mean man	2	11.29	2	5538	5000	5595
runny nose	1	5.81	72	112	4052	2551(38395)
round nose	2	9.4	6	5456	4052	9989
funny pajamas	1	9.64	1	4519	797	4962 (38395)
cuddly pajamas	2	8.05	3	66	797	66 (87552)
bumpy road	1	5.29	31	169	1217	536 (8898)
tiny road	2	8.66	3	1040	1217	6913 (38395)
dirty shoes	1	9.23	5	2330	2993	2332(26044)
new shoes	2	5.26	78	3699	2993	3908
huge tower	1	10.16	1	276	1147	277
tall tower	2	4.58	48	773	1147	774
old woman	1	3.46	29	2739	320	2797
kind woman	2	8.32	1	3583	320	3603

To check whether the picture stimuli would elicit mention of the features even in the absence of having heard an adjective, a separate group of 24 children (mean age 55 months; Range: 48 – 65 months) were presented with the stimuli pictures following the same procedure as in study 2. Each participant saw a randomly selected set of stimuli, with each noun occurring only once, appearing in a random order. For every other item the adjective was replaced with a cough such that it sounded as if the experimenter had intended to produce an adjective but it was unclear what this was. In this way, children were primed to produce adjective-noun combinations but on cough + noun trials the only thing that could have influenced adjective production was the image (not the experimenter's choice of modification). When children came to describe pictures following cough + noun trials, they were equally likely to produce a noun as on adjective + noun trials. However, they were less likely to produce an adjective (Adjective + noun stimuli: 72% child adjective provision. Cough + noun stimuli: 14% adjective provision) and less likely to reference the same feature as the experimenter (Adjective + noun stimuli: 68% same feature. Cough + noun stimuli: 10% same feature). Thus children were very unlikely to happen upon mentioning the same feature as the experimenter if they hadn't already heard an adjective. When they did so this was not correlated with the adjective information content of the adjective that had been replaced by the cough. This suggests that children's adjective provision in the main experiments was driven by the descriptions they had heard the experimenter produce.

Procedure. The procedure was the same as in the picture-visible-to-addressee condition from Experiment 1, with the exception that rather than E1 holding up and describing pictures in the first stage, a slide show on a computer screen was employed to play the recorded descriptions using a media enhanced pdf presentation.

Coding. Children's responses were coded (using the same criteria as for Experiment 1) by a research assistant blind to the purpose of the study. Responses from 11 children were reliability coded by a second research assistant (also blind to the hypotheses) who agreed with the first coder on 95% of responses for the first coding decision (i.e., whether an adjective was produced, $\kappa = 0.905$) and 89% of responses for the second coding decision (i.e., whether it was the same adjective as the experimenter, $\kappa = 0.765$).

Results

Of the 119 responses children made in total, 21% of these contained no adjective, 55% contained the same adjective as the experimenter had used, and 24% contained a different adjective. For the 29 responses in which a different adjective was used, 14 unambiguously referred to the same semantic feature as the experimenter's adjective.

The same analyses were performed as for Experiment 1, using the same models, including the same random effects. Bayes factors are a particularly appropriate statistic for evaluating replication as they allow us to assess the evidence in favour of the null hypothesis as well as the hypothesis (Dienes, 2016). As the priors used to calculate these need to be informative, we updated the priors for the slopes to include the new evidence from experiment one. Means of zero were retained but the standard deviations for the prior on these two decisions were set equal to the respective slopes estimated from the Experiment 1 data.

The inferred parameters for the decisions of interest are shown in table 2. The model is visualized in figure 3. The lines represent the fitted models with intercepts adjusted to the participant harmonic mean where relevant random terms were included. Figure 3a plots the proportion of children who produced an adjective as a function of information. For those cases in which an adjective was produced, figure 3b shows whether or not it was the same adjective as the experimenter's. The likelihood of a child producing an adjective increases by 24% for each additional bit of information given by the experimenter's adjective. The Bayes factor of 10.78 tells us there is positive evidence that the probability of children producing an adjective increases with information content.

While this experiment was powered to look for replication of the above effect of information content on adjective production, the model also tells us that the likelihood of a child who produces an adjective producing the same adjective as the experimenter, in contrast, decreases by 8% for each bit of information it provides. The Bayes factor of 0.82 tells us that the data is by itself is inconclusive as to whether adjectives that have lower information content are more likely to be directly reproduced than those who have higher information content (the data by itself provides no clear evidence for or against the proposal).

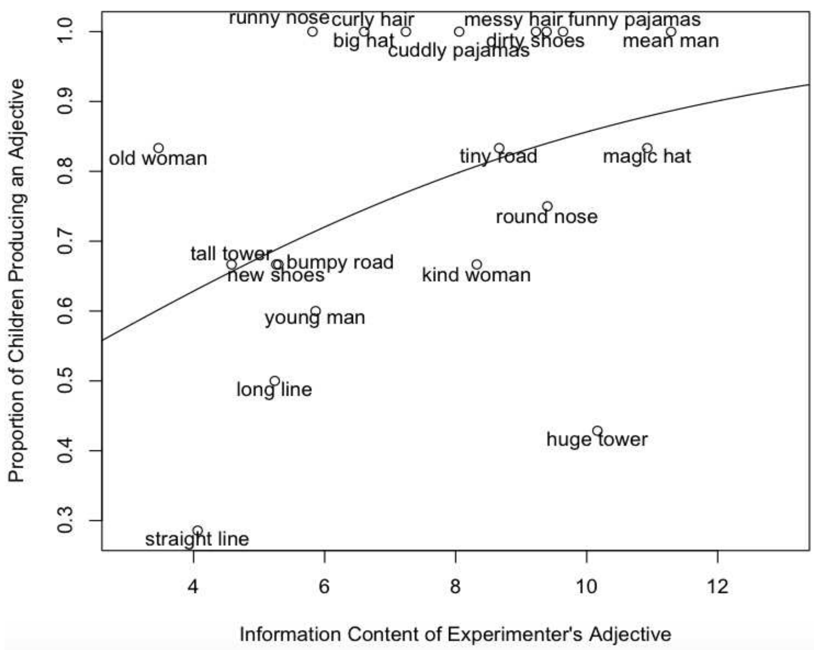
The above Bayes factor concerns whether the slope for the second decision differed from zero. As for Experiment 1, we also tested whether the slope for the second decision differed from the slope of the first decision. We built a model including a shared slope term and a same-adjective decision-specific term (with doubled standard deviation on the prior to allow the slopes to vary in direction). We observe a positive shared slope (mean estimate = .23, SD =.1 , $p < .001$) with a negative same-adjective decision specific term (mean estimate = -.31, SD = .18, p = .036, Bayes factor = 3.18). This is positive evidence that adjective information content has a different relationship to deciding whether to produce an adjective than it does to deciding whether to produce the same adjective as the experimenter.

Table 4.

Summary of fixed effects for a) the proportion of participants who produced an adjective (adjective coded as 1, no adjective coded as 0); b) the proportion of the participants who produced the same adjective as the experimenter (for cases in which an adjective was produced; same adjective coded as 1, different adjective coded as 0) for Experiment 2.

	Did the child use an adjective?				Was it the same adjective as E1?			
	β	SD	*p*	Bayes Factor	*β*	SD	*p*	Bayes Factor
Intercept	0.78 (-0.07-1.70)	0.55	.065		1.16 (0.38-1.99)	0.50	.007	
Information content	0.21 (0.06-0.37)	0.09	.010	10.78	-.08 (-0.30-0.13)	0.14	.266	0.82

a)



b)

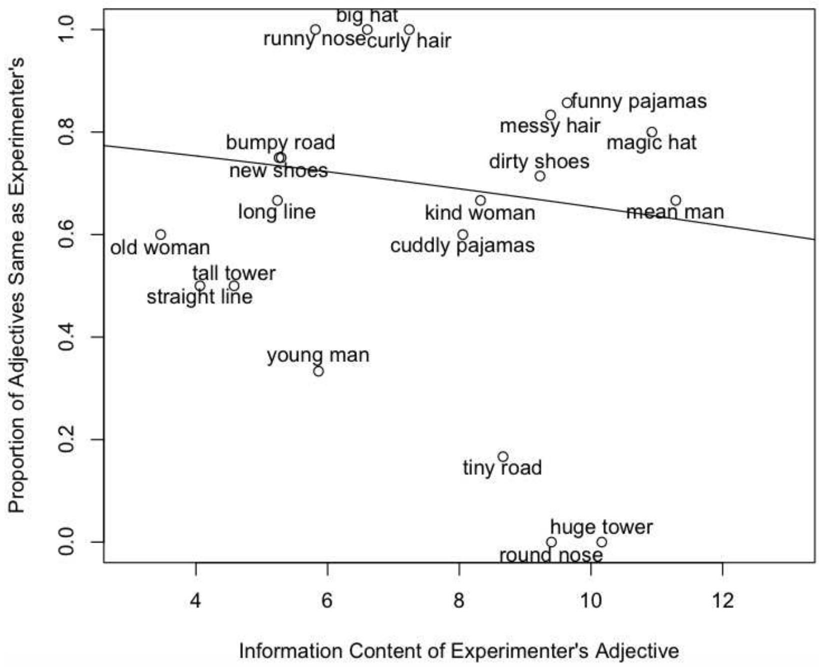


Figure 3. The relation (in Experiment 2) between adjective informativeness and a) the proportion of participants who produced an adjective; b) the proportion of the participants who produced the same adjective as the experimenter (for cases in which an adjective was produced).

Finally, we assessed whether when children were attempting to reference the same feature as the experimenter, they did so by using a less informative (more frequent) adjective given the noun. Across the two experiments, children's alternative adjectives had a mean information content of 7.05, compared to a mean information content of 8.52 for the experimenter's adjectives for the same items. We fitted a multilevel linear effects model with information content as the outcome and speaker (experimenter or child, coded as 0 and 1 respectively) as a predictor. Random effects of participant and target adjective on the intercept were included (the random effects structure was chosen in the same manner as for the main experiments, except that we didn't consider random slopes due to the small sample - only half of the phrases were subject to substitution and for half of these substitution happened only once). To allow estimation of a Bayes factor, an informative prior was used for the speaker parameter. We assumed that the change in information content between the experimenter's adjective and the adjective the child used was likely to be at most the range of the information content of the stimuli, and thus we assigned a normal distribution with a mean of zero and the standard deviation set to equal the range. The parameter for speaker (mean estimate = -1.46, SD = .48, p = .002, Bayes factor = 9.42) supports the hypothesis that when children were attempting to reference the same feature as the experimenter but with other means, they opted to replace the unexpected adjective with more expected words.

General Discussion

In the current experiments, children heard someone give more or less informative descriptions of pictures. They then needed to describe these pictures to another person. As the information content of the adjectives in the initial descriptions increased, so did children's tendency to themselves use an adjective. Thus information theory can explain what children choose to say.

Most responses with an adjective directly reused the experimenter's phrase. While reuse may be a shortcut to production, there was evidence of a counter pressure to taking it. Whereas increasing information content made children more likely to produce an adjective, it bore a different relation to the tendency to produce the same adjective as the experimenter (which decreased in Experiment 1 and tended to do so in Experiment 2). We ascribe this to the difficulty children encounter in producing low frequency word combinations. When children chose alternative means of referencing the same feature as the experimenter, the adjectives they used were more frequent given the noun. This trade-off has potential consequences for children's developing grammatical productivity. If children strive to be informative, they will need to stray off well-known linguistic territory, pushing them to produce novel word combinations and yet this creativity is constrained by ease of production.

Were children striving to be informative for the benefit of their addressee? Children this age are certainly capable of audience design in some cases (e.g., Saylor et al., 2006). However, there was no evidence of adaptation to a specific listener in Experiment 1. We therefore assume that children were adapting to a generic listener (i.e., what O'Neill, 2012 refers to as cognitive - rather than social or mindful - pragmatics). It will be important to determine what underpins this type of pragmatic skill, also observed in adult language production where, for example, rate of information flow is managed right from the phonological to the syntactic level (Aylett & Turk, 2004; Jaeger, 2010). While much research has explored social cognition as an explanatory factor in pragmatic competence, recent theoretical accounts suggest we have underestimated the role of other cognitive processes in supporting interaction

(Apperly, 2010). Work on individual differences could test the prediction that the ability to comment on the unexpected depends on cognitive capacities such as attending to similar features as others and learning about statistical regularities in the environment. To the extent that we track the same probabilities in the world, we find the same things noteworthy and are thus able to have mutually satisfying conversations.

Author Contributions

D. Matthews and C. Bannard developed the study concept. All authors contributed to the study design. Data collection was conducted in part by M. Rosner and supervised by C. Bannard. C. Bannard performed the data analysis. D. Matthews & C. Bannard drafted the manuscript. All authors approved the final version of the manuscript for submission.

References

Ambridge, B., Kidd, E., Rowland, C., & Theakston, A. (2015). The ubiquity of frequency effects in first language acquisition. *Journal of Child Language, 42*(02), 239-273. doi:10.1017/S030500091400049X

Apperly, I. A. (2010). *Mindreaders: the cognitive basis of" theory of mind"*. Hove: Psychology Press.

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and speech, 47*(1), 31-56.

Baayen, R., Piepenbrock, R., & Gulikers, L. (1995). CELEX2 LDC96L14. *Web Download. Philadelphia: Linguistic Data Consortium*.

Baker, N. D., & Greenfield, P. M. (1988). The development of new and old information in young children's early language. *Language Sciences, 10*(1), 3-34. doi:10.1016/0388-0001(88)90003-4

Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations: Research article. *Psychological Science, 19*(3), 241-248. doi:10.1111/j.1467-9280.2008.02075.x

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.

Bates, E. (1976). *Language and Context: The acquisition of pragmatics*. New York: Academic Press.

Boersma, P., & Weenink, D. (2016). Praat: doing phonetics by computer (Version 6.0.19).

Chafe, W. L. (1974). Language and consciousness. *Language*, 111-133. doi:10.2307/412014

Clark, E. V. (2007). Conventionality and contrast in language and language acquisition. *New Directions for Child and Adolescent Development, 115*, 11-23.

Dienes, Z. (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology*.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Boca Raton, FL: CRC Press.

Fox, J. (1997). *Applied regression analysis, linear models, and related methods*: Sage Publications.

Frank, M. C., & Goodman, N. D. (2014). Inferring word meanings by assuming that speakers are informative. *Cognitive psychology, 75*, 80-96. doi:10.1016/j.cogpsych.2014.08.002

Greenfield, P. M. (1979). Informativeness, presupposition and semantic choice in single word utterances. In E. Ochs & B. Schieffelin (Eds.), *Developmental pragmatics.* (pp. 159-166). New York: Academic Press.

Greenfield, P. M. (1980). Going beyond information theory to explain early word choice: a reply to Roy Pea. . *Journal of Child Language, 7*(1), 217-221. doi:10.1017/S0305000900007145

Greenfield, P. M., & Dent, C. H. (1982). Pragmatic factors in children's phrasal coordination. *Journal of Child Language, 9*(2), 425-443. doi:10.1017/S0305000900004797

Greenfield, P. M., & Smith, J. (1976). *The structure of communication in early language development.* New York: Academic Press.

Greenfield, P. M., & Zukow, P. G. (1978). Why Do Children Say What They Say When They Say It? An Experimental Approach to the Psychogenesis of Presupposition. In K. Nelson (Ed.), *Children's Langauge* (Vol. 1). New York: Gardner Press.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41-58). New York: Academic Press.

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive psychology, 61*(1), 23-62.

Jurafsky, D. (2003). Probabilistic modeling in psycholinguistics: Linguistic comprehension and production. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probabilistic Linguistics*. Cambridge, MA: MIT Press.

Kaland, C., Krahmer, E., & Swerts, M. (2014). White Bear Effects in Language Production: Evidence from the Prosodic Realization of Adjectives. *Language and Speech, 57*(4), 470-486. doi:10.1177/0023830913513710

Karmiloff-Smith, A. (1979). *A functional approach to child language.* Cambridge, UK: Cambridge University Press.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*(430), 773-795.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (Vol. 2: The database). London: Lawrence Erlbaum Associates.

Menig-Peterson, C. L. (1975). The modification of communicative behavior in preschool-aged children as a function of the listener's perspective. *Child Development*, 1015-1018. doi:10.2307/1128416

Nelson, J. D., Divjak, B., Gudmundsdottir, G., Martignon, L. F., & Meder, B. (2014). Children's sequential information search is sensitive to environmental probabilities. *Cognition, 130*(1), 74-80. doi:10.1016/j.cognition.2013.09.007

O'Neill, D. K., & Happe, F. G. E. (2000). Noticing and commenting on what's new: differences and similarities among 22-month-old typically developing children, children with Down syndrome and children with autism. *Developmental Science, 3*(4), 457-478. doi:10.1111/1467-7687.00139

O'Neill, D. K. (2012). Components of pragmatic ability and children's language development. In H. J. Schmid (Ed.), *Cognitive Pragmatics.* (pp. 261-288). Berlin: Mouton de Gruyter.

Pea, R. D. (1979). Can information theory explain early word choice. *Journal of child language, 6*(03), 397-410. doi:10.1017/S0305000900002440

Perner, J., & Leekham, S. (1986). Belief and quantity: three year olds' adaptation to listener's knowledge. *Journal of Child Language, 13*(2), 305-316.

Plummer, M. (2003). *JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling.* Paper presented at the Proceedings of the 3rd international workshop on distributed statistical computing.

Salomo, D., Lieven, E., & Tomasello, M. (2009). Young children's sensitivity to new and given information when answering predicate-focus questions. *Applied Psycholinguistics, 31*(01), 101-115. doi:10.1017/S014271640999018X

Saylor, M. M., Baird, J. A., & Gallerani, C. (2006). Telling others what's new: Preschoolers' adherence to the given-new contract. *Journal of Cognition and Development, 7*(3), 341-379. doi:10.1207/s15327647jcd0703_7

Shannon, C. E. (1948). A mathematical theory of communication. In (Vol. 27, pp. 379–423 and 623-656). *Bell System Technical Journal*

Tutz, G. (1991). Sequential models in categorical regression. *Computational Statistics & Data Analysis, 11*(3), 275-295. doi:10.1016/0167-9473(91)90086-H

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive psychology, 60*(3), 158-189.

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Cambridge, MA: Addison-Wesley Press.