Integrating Pragmatic Reasoning in an Efficiency-Based Theory of Utterance Choice

Dissertation
zur Erlangung des akademischen Grades
eines Doktors der Philosophie
der Philosophischen Fakultät
der Universität des Saarlandes

vorgelegt von
Ekaterina Kravtchenko
aus Moskau, Russland

Saarbrücken, 2022

Dekan: Prof. Dr. Augustin Speyer
Erstberichterstatterin: Prof. Dr. Vera Demberg
Zweitberichterstatter: Dr. Heiner Drenhaus
Tag der letzten Prüfungsleistung: 19. Oktober 2021

Abstract

This thesis explores new methods of accounting for discourse-level linguistic phenomena, using computational modeling. When communicating, efficient speakers frequently choose to either omit, or otherwise reduce the length of their utterances wherever possible. Frameworks such as Uniform Information Density (UID) have argued that speakers preferentially reduce or omit those elements that are more predictable in context, and easier to recover.

However, these frameworks have nothing to say about the effects of a linguistic choice on how a message is interpreted. I run 3 experiments which show that while UID posits no specific consequences to being "overinformative" (including more information in an utterance than is necessary), in fact *overinformativeness* can trigger pragmatic inferences which alter comprehenders' background beliefs about the world.

In this case, I show that the Rational Speech Act (RSA) model, which models back-and-forth pragmatic reasoning between speakers and comprehenders, predicts both efficiency-based utterance choices, as well as any consequent change in perceived meaning. I also provide evidence that it's critical to model communication as a *lossy* process (which UID assumes), which allows the RSA model to account for phenomena that it otherwise is not able to.

I further show that while UID predicts increased use of pronouns when referring to more contextually predictable referents, existing research does not unequivocally support this. I run 2 experiments which fail to show evidence that speakers use reduced expressions for predictable elements. In contrast to UID and similar frameworks, the RSA model can straightforwardly predict the results that have been observed to date.

In the end, I argue that the RSA model is a highly attractive alternative for modeling speaker utterance choice at the discourse level. When it reflects communication as a *lossy* process, it is able to predict the same predictability-driven utterance reduction that UID does. However, by additionally modeling back-and-forth pragmatic reasoning, it successfully models utterance choice phenomena that simpler frameworks cannot account for.

Kurzzusammenfassung

Diese Arbeit erforscht neue Methoden, linguistische Phänomene auf Gesprächsebene per Computermodellierung zu erfassen. Effiziente Sprecher:innen entscheiden sich bei der Kommunikation häufig dazu, wenn immer es möglich ist, Äußerungen entweder ganz auszulassen oder aber ihre Länge zu reduzieren. Modelle wie *Uniform Information Density* (UID) argumentieren, dass Sprecher:innen vorzugsweise diejenigen Elemente auslassen, die im jeweiligen Kontext vorhersagbarer und einfacher wiederherzustellen sind.

Allerdings sagen diese Modelle nichts über die Auswirkungen einer linguistischen Entscheidung bezüglich der Interpretation einer Nachricht aus. Ich führe drei Experimente durch, die zeigen, dass wenngleich UID keine spezifischen Auswirkungen von "Überinformation" (einer Äußerung mehr Information als nötig geben) postuliert, Überinformationen doch pragmatische Schlussfolgerungen, die das gedankliche Weltmodell der Versteher:innen ändern können, auslöst.

Für diesen Fall zeige ich, dass das *Rational-Speech-Act*-Modell (RSA), welches pragmatische Hin-und-Her-Schlussfolgerungen zwischen Sprecher:innen und Versteher:innen modelliert, sowohl effizienzbasierte Äußerungsauswahl als auch jegliche resultierende Verständnisänderung vorhersagt. Ich liefere auch Anhaltspunkte dafür, dass es entscheidend ist, Kommunikation als *verlustbehafteten* Prozess zu modellieren (wovon UID ausgeht), was es dem RSA-Modell erlaubt, Phänomene einzubeziehen, wozu es sonst nicht in der Lage wäre.

Weiterhin zeige ich, dass obschon UID beim Bezug auf kontextuell vorhersagbarere Bezugswörter eine erhöhte Nutzung von Pronomen vorhersagt, dies von existierender Forschung nicht einstimmig gestützt wird. Ich führe zwei Experimente durch, die keine Anhaltspunkte dafür, dass Sprecher:innen reduzierte Ausdrücke für vorhersagbare Elemente verwenden, finden. Im Gegensatz zu UID und ähnlichen Modellen kann dass RSA-Modell direkt die bislang beobachteten Resultate vorhersagen.

Schließlich lege ich dar, warum das RSA-Modell eine höchst attraktive Alternative zur Modellierung von Sprachäußerungsentscheidungen auf Gesprächsebene ist. Wenn es Kommunikation als einen *verlustbehafteten* Prozess widerspiegelt, kann es

dieselbe vorhersagebasierte Äußerungsreduktion vorhersagen wie auch UID. Modelliert man jedoch zusätzlich pragmatische Hin-und-Her-Schlussfolgerungen, modelliert RSA erfolgreich Phänomene bei Äußerungsentscheidungen, die einfachere Modelle nicht abbilden können.

Ausführliche Zusammenfassung

In dieser Arbeit untersuche ich alternative Möglichkeiten zu modellieren, wie Sprecher:innen Entscheidungen zwischen bedeutungsäquivalenten Äußerungen treffen und wie diese Äußerungsentscheidungen die Bedeutung beeinflussen, die Versteher:innen vermittelt wird. Ich gehe davon aus, dass rationale Sprecher:innen sprachliche Entscheidungen treffen und ihre Äußerungen so strukturieren, dass sie den Zielen der Vermittlung einer bestimmten Information an eine:n Hörer:in am ehesten entsprechen, wobei sie den relevanten Kontext berücksichtigen, wie etwa kognitive Ressourcenbeschränkungen seitens der Gesprächspartner:innen. Insbesondere können Sprecher:innen versuchen, den Aufwand, den sie für die Übermittlung von Informationen betreiben, möglichst gering zu halten. Insbesondere sollten sie zumindest bis zu einem gewissen Grad versuchen, dies auf eine *optimale* Art und Weise zu tun, indem sie innerhalb der gegebenen Einschränkungen das erfolgreichst mögliche Ergebnis erzielen, das über die einfache Übermittlung der relevanten Informationen hinausgeht.

Ich beginne mit Blick auf eine prominente computergestützte Theorie der Sprachproduktion (*Uniform Information Density*, oder UID), gemäß derer rationale Sprecher:innen bevorzugt solche optionale Elemente in ihrer Sprache auslassen oder kürzen, die mehr vorhersagbare oder wiederherstellbare Informationen kodieren. Ich argumentiere, dass diese Theorie deskriptiv unzureichend ist, wenn es darum geht, sprachliche Entscheidungen zu berücksichtigen, die pragmatische Interpretationen bei dem:der Zuhörer:in auslösen können, und dass sie infolgedessen in einigen Fällen keine korrekten empirischen Vorhersagen macht. Ich schlage dann vor, dass das *Rational-Speech-Act*-Modell (RSA) — ein computergestütztes Modell der zielgerichteten pragmatischen Hin-und-Her-Argumentation zwischen einem:einer Sprecher:in und einem:einer Hörer:in — verwendet werden kann, um mit kleinen Modifikationen diejenigen Phänomene zu modellieren, die UID berücksichtigt. Darüber hinaus argumentiere ich, dass es konzeptionell geeigneter und in der Lage ist, Vorhersagen über den Sprachgebrauch zu treffen, die von UID nicht berücksichtigt werden.

Viele der jüngsten Versuche, Kommunikation als rationale Handlung zu formalisieren, basieren auf der Informationstheorie. In diesem Rahmen ist eine optimale Effizienz und Wiederherstellung des Signals (durch einen Decoder oder "Hörer:in")

gegeben, wenn die Wahrscheinlichkeit des Auftretens eines beliebigen Segments so nahe wie möglich an der Kapazität des Kommunikationskanals liegt — oder so unvorhersehbar, wie der Decoder oder Hörer:in damit umgehen kann (wobei die Nachricht noch korrekt wiederhergestellt wird), jedoch nicht darüber hinaus. Diese Erkenntnis legt nahe, dass menschliche Sprecher:innen in der Lage sein sollten, ihre kommunikative Effizienz zu maximieren, indem sie unnötige Redundanz oder übermäßig vorhersehbare Elemente in ihrer Sprache eliminieren (d. h. alles, was nicht für die Signalwiederherstellung notwendig ist). Umgekehrt sollten sie zusätzliche Redundanz einfügen, wenn ein Element ansonsten zu (kontextuell) unvorhersehbar ist, um eine genaue Wiederherstellung durch den:die Hörer:in zu gewährleisten.

In den letzten Jahrzehnten kamen vermehrt rationale Modelle der Sprachproduktion auf. Eine Familie von eng verwandten Theorien der Sprachproduktion, die von der Informationstheorie inspiriert sind, schlagen vor, dass Sprecher:innen darauf abzielen, ihre Äußerungen so *effizient* wie möglich zu gestalten, indem sie den Ausgleich zwischen zwei konkurrierenden Bedürfnissen ausgleichen:

1. Das Bedürfnis, beim Sprechen durch den Einsatz eines Minimums an Zeit und Artikulationsaufwand zur Kommunikation einer bestimmten Information Aufwand zu sparen. Mit anderen Worten, die Sprecher:innen wollen so wenig Arbeit wie möglich verrichten.

2. Das Bedürfnis, sicherzustellen, dass das beabsichtigte Signal von dem:der Hörer:in bestmöglich wiedergegeben werden kann, so dass das grundlegende Ziel der sprachlichen Interaktion erreicht wird. Dabei wird typischerweise davon ausgegangen, dass der Kommunikationskanal zwischen Sprecher:in und Hörer:in verrauscht oder verlustbehaftet ist (entweder in Bezug auf kognitive Ressourcenbeschränkungen oder Störfaktoren aus der Umgebung), und dass es im Allgemeinen keine Garantie für den:die Hörer:in gibt, das beabsichtigte Signal (und die zugehörige Nachricht) wiederherzustellen, selbst wenn es klar kommuniziert wird.

Mein Schwerpunkt liegt auf der Uniform Information Density (UID), der wohl produktivsten dieser Theorien, die ein breites Spektrum an sprachlichen Phänomenen auf verschiedenen Ebenen der Produktion abdeckt. Diese Theorie bietet einen guten Ausgangspunkt, das Verhalten von Hörer:innen und Sprecher:innen in folgenden Bereichen zu erklären: a) Verwendung von informationstheoretisch redundanten Äußerungen und b) Verwendung von Personalpronomen anstelle von eindeutigen Referenzen.

UID geht von einer Menge an Äußerungsalternativen aus, die in ihrer Bedeutung gleichwertig sind (z. B. “Ich glaube, er ist verrückt” versus “Ich glaube, dass er verrückt ist”), und schlägt vor, dass Sprecher:innen grundsätzlich kürzere und effizientere Äußerungen wählen, wenn das Signal oder die Bedeutung, die kommuniziert werden soll, im Zusammenhang ausreichend vorhersehbar ist. Mathematisch ausgedrückt ist

die Wahrscheinlichkeit des Auftretens eines optionalen Elements i in der Sprache proportional zu seiner *Information* — oder der negativen logarithmischen Wahrscheinlichkeit des Auftretens im Kontext der sprachlichen Information, die es im Kontext liefern würde:

$$P(\text{Element}_i) \propto I(\text{Element}_i) = -\log_2 P(\llbracket \text{Element}_i \rrbracket \mid \text{Kontext})$$

UID geht davon aus, dass die Kommunikation über einen verrauschten Kanal stattfindet und dass das beabsichtigte Signal entweder durch Störfaktoren aus der Umgebung oder durch interne Prozesse zwischen den Gesprächspartner:innen verschlechtert werden kann. Infolgedessen reduzieren Sprecher:innen bevorzugt diejenigen Äußerungen, die trotz einer erwarteten Verschlechterung des Signals oder der Nachricht von dem:der Hörer:in wahrscheinlich wiederhergestellt werden können, da sie im Kontext eine hohe Vorhersagbarkeit (und infolgedessen Wiederherstellbarkeit) aufweisen. Mit anderen Worten: Je vorhersehbarer die Bedeutung einer bestimmten Äußerung oder eines Äußerungselements im Kontext ist, desto wahrscheinlicher ist es, dass der:die Sprecher:in diese Äußerung oder dieses Äußerungselement reduziert oder auslässt (sofern die Grammatik es erlaubt). Das allgemeine Ziel von Sprecher:innen ist es, Kommunikation robust und gleichzeitig effizient zu gestalten.

Belege für eine robuste und effiziente Kommunikation als allgemeines Ziel von Sprecher:innen finden sich auf mehreren Ebenen der Produktion. Auf der phonetischen und phonologischen Ebene reduzieren Sprecher:innen bevorzugt Vokale und lassen optionale Konsonanten in kontextuell vorhersehbaren Wörtern weg. Auf der syntaktischen Rollenebene bevorzugen Sprecher:innen, wo es grammatisch möglich ist, reduzierte oder nicht vorhandene morphologische Marker, wenn die syntaktische Rolle im Kontext vorhersehbarer ist. In ähnlicher Weise lassen Sprecher:innen ganze Wörter aus, die das Einsetzen verschiedener syntaktischer Strukturen signalisieren, wenn diese Strukturen kontextuell vorhersehbar sind — wie etwa optionale Konjunktionen oder optionale Relativpronomen. Auf lexikalischer Ebene neigen Sprecher:innen dazu, kürzere Wörter in Kontexten zu verwenden, in denen ihre Bedeutung besser vorhersagbar ist, und es gibt sprachübergreifende Belege dafür, dass die Wortlänge mit der durchschnittlichen Vorhersagbarkeit korreliert, die ein bestimmtes Wort im Kontext hat.

Eine Kernaussage des UID-Modells ist, dass es auf allen Ebenen der Produktion gilt. Es ist jedoch nicht klar, ob UID, so wie es derzeit formalisiert ist, die richtigen Vorhersagen auf der Diskursebene macht, wo die Annahme der Bedeutungsäquivalenz zwischen verschiedenen Arten der Kodierung derselben Information dazu neigt, nicht mehr gültig zu sein (nicht zuletzt, weil stark unterschiedliche Arten der Formulierung derselben Nachricht unterschiedliche pragmatische Schlussfolgerungen auslösen können), und Äußerungselemente für die Interpretation stärker vom Diskurskontext abhängen können. Während es auf anderen ("niedrigeren") Produktionsebenen möglich ist, die Unterscheidung zwischen der Vorhersagbarkeit von Signal und Bedeutung in

einem gegebenen Kontext zu vernachlässigen (da sie tendenziell eng miteinander verbunden sind), wird dies auf der Diskursebene zu einem dringlicheren Problem. Es ist auch weniger klar, wie man die Vorhersagbarkeit des betreffenden Elements messen kann, insbesondere wenn die Form der Äußerung und ihre Bedeutung weniger eng oder intrinsisch miteinander verbunden sind.

Darüber hinaus sagt das UID-Modell zwar voraus, dass Sprecher:innen eine robuste, aber dennoch effiziente Kommunikation anstreben, aber es setzt nur deutliche Einschränkungen für die Tendenz von Sprecher:innen, Äußerungen zu reduzieren. Wenn ein:e Sprecher:in ein sprachliches Element reduziert, das im Kontext nicht vorhersehbar ist, beeinträchtigt dies die Kommunikation dadurch, dass es dem:der Hörer:in übermäßig schwer gemacht wird, das beabsichtigte Signal oder die Bedeutung wiederherzustellen. Wenn ein:e Sprecher:in es jedoch versäumt, *effizient* zu sein, hat dies keine offensichtlichen Konsequenzen in Bezug auf das Ziel einer erfolgreichen, rationalen Kommunikation — der:die Hörer:in sollte immer noch in der Lage sein, die Bedeutung der Äußerung wiederzugewinnen, und das vielleicht sogar *besser*, selbst wenn der:die Sprecher:in mehr Aufwand betreibt, als auf seiner:ihrer Seite unbedingt notwendig ist.

Dies führt zu einem Konflikt mit der folgenden Vorhersage: Wenn ein:e Sprecher:in übermäßig detailliert oder wortreich ist, um eine Bedeutung auszudrücken, die eigentlich leicht wiederhergestellt werden kann, wird der:die Hörer:in wahrscheinlich daraus schließen, dass es einen *Grund* für die mangelnde Effizienz bei dem:der Sprecher:in gibt, wobei der *Grund* dann Teil der Botschaft wird, die er:sie ableitet. Mit anderen Worten, der kommunikative Erfolg wird nicht nur durch die Fähigkeit der Zuhörer:innen bestimmt, die wörtliche Bedeutung der Äußerung der Sprecher:innen korrekt zu erfassen, sondern auch durch die Fähigkeit der Zuhörer:innen, die gesamte Botschaft, die Sprecher:innen zu kommunizieren beabsichtigen, korrekt zu erschließen. Dies schließt jegliche pragmatische Bedeutung ein, die die Äußerung impliziert. Das UID-Modell, das keine Schnittstelle zur Pragmatik hat und sich nur mit der Wiederherstellung des beabsichtigten Signals und seiner wörtlichen Bedeutung zu befassen scheint, kann das Potenzial der Redundanz der Sprecher:innen, die Botschaft, die den Hörer:innen eine Aussage letztendlich vermittelt, zu verzerren, nicht berücksichtigen und ist dafür auch nicht geeignet.

Im ersten Teil der Arbeit untersuche ich, wie Versteher:innen (oder Hörer:innen) informationstheoretisch redundante Äußerungen interpretieren, wie z. B.:

(1) *John ging einkaufen.* ***Er bezahlte an der Kasse.***

Ich führe drei Experimente durch, die zeigen, dass bei gegebenem allgemeinen Weltwissen im Kontext des ersten Satzes in 1 die Versteher:innen typischerweise automatisch schlussfolgern, dass *John* gewohnheitsmäßig an der Kasse bezahlt, auch wenn dies nicht explizit erwähnt wird. Die explizite Erwähnung von *Bezahlen an der Kasse* durch den:die Sprecher:in ist daher redundant, und man kann vorhersagen,

dass die Hörer:innen versuchen würden, diese Redundanz zu verstehen. Tatsächlich zeige ich, dass Hörer:innen beim Hören des zweiten Satzes in 1 nicht nur die wörtliche Bedeutung verstehen (dass an der Kasse bezahlt wurde), sondern zusätzlich schlussfolgern, dass *John gewöhnlich nicht an der Kasse zahlt* — vermutlich, weil es sonst keinen Grund gäbe, dies explizit zu erwähnen.

Diese Art der pragmatischen Schlussfolgerung, die durch die informationstheoretische Redundanz von Sprecher:innen im Kontext ausgelöst wird, verändert die Bedeutung der Äußerung erheblich, und für den Fall, dass diese zusätzliche Bedeutung von den Sprecher:innen nicht beabsichtigt ist, stellt sie eine erhebliche Verzerrung der beabsichtigten Botschaft der Sprecher:innen oder der Fakten über die Welt dar, die die Sprecher:innen vermitteln wollten. Dies setzt eine im UID-Basismodell nicht berücksichtigte klare Grenze, wie redundant Sprecher:innen sein können, ohne die Bedeutung ihrer Äußerungen zu verändern. Darüber hinaus bietet UID keine Werkzeuge zur formalen Modellierung dieser Art von Ergebnissen und hat keine Schnittstelle zu nicht-literarischen, pragmatischen Äußerungsinterpretationen (nicht einmal auf konzeptioneller Ebene). Dies ist bestenfalls unbefriedigend.

Im nächsten Teil der Arbeit untersuche ich die Entscheidungen von Sprecher:innen beim Bezug auf Entitäten im Diskurs. Wenn sich der:die Sprecher:in zum Beispiel auf eine der Entitäten in "*John gab Bill das Buch*" bezieht, kann er:sie entweder den Eigennamen (*John* oder *Bill*) oder das Pronomen *er* in der dritten Person Singular verwenden. Wovon hängt diese Wahl ab? UID sagt voraus, dass Sprecher:innen Pronomen für besser vorhersehbare Referenten und Eigennamen für weniger vorhersehbare Referenten bevorzugen sollten. In dem gegebenen Kontext neigt das Verb "geben" beispielsweise zu Fortsetzungen, die das "Ziel" der übertragenden Aktivität erwähnen — mit anderen Worten: "Bill". Als der vorhersehbarere Referent sollte *Bill* daher im folgenden Diskurs grundsätzlich bevorzugt pronominalisiert werden. Empirische Belege für diesen Standpunkt sind jedoch spärlich und bestenfalls inkonsistent. Es wurde andererseits dargelegt, dass der Vorhersagbarkeitseffekt auf die Wahl des referierenden Ausdrucks nur bei der Verwendung bestimmter kontinuitätsbeeinflussender Verben, bestimmter Aufgabenparadigmen und/oder ab einem bestimmten Grad an referentieller Verwechselbarkeit auftritt — aber die Gründe dafür, warum der Effekt in manchen Kontexten auftritt, in anderen aber nicht, bleiben oft unklar.

Ich führe zwei Experimente durch, die zeigen, dass, wenn man Aufgabe und Versuchspersonen kontrolliert, die Vorhersagbarkeit des Referenten keinen Effekt auf die Wahl des referenziellen Ausdrucks hat — unabhängig vom Verb, das zur Kontinuitätsbeeinflussung verwendet wird, von der spezifischen Aufgabe, mit der sich die Teilnehmer beschäftigen, von der Form der Vorgängerwörter (Eigenname versus definite Beschreibung) oder vom Grad der vorhandenen referenziellen Mehrdeutigkeit (gleichgeschlechtliche versus gegengeschlechtliche Vorgängerwörter). Obwohl die Möglichkeit besteht, dass Effekte in interaktiveren und lebensnaheren Aufgaben oder in anderen Sprachen als Englisch beobachtet werden, deuten diese Ergebnisse darauf hin, dass UID in den meisten Kontexten die falschen Vorhersagen in Bezug auf die

Entscheidungen der Sprecher:innen bzgl. referentieller Ausdrücke macht. Obwohl UID keinen großen Anspruch auf Allgemeingültigkeit erhebt, stellt es keine Modellierungswerkzeuge zur Verfügung, mit denen sich erklären ließe, warum es in einigen Kontexten die richtigen Vorhersagen trifft, in anderen aber nicht (was entscheidend dazu hätte beitragen können, empirische und überprüfbare Vorhersagen über verwandte Phänomene zu treffen). Es stellt auch keine Werkzeuge zur Verfügung, die die widersprüchlichen empirischen Ergebnisse in diesem Forschungsbereich erklären könnten — ich argumentiere jedoch, dass das Rational Speech Act (RSA) Modell dies leistet.

Die oben diskutierten Ergebnisse deuten klar darauf hin, dass UID, so wie es derzeit formalisiert ist, keine adäquate Darstellung oder Werkzeuge für die Modellierung von Einschränkungen von Produktionen der Sprecher:innen oder der Auswirkungen der Redundanz der Sprecher:innen auf die genaue Nachrichtenübermittlung bietet. Dies ist nicht notwendigerweise ein Problem auf "niedrigeren" Ebenen der Produktion, wo UID ein einfaches, aber adäquates Modell der Äußerungsentscheidungen der Sprecher:innen zu sein scheint. Auf der Diskursebene wird jedoch deutlich, dass es grundsätzlich nicht in der Lage ist, die pragmatische Interpretation von Äußerungen zu berücksichtigen, da es keine echten Werkzeuge zur Modellierung der Äußerungsbedeutung bietet. Auf der Diskursebene sagt UID außerdem weitgehend fälschlicherweise voraus, dass Sprecher:innen kürzere oder einfachere referierende Ausdrücke für vorhersehbarere Referenten wählen sollten, und stellt keine Werkzeuge zur Verfügung, die widersprüchliche empirische Ergebnisse in diesem Bereich erklären könnten — auch in Fällen, in denen Effekte systematisch nachweisbar sind, möglicherweise aufgrund unterschiedlicher und kontextabhängiger pragmatischer Kompetenz von Agenten oder der Auswirkungen von Mehrdeutigkeit.

Im letzten Teil dieser Arbeit zeige ich, dass das Rational-Speech-Act-Modell (RSA) im Gegensatz dazu den Effekt, den Redundanz auf die Interpretation von Nachrichten hat, adäquat und explizit modellieren kann. Ich zeige weiter, dass das Hinzufügen einer Repräsentation des verrauschten Kanals zum Basismodell (was, wie ich argumentiere, konzeptionell und empirisch unabhängig davon motiviert ist) nicht nur die Tatsache berücksichtigt, dass wahrnehmungsmäßig markantere redundante Äußerungen stärkere pragmatische Inferenzen auslösen, sondern es dem RSA-Modell auch erlaubt, die von der UID vorhergesagten Präferenzen der Sprecher:innen für die Verwendung kürzerer und weniger kostspieliger (aber bedeutungsäquivalenter) Äußerungen für vorhersehbarere Bedeutungen allgemein zu repräsentieren.

Ich zeige weiter, dass das RSA-Modell mit verrauschten Kanälen — im Gegensatz zu UID — zusätzlich spekulativ die Tatsache erklären kann, dass die Vorhersagbarkeit des Referenten die Wahl des referierenden Ausdrucks in solchen experimentellen Kontexten zu beeinflussen scheint, die zu mehr pragmatischer Kompetenz seitens der oder zielgruppengerechterer Gestaltung für die Sprecher:innen oder Hörer:innen führen können, sowie in Kontexten, in denen referentielle Mehrdeutigkeit besteht. Diese Effekte können prinzipiell nicht im UID-Rahmen erklärt werden. Ob-

wohl abzuwarten bleibt, ob dieses Ergebnismuster in zukünftigen Arbeiten repliziert werden kann, machen die Modelle klare empirische Vorhersagen darüber, welche experimentellen Kontexte und Aufgabengestaltungen robustere Effekte hervorrufen sollten. Kurz gesagt, Experimente die so gestaltet sind, dass robuste Effekte konsistenter zu sehen sind, sind tendenziell viel lebensnäher und interaktiver. Ich argumentiere, dass es vernünftig ist, anzunehmen, dass solche Gestaltungen ein höheres Maß an pragmatischer Kompetenz auf Seiten der Sprecher:innen hervorrufen und/oder sie dazu veranlassen, ein höheres Maß an Rationalität oder Zielgerichtetheit bei der Wahl ihrer Äußerungen zu zeigen. Im RSA-Rahmen würde dies zum Auftreten von Effekten der Vorhersagbarkeit des Referenten auf die Wahl des referierenden Ausdrucks führen.

Letztlich argumentiere ich also, dass die UID-Hypothese ein nützliches und empirisch hoch produktives Modell für die Äußerungsentscheidungen von Sprecher:innen unterhalb der Diskursebene ist (z.B. phonetische/phonologische Veränderungen, Weglassen von optionalen grammatischen Partikeln, lexikalische Varianten unterschiedlicher Länge). Sowohl konzeptionell als auch empirisch scheint es jedoch auf der Diskursebene der Produktion unzureichend zu sein — wo die Unmöglichkeit, (pragmatische) Interpretationen der Hörer:innen von Äußerungen oder sogar die Äußerungsbedeutung selbst zu modellieren (sowie potenziell Grade der Rationalität der Sprecher:innen und Effekte von Mehrdeutigkeit), kritisch wird. Im Gegensatz zu UID setzt das Rational-Speech-Act-Modell klare Grenzen dafür, wie redundant ein:e Sprecher:in sein darf, ohne die Bedeutung der Äußerung wesentlich zu verändern — und ist darüber hinaus in der Lage, den Effekt der Redundanz auf die Interpretation der Äußerung zu formalisieren. Ebenso ist das RSA-Modell, wiederum im Gegensatz zu UID, prinzipiell in der Lage, das bestehende Muster positiver und negativer Befunde in Bezug darauf zu erklären, ob die Vorhersagbarkeit des Referenten die Wahl des referierenden Ausdrucks beeinflusst. Schließlich ist das RSA-Modell in der Lage, unterschiedliche Kompetenzniveaus der Sprecher:innen (und Hörer:innen) zu berücksichtigen, was spekulativ erklären könnte, warum stärkere und konsistentere Effekte der Vorhersagbarkeit des Referenten auf die Wahl des referierenden Ausdrucks in lebensnäheren und interaktiven experimentellen Gestaltungen beobachtet werden.

Acknowledgements

The work that I did on this thesis would not have been possible without the support of my mentors and colleagues.

My first thanks go to my advisor Vera Demberg, who gave me the freedom to pursue the projects that I was interested in, even when they went beyond the initial scope of the research program. She supported me in gaining expertise in fields and topics unfamiliar to both of us, while helping me always tie the work back to our initial focus on communicative efficiency and information density. She oversaw multiple iterations of many of the thesis chapters, and showed me how to create a coherent narrative out of disparate research findings.

I also profited greatly from help from other current and former colleagues on various parts of this thesis. Pranav Anand at UCSC initially introduced me to pragmatics, and was always willing to consult with me as I worked to situate my findings in a broader field that I at first had only the most basic familiarity with. Leon Bergen at UCSD gave me extremely useful feedback on some initial findings, as well as first suggesting that I use the RSA model to account for some of my results. Judith Degen at Stanford provided me with highly useful feedback on my first findings and initial models. Michael Franke at Osnabrück spent invaluable time with me working out the kinks in the final versions of various models, as well as discussing how the work could be situated in theory. A chance but extremely productive discussion with Oliver Bott at Tübingen took my research on referring expressions in an unexpected direction. I further thank Hannah Rohde at Edinburgh, University of Rochester BCS and Linguistics, and members of the Institut Jean Nicod in Paris for extensive feedback on ongoing research and many fruitful discussions.

My colleagues at Saarland, including Dave Howcroft, Asad Sayeed, Alessandra Zarcone, and Elisabeth Rabs in particular, were always available for interesting discussions of our field and general support throughout the years. This period would have been far more challenging (and boring!) without you.

Deepest thanks to my family and friends, who stood beside me all this time. Being so far away from many of you, scattered all over multiple continents, was challenging,

but you always welcomed me back whenever I was able to visit.

My final thanks goes to Alex, for always being there, providing me with moral and practical support, and for your unwavering confidence in me. Thank you so much for everything.

Contents

Chapter 1

Introduction

Given alternative ways of formulating an utterance with the same meaning – removing or omitting an optional word, shortening or lengthening a syllable, or including or leaving out a piece of information the listener should be able to infer independently – how and why do speakers make the linguistic choices that they do? Do speakers formulate their utterances in a principled, rational (or goal-directed) manner, and to what degree (Zipf, 1949; Aylett & Turk, 2004; Levy & Jaeger, 2007; Jaeger, 2010; Frank & Goodman, 2012)? Fundamentally, a rational speaker could be said to make linguistic choices, and structure their utterances in a way that is most in accordance with the goals of communicating a given piece of information to a listener, while taking into account cognitive resource limitations as well as any relevant particularities of the environment. The speaker may strive to meet specific goals such as conserving the amount of effort that they expend on transmitting information, or tailoring their linguistic choices to the listener's presumed mental state or existing beliefs. Furthermore, the speaker should attempt at least to some degree to do so in an *optimal* manner, achieving a maximally successful outcome within given constraints – above and beyond simply encoding and transmitting the information that they intend to communicate.

Although the idea that speakers should attempt to maximize the utility of their utterances appears intuitive, it can be difficult to formulate, and yet more difficult to test systematically. Exactly which communicative goals do speakers tend to share, and how do they influence the manner in which they communicate – particularly when goals conflict? How does one formulate a set of principles that ensure successful communication? In the presence of limited cognitive resources and limited information about the listener's mental state and prior beliefs about the world, to what extent do speakers in fact make situationally optimal communicative choices, and do they attempt to tailor their choices to the listener? Do speakers go as far as considering the possible listener interpretations of various alternative ways of encoding a given unit of information, and selecting the one that is most likely to result in the correct in-

terpretation? And do they strive for an optimal, or simply a good-enough approach? These are empirical questions that are difficult to answer without a framework that makes clearly falsifiable, and, critically, quantifiable (i.e., testable) predictions.

Throughout this thesis, I start by looking at a prominent computational theory of language production (Uniform Information Density, or UID), which argues that rational speakers preferentially omit or shorten optional elements in their speech that encode more predictable, or recoverable information (Levy & Jaeger, 2007; Jaeger, 2010). I argue that this theory is descriptively inadequate when accounting for linguistic choices which may trigger pragmatic interpretations on the part of the listener, and that in some cases it fails to make correct empirical predictions as a result. I then propose that the Rational Speech Act (RSA) model – a computational model of goal-directed back-and-forth pragmatic reasoning between a speaker and a listener – can not only be used to model, with minor modifications, the phenomena that UID accounts for, but is more conceptually adequate, and capable of making predictions about language use that UID fails to account for. In the rest of this chapter, I focus on introducing the UID model, as well as the linguistic phenomena that challenge its descriptive adequacy. I then propose that the RSA model can incorporate critical UID assumptions with some conceptually and empirically motivated modifications, improves on modeling the speaker-listener interaction (as well as explicitly modeling utterance meaning), and is able to account for discourse-level speaker choices that UID is otherwise unequipped to account for.

1.1 The Uniform Information Density Model

Many of the recent attempts to formalize the idea of communication as a rational act are grounded in information theory (Shannon, 1948). Shannon proposed a "Mathematical Theory of Communication" which provided a set of methods for determining the most efficient way of transmitting information through a noisy, or lossy, communication channel. He demonstrated that optimal efficiency and signal recoverability (by a decoder, or "listener") occurs when the probability of occurrence of any given segment is maximally close to channel capacity – or about as unpredictable as the decoder or listener can handle (while still recovering the message correctly), but no more. This finding suggests that a human speaker should be able to maximize their communicative efficiency by eliminating unnecessary redundancy, or overly predictable elements, in their speech (i.e., anything not necessary for signal recovery) – while inserting additional redundancy where an element may otherwise be too (contextually) unpredictable to ensure accurate recovery by the listener. A basic schematic illustrating the principles of the noisy channel can be seen in Figure 1.1.

I will discuss the noisy channel further in the following chapter, and for the time being, will move on to discussing theories of language production inspired by the noisy channel theorem. In the past several decades, rational models of language production

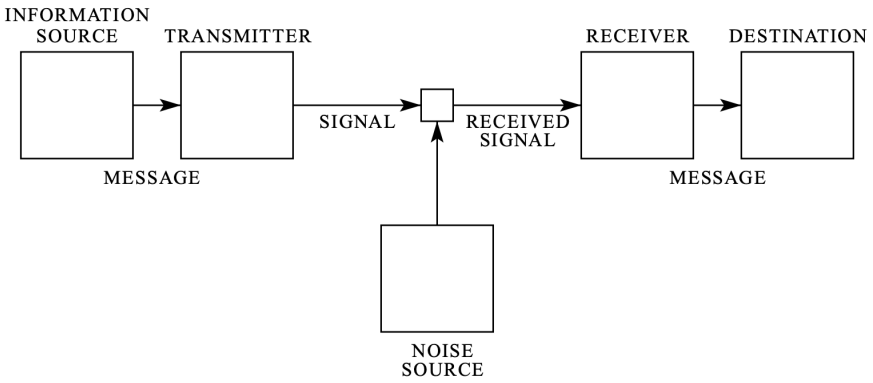

Fig. 1—Schematic diagram of a general communication system.

Figure 1.1: Schematic illustrating the noisy channel, from Shannon (1948).

(as well as language processing[1]) have proliferated. A family of closely-related theories of language production (Aylett & Turk, 2004; Levy & Jaeger, 2007; Jaeger, 2010; Genzel & Charniak, 2002), conceptually differing mostly only in their scope, and based in Shannon (1948)'s observations, have proposed that speakers aim to make their utterances maximally *efficient* by balancing out two competing pressures:

1. The pressure to conserve speaker effort, by expending minimal time and articulatory effort on communicating a given piece of information. In other words, speakers are lazy and want to do the minimum amount of work possible.

2. The pressure to ensure that the intended signal is maximally recoverable by the listener, or that the basic goal of the linguistic interaction is achieved. This typically assumes that the communication channel between speaker and listener is noisy, or lossy (either in terms of cognitive resource limitations, or environmental 'noise'), and that the listener is in general not guaranteed to recover the intended signal (and corresponding message) even if communicated clearly.

I focus on Uniform Information Density, or UID (Levy & Jaeger, 2007; Jaeger, 2010), arguably the most productive of these theories, which covers a wide range of linguistic phenomena at various levels of production. This theory provides a good starting point for accounting for listener and speaker behavior in the domains of: a) use of informationally redundant utterances; and b) use of personal pronouns in place of unambiguous proper-name references.

UID assumes, as a starting point, a set of utterance alternatives equivalent in meaning (e.g. "*I think he's crazy*" vs. "*I think that he's crazy*"), and proposes that,

[1]I will generally devote little space to the corresponding theories of language processing, as they are primarily concerned with linking the predictability of linguistic elements to processing difficulty (e.g., as measured by reaction times). As I do not present any measures of processing difficulty throughout this work, but rather focus on production choices and their pragmatic interpretations, I do not for the most part consider them directly relevant.

all things being equal, speakers will choose shorter and more efficient utterances when the signal or meaning to be communicated is sufficiently predictable in context. Mathematically speaking, the probability of occurrence of an optional element i in speech is proportional to its *information* – or the negative log probability of occurrence, in context, of whatever linguistic information[2] that it would provide in context:

$$P(\text{element}_i) \propto I(\text{element}_i) = -\log_2 P(\llbracket\text{element}_i\rrbracket \mid \text{context}) \tag{1.1}$$

UID assumes that communication takes place in a noisy channel, and that the intended signal may become degraded through either environmental noise, or interlocutor-internal processes[3]. As a result, speakers preferentially reduce those utterances which are likely to be recovered by the listener *despite* some expected signal or message degradation, due to their high predictability (and, as a consequence, recoverability) in context. In other words, *the more predictable the meaning of a given utterance or utterance element is in context, the more likely the speaker is to reduce or omit that utterance or utterance element (grammar permitting)*. The overarching aim of speakers is to make communication robust, yet efficient.

Evidence for robust and efficient communication, as an overarching speaker goal, is seen at multiple levels of production. At the phonetic and phonological levels, speakers preferentially reduce vowels and omit optional consonants in contextually predictable words (Aylett & Turk, 2004, 2006; Bell et al., 2009; Gahl et al., 2012, 2004; Jurafsky et al., 2001). At the syntactic role level, where grammatically licensed, speakers prefer reduced or null morphological markers when the syntactic role is more predictable in context (Kurumada & Jaeger, 2015; Norcliffe & Jaeger, 2014). Speakers similarly omit whole words signaling the onsets of various syntactic structures when those structures are contextually predictable – such as optional complementizers (Jaeger, 2010), or optional relativizers (Jaeger, 2011; Wasow et al., 2011). At the lexical level, speakers tend to use shorter words in contexts where their meaning is more predictable (Mahowald et al., 2013), and there is evidence cross-linguistically that word length correlates with the average predictability that a given word has in context (Piantadosi et al., 2011).

A core prediction of UID is that it applies at all levels of production. However, it is not clear if UID, as currently formalized, makes the correct predictions at the discourse level, where the assumption of meaning-equivalence among different ways of encoding the same information tends to break down (not least because substantially different ways of formulating the same message may trigger different pragmatic inferences), and utterance elements may depend more on discourse context for interpretation. For

[2]In the non-probabilistic sense; e.g. something may serve as a signal of grammatical role, clause onset, word identity/meaning (in cases where a phoneme may be optionally dropped or reduced), etc.

[3]To note, although UID takes the concept of the noisy channel as inspiration, it does not attempt to formalize it in any way, and remains agnostic on the source of 'noise' in general, or in a given interaction.

instance, while *Mary* and *she* may refer to the same individual in a given sentence, *she* is more ambiguous as to its meaning, and could technically refer to any female entity, which only context disambiguating the intended referent. An ability to recover the signal and literal meaning of the utterance, in this case, is insufficient to correctly interpret the message – as a result, it is difficult to predict exactly how UID predictions will play out. Further, while at other ("lower") levels of production it may be possible to hand-wave away the distinction between the predictability of signal and meaning in a given context (given that they tend to be closely linked), this becomes more of a pressing problem at the discourse level. It also becomes less clear how to measure the predictability of the element in question, particularly as the form of the utterance and its meaning are less closely or intrinsically linked[4].

Additionally, while UID predicts that speakers aim for robust, yet efficient, communication, it places clear constraints only on a speaker's tendency to *reduce* utterances. If a speaker reduces a linguistic element that is unpredictable in context, this presumably impairs communication by making it excessively difficult or even impossible for the listener to recover the intended signal or meaning. However, if a speaker fails to be *efficient*, this is met with no apparent consequences with respect to the goal of successful, rational communication – the listener should still be able to recover the utterance's meaning, and is in fact perhaps *better* able to recover it, even if the speaker expends more effort than is strictly necessary on their end.

This runs into conflict with the following prediction (which I confirm in Chapter 4: if a speaker is excessively detailed or wordy in expressing a meaning that is otherwise easily recovered, the listener is then likely to infer that there is a *reason* for the speaker's lack of efficiency, the *reason* then becoming part of the message that they infer. In other words, communicative success is determined not only by the listener's ability to correctly recover the literal meaning of the speaker's utterance, but also by the listener's ability to correctly infer the entire message that the speaker intends to communicate, which critically includes any pragmatic meaning that the utterance implicates. UID, which does not interface with pragmatics, and appears to concern itself solely with recovery of the intended signal and its literal meaning, cannot account for, and is critically unequipped to model the potential for speaker redundancy to distort the message that an utterance ultimately conveys to the listener.

In the following section, I briefly outline the empirical challenges to UID that I present in this thesis. I then propose that if an assumption of a noisy channel is integrated, then the Rational Speech Act (RSA) model, a probabilistic model of iterative pragmatic reasoning, is able to both predict the same effect on utterance choice that UID models, and to represent the potential for redundancy to distort the

[4]UID is often not explicit about exactly what sort of predictability is being measured or modeled – for instance, whether it's from the perspective of the speaker or the presumed perspective of the listener. Often there is an assumption that the different 'predictabilities' should be sufficiently similar that one can simply remain agnostic while approximating a measure based, for instance, on occurrences within a corpus, or predictions of a language model.

speaker's intended message, by modeling both literal utterance meaning and potential pragmatic interpretations (as well as by explicitly modeling aspects of the speaker and listener's mental states). The RSA model represents the process by which speaker and listener agents reason about each others' beliefs and intentions, in the context of an overarching goal of successfully communicating (or inferring) both. Fundamentally it is based in pragmatic, or Gricean principles of rational communication. Pragmatic theory, including the Gricean programme, is discussed in more detail in the following chapter.

1.2 Challenges to UID at the Discourse Level

In Chapters 3-4, I look at how comprehenders (or listeners) interpret informationally redundant utterances, such as:

(1) *John went shopping.* ***He paid the cashier.***

Given common world knowledge, I show that in the context of the first sentence in 1, comprehenders typically infer automatically that *John* habitually pays the cashier, even if this is not explicitly mentioned. Explicit mention of *cashier-paying* by the speaker is, therefore, redundant, and one may predict that listeners would attempt to make sense of this redundancy. I in fact demonstrate that, upon hearing the second sentence in 1, comprehenders not only receive the literal meaning (that the cashier was paid), but additionally infer that *John does not habitually pay the cashier* – presumably, because otherwise, there would be no reason to mention this explicitly.

This type of pragmatic inference, triggered by the speaker's informational redundancy in context, substantially alters the utterance's meaning, and in the case where this additional meaning is not intended by the speaker, it constitutes a substantial distortion of the speaker's intended message, or of the facts about the world that the speakers intended to convey. This places a clear limit, not accounted for by the base UID model, on how redundant a speaker may be without altering the meaning of their utterances. Further, UID provides no tools for formally modeling this type of outcome, and does not interface with non-literal, pragmatic utterance interpretations (even on a conceptual level). This is, at best, unsatisfying.

In Chapters 5-6, I further look at speaker choice of how to refer to entities in the discourse. For example, when referring back to either one of the entities in "*John gave the book to Bill*," the speaker may choose to use the proper name (*John* or *Bill*), or the third-person singular pronoun *he*. What governs this choice? UID predicts that speakers should preferentially choose pronouns for more predictable referents, and proper names for less predictable referents (Tily & Piantadosi, 2009). In the given context, for instance, the verb *gave* biases towards continuations which mention the *goal* of the transferring activity – in other words, *Bill*. As the more predictable

referent, *Bill* should therefore be preferentially pronominalized in the following discourse, all things being equal. However, empirical evidence for this position is scarce and, at best, inconsistent (Arnold, 2001; Rosa & Arnold, 2017; Fukumura & van Gompel, 2010; Rohde & Kehler, 2014; Bott et al., 2018, p.c. July 25, 2018). It has been alternately argued that the effect of predictability on referring expression choice emerges only given the use of certain continuation-biasing verbs, certain task paradigms, and/or a certain degree of referential ambiguity (Rosa & Arnold, 2017; Bott et al., 2018, p.c. July 25, 2018) – but the reasons for why the effect should emerge in some contexts but not others often remain elusive.

In Chapter 6, I show that when one controls for task and experiment population, referent predictability has no effect on referring expression choice – independently of the verb used to bias continuations, the specific task participants engage in, the form of the antecedents (proper name vs. definite description), or the degree of referential ambiguity present (same-gender vs. opposite-gender antecedents). While there remains a possibility that effects would be observed in more interactive and naturalistic tasks, or in languages other than English, these results suggest that in most contexts, UID makes the wrong predictions with respect to speaker choice of referring expressions. Although UID does not make very strong claims to universality, it does not provide modeling tools that may be used to account for why it makes the correct predictions in some contexts, but not others (which, critically, could be used to make empirical and testable predictions about related phenomena). It also provides no tools which could potentially account for the conflicting empirical results in this research area specifically – however, I argue in the next section that the Rational Speech Act (RSA) model does.

1.3 The Rational Speech Act Model as a Solution

The findings discussed in the previous section clearly suggest that UID, as currently formalized, does not adequately represent or provide tools for modeling constraints on speaker production, or the effects of speaker redundancy on accurate message transmission. This is not necessarily a problem at "lower" levels of production, where UID appears to be a simple but adequate model of speaker utterance choice. However, at the discourse level it is clear that it is fundamentally unequipped to take the pragmatic interpretation of utterances into account, as it provides no real tools for modeling utterance meaning. At the discourse level, UID further largely incorrectly predicts that speakers should choose shorter, or simpler referring expressions for more predictable referents, and provides no tools that may account for conflicting empirical results in this area – including in cases where effects *may* be systematically detectable, potentially due to varying and context-dependent pragmatic sophistication of agents, or the effects of ambiguity.

In Chapter 8, I propose that the Rational Speech Act (RSA) model can, in contrast,

adequately and explicitly model the effect that utterance redundancy has on message interpretation. I further show that adding a representation of the noisy channel to the base model (which I argue is independently conceptually and empirically motivated) not only accounts for the fact that more perceptually prominent redundant utterances trigger stronger pragmatic inferences, but allows the RSA model to generally represent UID-predicted speaker preferences for using shorter and less costly (but meaning-equivalent) utterances for more predictable meanings.

In Chapter 9, I further propose that the noisy channel RSA model can additionally – in contrast to UID – speculatively account for the fact that referent predictability does appear to effect referring expression choice in experimental contexts that may prompt more pragmatic sophistication or audience design on the part of speakers of listeners, as well as in contexts where there is referential ambiguity. These effects, principally, can not be accounted for in the UID framework. Although it remains to be seen whether this pattern of results will replicate in future work, the models make clear empirical predictions for which experimental contexts and task designs should prompt more robust effects. Briefly, those experimental designs in which robust effects are more consistently seen have tended to be far more naturalistic and interactive (Rosa & Arnold, 2017). I argue that it is reasonable to assume that such designs may prompt a greater level of pragmatic sophistication on the part of speakers (cf. Lockridge & Brennan, 2002), and/or prompt them to exhibit a greater degree of *rationality*, or goal-directedness, in their utterance choices. In the RSA framework, this would lead to an emergence of effects of referent predictability on referring expression choice.

Ultimately, then, I argue that the UID hypothesis is a useful and empirically highly productive model of speaker utterance choice below the discourse level (e.g., phonetic/phonological alterations, omission of optional grammatical particles, lexical variants of different lengths). However, both conceptually and empirically it appears inadequate at the discourse level of production – where an inability to model (pragmatic) listener interpretations of utterances, or indeed utterance meaning itself (as well as, potentially, degrees of speaker rationality and effects of ambiguity) becomes critical. In contrast to UID, the Rational Speech Act model places clear limits on how redundant a speaker may be without substantially altering utterance meaning – and further, is equipped to formalize the effect of redundancy on utterance interpretation. Similarly, again in contrast to UID, the RSA model is also principally able to account for the existing pattern of positive and negative findings with respect to whether referent predictability affects referring expression choice (although, to note, my own results do not themselves confirm this pattern). Lastly, the RSA model is able to accommodate different levels of speaker (and listener) sophistication, which speculatively may account for why stronger and more consistent effects of referent predictability on referring expression choice are seen in more naturalistic and interactive experimental designs.

In the following section, I list the contributions that this thesis makes.

1.4 List of Contributions

1.4.1 Part 1 (Chapters 3-4): Interpretation of Informationally Redundant Utterances

1. Informationally redundant utterances (IRUs) trigger pragmatic inferences regarding activity *habituality*, which substantially alter the listener's beliefs about the background world state.

2. More perceptually prominent, or effortful, IRUs trigger *stronger* inferences, even as the truth-conditional utterance meaning does not change.

- **Take-Away:** UID does not propose any specific constraints on speaker redundancy, and due to its failure to model utterance meaning, it includes no mechanism that can account for the potential of redundancy to distort a speaker's message. Any theory or framework which accounts for the effect of redundancy on utterance meaning must take into account pragmatic reasoning.

1.4.2 Part 2 (Chapters 5-6): Referring Expression Choice

1. There is no consistent influence of referent predictability on referring expression (RE) choice, in contrast to UID predictions.

2. The effect does not consistently emerge in the context of specific verb types, contra Rosa & Arnold (2017).

3. The effect does not consistently emerge in the context of constrained-choice passage completion tasks coupled with referential ambiguity (same-gender antecedents), contra Bott et al. (2018).

4. The effect does not emerge when antecedents are lengthier definite descriptions (which should prompt more pronominalization, in the interest of conciseness). References back to definite descriptions are more likely to be pronominalized overall, however.

- **Take-Away:** UID predicts that referring expression choice should be reliably modulated by predictability; this prediction is not supported. UID similarly cannot account for the empirical pattern of positive findings to date, which rest on factors such as referential ambiguity and varying levels of speaker rationality, as discussed further in Chapter 9. More generally, this appears to be more evidence that UID is a poor model for production at the discourse level.

1.4.3 Part 3 (Chapters 7-9): Rational Speech Act (RSA) Model

1. **(IRU)** The RSA model, assuming joint listener reasoning about utterance meaning and background world states, can formally represent the process by which IRUs trigger *habituality* inferences.

2. **(IRU)** A *noisy channel* RSA model incorporating joint reasoning generates stronger inferences for more perceptually prominent, or effortful utterances. This, alongside conceptual and empirical motivation, suggests that *noisy channel* machinery should be added to the standard RSA toolkit.

3. **(RE)** The RSA model can represent speakers with different degrees of pragmatic sophistication or *rationality* (i.e., inclination towards audience design), which may account for more stable and stronger effects of predictability or utterance choice in more naturalistic, interactive tasks.

4. **(RE)** The RSA model, which unlike UID considers the utility of utterance meaning to the listener, as well as listener expectations, straightforwardly accounts for why more stable effects of predictability on referring expression choice are found when pronominal reference is *ambiguous* (same-gender antecedents).

5. **(RE)** Incorporating *noisy channel* machinery into the RSA model accounts for apparently weaker, but residually present effects of referent predictability in cases where pronominal reference is *not* ambiguous (opposite-gender antecedents).

6. **(RE)** Unlike UID or similar production models, the RSA model accounts for speaker/listener asymmetry with respect to taking into account referent predictability in production and interpretation, respectively (cf. Rohde & Kehler, 2014).

7. **(RE)** In contrast to existing Bayesian accounts of referring expression choice (cf. Rohde & Kehler, 2014), RSA (as well as UID, although this cannot be formalized to the same extent) can principally account for the emergence of the proposed grammatical constraint on speaker referring expression choice.

- **Take-Away:** Generally, the RSA model, particularly after incorporating the assumption of a noisy communication channel, is able to account for UID phenomena at the discourse level, whereas UID is principally unable to account for phenomena accounted for by the RSA model. The RSA model is likely a better tool for modeling speaker utterance choices at the discourse level, where alternative utterances are less likely to be meaning-equivalent, and lengthier ways of expressing the same meaning is more likely to trigger pragmatic inferences on the part of the listener. At "lower" levels of production, UID likely remains an adequate model of utterance choice.

Chapter 2

Background

This thesis draws upon two major theoretical traditions: that of information-theoretic constraints on utterance choice and comprehension on the one hand, and that of Gricean rationality on the other. Phenomenologically, it further draws upon two large bodies of literature: that which seeks to contextualize and account for speaker use of informationally redundant utterances, and that which attempts to account for speaker choice of referring expression.

Ultimately, I argue that the basic predictions of both efficiency-based theories of utterance choice, and pragmatic theories of utterance interpretation, fall out of the modified Rational Speech Act (RSA) model. I further argue that this model either better accounts for, or overcomes, several critical limits of simpler information-theoretic theories of utterance choice, such as the UID (Uniform Information Density) hypothesis.

In this section, I cover the background literature on efficiency-based theories of utterance choice, focusing on the UID hypothesis, as well as the basic theoretical literature on pragmatic theories of utterance interpretation. I also cover the background literature on the Rational Speech Act model, as well as recent modifications of this model which I draw on in my thesis. Literature concerning the production and interpretation of the specific phenomena that I study – informationally redundant utterances, and referring expressions – is covered in separate background chapters, 3 and 5.

2.1 Efficiency-Based Theories of Utterance Choice

Efficiency-based theories of utterance choice include a large family of hypotheses which aim to explain utterance choice at different levels of linguistic production, and vary in their focus on accounting for online production choices vs. conventionalized linguistic form. What these theories generally hold in common is that, at their level

of focus, they aim to explain linguistic production and/or form as arising from two competing pressures:

1. A listener-centric pressure to transmit a message with sufficient fidelity; i.e., to ensure that the listener is able to recover the form and meaning of the intended message, particularly in the context of a noisy channel.

2. A speaker-centric pressure to expend the minimum amount of energy possible in transmitting said message, with the aim to maximally conserve speaker time and effort.

Unsurprisingly, these two pressures are often in direct conflict: a speaker who wishes to transmit a message with maximum fidelity may choose to enunciate every element of the message with maximum clarity, to leave nothing unsaid, and to make the message as redundant and invulnerable to loss as possible. However, this speaker will likely also expend a tremendous amount of energy in accomplishing this goal, and in most cases much of this effort will have been unnecessary: most listeners may have been able to recover the intended message given a far less detailed and redundant signal. Similarly, a speaker who prioritizes the conservation of their own effort may, at the extreme, simply say nothing. Needless to say, this strategy will typically not result in the listener inferring the intended message, unless they have already inferred it through other means.

The Uniform Information Density (UID) hypothesis can be regarded as an umbrella theory which encompasses the base arguments and predictions of similar, domain-specific theories. For this reason, in the rest of this section, I will in general refer to UID, in lieu of referring to each of these theories specifically. Much like similar theories, UID attempts to relate production choices (specifically, the tendency to use reduced variants of linguistic elements, where possible) to the predictability or recoverability of the element in context.

The base reasoning behind UID is that speakers will balance off the two competing pressures above by omitting or reducing linguistic elements only, or preferentially, in contexts where the listener is sufficiently likely to recover these elements. In other words, the more predictable or recoverable a particular message is in context, the less need there is for a speaker to make that message explicit, and the more likely the speaker will be to reduce the signal used to communicate said message wherever possible, or in fact to omit it altogether.

This hypothesis is grounded in the notion of the *noisy channel*, initially proposed by Shannon (1948). A schematic illustrating the concept can be seen in Figure 2.1.

The base idea is that any signal produced by a speaker – attempting to convey a certain message – is subject to corruption by noise. In the context of language, the noise may take the form of environmental noise distorting the auditory signal, any other conditions of the environment or interlocutors that may result in perceiving a distorted signal, or perhaps even cognitive factors distorting speech production or

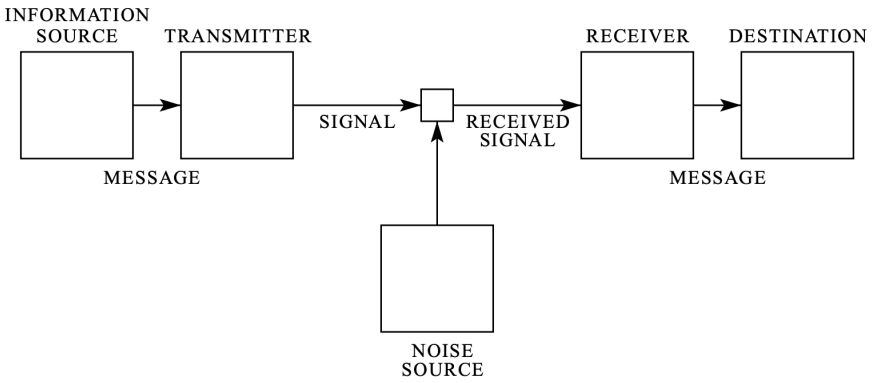

Fig. 1—Schematic diagram of a general communication system.

Figure 2.1: Schematic illustrating the noisy channel, from Shannon (1948).

comprehension (e.g., memory distortions). Shannon demonstrated that optimal signal recoverability and efficiency (or least required effort) occurs when the contextual predictability of each element in the signal is as close as possible to channel capacity, or about as unpredictable as can be without sacrificing the ability of the listener (or receiver) to recover the intended message - as the more predictable an element is in context, the easier it is to recover. A highly predictable, easily recoverable element may be (when possible) shortened, underarticulated, or omitted altogether when the context otherwise conveys the 'missing' information.

The main distinction between UID and similar theories is that it 1) aims to be linguistic domain-general, and 2) is stated in explicit, formal terms. While theories such as the Smooth-Signal Reduction Hypothesis (Aylett & Turk, 2004) aim to account only for vowel shortening within words, UID aims to account for utterance choice at all levels of linguistic production: phonetics/phonology, morphology, choice of lexical form, signaling of syntactic structure, and signaling of discourse-level phenomena. Second, while most similar theories express this hypothesis informally, as described above, UID states its predictions in formal terms, directly relating empirical measures of predictability to the likelihood of using a more effortful (or explicit) version of a linguistic element to convey a given unit of information. In the following equation, assume we are talking about a case where an element may be either produced in full, or omitted entirely[1]:

$$P(\text{element}_i) \propto I(\text{element}_i) = -\log_2 P(\text{info conveyed by element}_i \mid \text{context}) \quad (2.1)$$

In other words, the likelihood of producing a given linguistic element i is proportional to the information (in the information-theoretic sense) carried by form i, which is equal to the negative log-probability of the linguistic information that it conveys in context. The lower the information carried by the linguistic element, the less likely

[1]This may be a word that can be optionally omitted, a full (vs. reduced) vowel, a longer (vs. shorter) version of a given expression, and so forth.

it is to be produced. Conversely, the more information the element carries, the more likely it is to be produced relative to any meaning-equivalent lower-information alternative Context is typically defined as the preceding series of linguistic elements: $context = element_n...element_{i-1}$. Non-linguistic context is typically removed from consideration, due to the difficulty in measuring and quantifying it.

UID further proposes that speakers aim to keep the density of information transmitted maximally uniform – in essence ensuring that every component of the message is roughly equally recoverable, to the degree that the grammar, or other constraints on linguistic production, allow. In other words, drawing from information theory, UID postulates that there is an optimal 'channel capacity,' or lower limit on how recoverable each element of the message must be to reasonably ensure accurate message transmission. It is therefore in the speaker's best interest to use reduced forms of words or utterances wherever possible, so long as the intended message remains sufficiently (and maximally equally) recoverable to the listener.

The validity of this particular claim is beyond the scope of this thesis, but it is relevant to point out that the cost of exceeding channel capacity – using reduced forms of words or utterances when the information they are meant to convey is not sufficiently recoverable in context, given the communication channel – is clear. In contrast, the cost of essentially undershooting channel capacity, by using more "full" forms that lead to very easy recovery of the intended information in context, is less so. While in one case, the speaker risks the non-transmission of the intended message, in the other case, the speaker simply risks expending a bit more energy on speech than strictly necessary – a rather soft and violable limit.

The question then arises of whether speakers are in fact 'optimal' in conserving articulatory energy wherever possible. There is in fact a fair amount of evidence that speakers are *not* always optimal, and that they frequently expend more effort than strictly necessary in communicating a message (Baker et al., 2008; Walker, 1993). The question of whether the redundancy typically found in speech is evidence of non-optimality is further addressed in Chapter 3, but intuitively the costs of excessive articulatory effort, in terms of impairing communication, would appear to be minimal, and if anything, it is highly plausible that more effortful and redundant speech assists signal recovery.

First, I will cover the empirical evidence for UID, and then I will raise some remaining empirical and theoretical questions as to:

1. The degree to which UID holds above the level of syntactic structure and minor linguistic shortenings or omissions, at which the assumption of meaning-equivalence among linguistic alternatives begins to falter, and reliance on world knowledge and pragmatic reasoning in recovering the intended message increases. I argue that the evidence for this is limited, lacking, or mixed.

2. The degree to which redundancy impairs efficient communication. I argue that while unnecessary redundancy does not generally impair (and, if anything, im-

proves) transmission of an utterance's literal meaning, excessive redundancy may trigger unintended pragmatic inferences, which may substantially distort the intended message. In Chapter 4, I present empirical evidence for this, and in Chapter 8 a formal model of how such inferences may be generated.

2.1.1 Reduction at the Level of Lexical/Phonological Form and Syntactic Structure

At the level of lexical/phonological form and syntactic structure, there is abundant evidence for the reduction of linguistic elements which signal particularly predictable sounds, grammatical roles, word meanings, or syntactic structures. In other words, the more predictable a given component of the intended message – whether it's the identity of a particular word, a particular semantic meaning, or a particular syntactic structure – the less signal is used to communicate it, whenever signal reduction is a licensed alternative.

At the phonetic level, speakers have been shown to reduce vowels in contextually predictable words, whether by reducing vowel duration (Aylett & Turk, 2004), or the level of distinguishing articulatory detail (Aylett & Turk, 2006; Gahl et al., 2012). Optional consonants are similarly more likely to be omitted in more predictable words. For example, t/d-deletion, which occurs frequently in English, is more likely to occur in more predictable words (Gahl et al., 2004; Jurafsky et al., 2001). Interestingly, those consonants that on average occur in more predictable contexts are more likely to be omitted, regardless of how predictable the word is in its local context (Cohen Priva, 2008, 2015; Seyfarth, 2014).

At the level of grammatical roles, speakers have similarly been shown to prefer reduced or null morphological markers, when such are available, in context where the information encoded by those markers is particularly contextually predictable. Frank & Jaeger (2008) demonstrate that speakers are more likely to reduce forms such as "did not" to "didn't," and "have not" to "haven't," when the negation is contextually predictable. Similarly, optional case markers may be omitted in Japanese when the grammatical function that they encode is particularly predictable Kurumada & Jaeger (2015), and head-marking morphology in Yucatec Maya may be omitted when the grammatical function of the word affected is more predictable (Norcliffe, 2009; Norcliffe & Jaeger, 2014).

Whole words optionally signaling syntactic structures may similarly be omitted. The function word *that*, which optionally signals the onset of a complement clause, is preferentially omitted in contexts where the clause onset is more predictable (Jaeger, 2010). The word *that* may also signal the onset of a relative clause, and similarly is more likely to be omitted when the onset of the relative clause is predictable in context (Jaeger, 2011; Wasow et al., 2011). Beyond this, there is still limited evidence for the omission or reduction of function words (Jaeger & Buz, 2017).

Beyond functional elements, there is similarly evidence that speakers preferentially use shortened variants of content words in contexts where their meaning is particularly predictable. For example, speakers may prefer 'math' in place of 'mathematics,' or 'chimp' instead of 'chimpanzee' (Mahowald et al., 2013). A corpus study showed that the contextual predictability of longer word forms is significantly lower than that of shorter forms. Similarly, a sentence completion study showed that participants are more likely to use short forms in contexts where their meaning is more contextually predictable. Similarly, there is evidence that cross-linguistically, those words that are on average more predictable in context are also shorter (Piantadosi et al., 2011), improving on a previous observation that more frequent words tend to be shorter (Zipf, 1949)).

In short, there is plentiful evidence, from multiple production levels, that speakers preferentially reduce those linguistic elements which convey more contextually predictable information, wherever this is permitted by the grammar. However, in all of these instances, the intended message is closely linked to the semantic meaning or syntactic structure of the utterance, whether the element encoding that information is clearly or explicitly articulated, or not. Above the level of elements whose length or inclusion only minimally (if at all) alter the literal meaning and syntactic structure of the utterance, it is far less clear whether UID principles apply.

2.1.2 Reduction at the Discourse Level

It's less clear how the principle of UID plays out above the level of lexical/phonological form and markers signaling syntactic structure, where expressed or intended meanings are often:

1. Not clearly tied to specific syntactic structures, grammatical functions, or lexical items, making it more difficult to test whether UID predictions play out;

2. Implied by the totality of the utterance, rather than being a part of the utterance's semantic meaning or structure; or

3. Dependent on material outside the clause for interpretation.

When alternative sets are clearly defined – such as {<null>, <discourse relation marker>} or {<null>, <pronoun>, <proper name>, <definite description>} – alternatives are often not meaning-equivalent as is typically assumed under UID, and the null or reduced forms depend strongly on world knowledge, or the whole discourse context, for interpretation. Further, many things may not necessarily need to be expressed for the listener to infer the intended meaning, provided the context or background knowledge is sufficiently supportive of the intended interpretation. At this level, it becomes a bit more difficult to determine whether UID predictions hold – even if, intuitively, one would expect that very highly predictable meanings (e.g., "the

sky is blue") are less likely to be expressed in isolation than relatively unpredictable meanings.

The best-investigated test case is that of referring expression choice – where a related domain-specific theory – the *Expectancy Hypothesis* (Arnold, 2001) predicts that shorter and less complex referring expressions should be used for more predictable referents. The essence of this hypothesis is supported by the observation that, on average, those factors that predict pronoun usage also tend to be correlated with a higher likelihood of referent mention (Arnold, 2008). However, it is to date unclear whether the predictability of a referent in its *local* context influences choice of referring expression.

This question has primarily been investigated using two research paradigms: one which attempts to correlate referent predictability in corpora with referring expression choice; and one which uses passage completion, with tightly controlled prompts, to determine whether the likelihood of referent mention correlates with referring expression choice. The published results, using both paradigms, are highly mixed, and the discrepancy remains to date unresolved. Details of both paradigms are discussed in Section 5.2, and only empirical observations are discussed here.

Empirically, there is evidence both for (Tily & Piantadosi, 2009; Kravtchenko, 2014; Resnik, 1996, indirectly) and against (Modi et al., 2017) the UID hypothesis in the corpus-based paradigm, which is difficult to reconcile given substantial differences in corpus size, text type, predominant varieties of referring expressions seen in the corpus, and relatively weak results in the positive studies. Similarly, there is evidence both for (Arnold, 2001; Rosa & Arnold, 2017; Bott et al., 2018, p.c. July 25, 2018) and against (Fukumura & van Gompel, 2010; Rohde & Kehler, 2014, Kravtchenko & Demberg, in prep) the same hypothesis in the passage completion literature - similarly difficult to reconcile given to subtle differences in stimulus design, the verb type used in prompts, task type, referring expression ambiguity, and the degree to which the task simulated natural discourse and/or dialogue.

To sum up the argument I make in Chapter 6, I argue that the effects of predictability on referring expression production are relatively weak locally, although there is evidence that the predictability of referents averaged across local contexts influences conventionalized patterns of referring expression choice. These effects, furthermore, appear to arise predominately in highly interactive contexts where the speaker may be prompted to do more audience design; in contexts where speakers are given reason to refer to contextually unpredictable referents; in contexts which better simulate natural discourse; and in contexts where pronouns are maximally ambiguous, for principled reasons explored in Chapter 9.

In short, to the extent that there is a local effect of predictability on referring expression choice, this effect appears to be substantially more weak and fragile than would be expected under the UID paradigm. In Chapter 7, I discuss in more detail some of the reasons why UID may be expected to falter at the discourse level.

2.1.3 Limits of the UID Model

In this section, I discuss what I see as limits of the UID model, as currently formulated. The lack of clear support for the UID model at the discourse level of production may be partially attributed to the increasing difficulty of recovering the intended message as alternate forms of messages become less meaning-equivalent, and recovery of the intended message relies increasingly on pragmatic reasoning and world knowledge. I discuss this problem in the next section.

Second, the lack of clear penalization of redundancy raises questions as to whether this is a soft, violable constraint which is adhered to or violated by speakers, with no ill consequences for the ultimate success of the communicative act. I argue that in fact, there is a 'hard limit' of sorts on how redundant one may be without distorting the pragmatic, if not surface, meaning of the message.

Message Recovery

In UID, communicative success is determined by whether the intended message is transmitted with sufficient fidelity. One of the primary issues with this measure of communicative success is that the degree to which the transmitted message matches that intended is determined *solely* by whether the listener is expected to be reasonably successful in recovering the intended linguistic element and corresponding meaning. However, communicative success is arguably better measured by the degree to which the listener recovers the intended *message*, which includes pragmatic as well as literal interpretations.

That is, communicative success in the UID paradigm displays complete disengagement from the question of message interpretation beyond the literal meaning of the utterance's surface form. This is despite the straightforward observation that any unintended pragmatic meanings triggered by the utterance's semantic meaning and context will necessarily distort the intended message.

This raises two primary questions:

1. How do listeners treat deviations from "optimal" speaker behavior, as defined by UID, which do not impair the recovery of the utterance's surface form or semantic meaning? Do these deviations nevertheless distort the intended message, placing a clear limit on how 'redundant' speakers may be?

2. If so, is it possible to account explicitly for the degree to which the message may be distorted, and the point at which redundancy may begin to distort the intended message, in a UID-like framework?

In the following section, I address the first question in more detail. The second question is addressed in Section 2.3.3.

Penalization of Redundancy

In the standard UID framework, the only consequence of excessive redundancy is that the speaker expends excessive effort on message transmission - an arguably rather non-serious consequence which should minimally impair the success of the communicative act. There is, in fact, abundant evidence that speakers are frequently not as optimally concise as might be expected under these frameworks (Baker et al., 2008; Walker, 1993), and that speakers are in general not as rational in this particular respect as would be expected under a UID framework (cf. Rohde & Kehler, 2014).

There are two likely consequences to excessive redundancy which *may* impair the communicative act, and place a relatively hard limit on how redundant a speaker may be while preserving message fidelity: the transmission of less useful information per time unit, assuming a listener with bounded time, attention, and memory; and triggering unintended pragmatic inferences which alter the message meaning. The first potential limit is not empirically investigated in this thesis, but it remains highly plausible that a given listener is *not* indifferent to the amount of time a message takes to transmit (given limited time to receive a message and act on it), to the length of time that they must sustain attention, and to how much (potentially non-useful) information they may need to store in memory.

The second consequence – that of message distortion through unintended pragmatic inferences – is arguably far more serious. For instance, consider the following utterances:

(1) John went grocery shopping. He paid the cashier.

(2) The sky is blue.

In the case of the first utterance, 1 communicates essentially the same information as the simpler variant, "*John went grocery shopping*". The listener may as a result wonder why the additional information is communicated, and may in fact begin to suspect, for example, that it is somehow unusual and noteworthy for *John* to pay the cashier. Chapter 4 presents empirical evidence that this is in fact the case, and that this degree of redundancy, save situations in which the speaker intentionally utilizes it to communicate something "extra," is truly excessive by virtue of introducing an additional meaning which substantially alters a listener's beliefs about the world.

In the case of the second utterance, 2 communicates something that is *a priori* accepted, and therefore sounds quite odd in isolation. Pragmatically, there may again be a variety of inferences triggered, depending on context: that the speaker believes the interlocutor just said something excessively obvious, that the sky is not usually blue in the speaker's typical location. If those inferences that are triggered are 1) unintended, and 2) communicate something that is not true, this would constitute substantial message distortion. If the listener is simply led to believe that the speaker is excessively redundant, and this is in general *false*, then the listener is liable to

misinterpret future utterances where the speaker utilizes redundancy to intentionally communicate a particular message, as frequently occurs (Walker, 1993).

The critical questions at this point are: how much redundancy is useful to listeners, how much is perhaps suboptimal but harmless, and how much should the theory posit a hard limit on? In Chapter 3, I discuss the empirical evidence for how much redundancy is typical in speech, and apparently tolerated by listeners; and in Chapter 8, I present a formal model which quantifies the effect of redundancy on message interpretation. I further argue that any theory that predicts UID-like effects should explicitly, and preferably formally, account for the influence of redundancy on message transmission.

2.2 Gricean Principles of Rational Communication

In contrast to UID, which concerns itself with the transmission and decoding of the utterance's intended form and semantic meaning, the study of pragmatic reasoning primarily concerns itself with the rules governing the transmission and decoding of non-literal, implied meaning, given a particular surface form. The overall goal of pragmatic theory is to describe and properly account for those principles which determine whether a particular speaker behavior is *rational* from a communicative standpoint, in the sense that the communicative act in its totality may be considered successful. In this respect, UID and pragmatic theory share the same goals, but pragmatics reaches beyond the question of whether a speaker is successful in transmitting the intended utterance and its literal meaning, to also account for the broader goals of communication as a transmission of truthful, necessary, and relevant information.

Pragmatic theory focuses primarily on the relationship between the semantic meaning of the utterance, and the sum total of what the utterance communicates in any given context, which includes its non-literal, or pragmatic meaning. The distinction made by Grice (1975) is between what is *said* ("*sentence meaning*"), and what is *mean* (*speaker meaning*). *Speaker meaning* is inferred by the listener by taking into account the general principles of successful communication, any alternative ways that the speaker might have expressed the same meaning, and the context in which the message was communicated.

For an example of how pragmatic reasoning may proceed, and a non-literal meaning may be inferred, consider the following:

(3) *Some* of the students passed the test.

(4) *Not all* of the students passed the test.

Utterance 3, at the level of truth-conditional meaning, communicates only that at least one of the students has passed the test. However, pragmatically, the utterance is typically interpreted to implicate 4: at least one of the students has passed the test, but not all of them did. This implicated meaning may be computed as follows: if, in

fact, *all* of the students had passed the test, the speaker could just have easily and efficiently (and unambiguously) said as much. The fact that they did not select this equally efficient but more informative utterance is strong evidence that the speaker either knows that *not all* of the students passed the test, or at minimum that the speaker does not have enough evidence to assert that all of the students passed. While one could argue that 4 is part of the literal meaning of 3, this view is difficult to reconcile with the coherence of the following sentence, which would otherwise be read as self-contradictory:

(5) Some, and in fact all, of the students passed the test.

The task of explaining exactly how such inferences arise, and which principles of communication are considered in inferring non-literal meanings, has primarily been tackled by Gricean, neo-Gricean, and Relevance theories.

2.2.1 Gricean and Neo-Gricean Programmes

Gricean and neo-Gricean programmes aim foremost to describe those principles of communication which speakers strive to adhere to, and to account for how apparent or potential deviations from those principles give rise to non-entailed, pragmatic meanings. First I will cover the basics of the original Gricean framework, and then I will provide a brief history of neo-Gricean approaches, which differ primarily in the specific taxonomy and priority that they assign to particular communicative principles, and/or in how they seek to account for the existence of said principles.

Although the distinction between literal and non-literal meaning did not originate with Grice (1975), he was among the first to propose a systematic program to account for how and why pragmatic inferences arise, as well as proposing a taxonomy of communicative principles and varieties of pragmatic inferences. The primary argument behind this programme is that speakers and listeners share a single joint goal: that of successful communication. The *rational* speaker aims to achieve this goal, which requires following a series of basic principles which increase the odds of the communication act being successful. The listener, further, generally assumes the speaker to be *rational* unless proven otherwise.

In Grice's framework, the aim of successful, rational communication is formulated by the *Cooperative Principle*, which dictates at a high level the basic conditions that a rational communicative act must meet:

> Cooperative Principle: "Make your contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged."
>
> – (Grice, 1975, p. 41)

This general principle is adhered to by following four *Maxims of Conversation*:

> Maxims of Quality: Try to make your contribution one that is true. (i) Do not say what you believe to be false. (ii) Do not say that for which you lack adequate evidence.
>
> Maxims of Quantity: (i) Make your contribution as informative as is required for the current purposes of the exchange. (ii) Do not make your contribution more informative than is required.
>
> Maxim of Relation: (i) Be relevant.
>
> Maxims of Manner: Be perspicuous. (i) Avoid obscurity of expression. (ii) Avoid ambiguity. (iii) Be brief (avoid unnecessary prolixity). (iv) Be orderly.
>
> – (Grice, 1975, pp. 45-46)

The comprehender, by default, assumes that the speaker is obeying these maxims. In the case where, given a particular utterance, one or more maxims would be violated if the literal meaning of the utterance was all that the speaker was intending to convey, the comprehender infers what else the listener may have intended to convey (or what would have to be true of the world), for the maxims (or, at least, the general communicative principle) to hold.

For instance, the implicature 4 arises from 3 due to the fact that 3 would violate Quantity-I, were the speaker to have used a less informative expression – *some* – to communicate the fact that *all* of the students passed the test. In this manner, Grice systematically accounts for how non-literal meaning, or a set of inferences and altered beliefs about the world, arises in response to observed speaker behavior, and the truth-conditional meaning of the speaker utterance.

Principles of Conversational Implicatures

The primary object of interest in the Gricean Programme is the *conversational implicature*, which is that part of utterance meaning which is implied, given the principles of cooperative communication, but not stated directly. Implicit in the term *implicature* is the notion that this implied meaning is something that the speaker (consciously or unconsciously) *intends* for the listener to comprehend – rather than a meaning that incidentally arises through the use of a particular utterance, in a particular context, but which is not intentionally communicated by the speaker. Grice defined an *implicature* among other things, as a non-truth-conditional aspect of *meaning*:

> S means p by 'uttering' U to A if and only if S intends: (i) A to think p, (ii) A to recognize that S intends (i), and (iii) A's recognition of S's intending (i) to be the primary reason for A thinking p.

Conversational implicatures are contrasted with *conventional implicatures*, which are indelibly associated with specific lexical items or syntactic construction – for

example, the connective *but*. These are cases where, although two alternative utterances (e.g., *and*, *but*) are truth-conditionally equivalent, one conventionally carries an additional meaning, independent of context. The *but* connective in 6 does not introduce any additional conditions on the sentence being true, compared to 7, and yet it clearly communicates that the state of being *unmotivated* is unexpected in the context of being *intelligent*:

(6) He is intelligent but unmotivated.

(7) He is intelligent and unmotivated.

Aside from non-conventionality, there are two other criteria that implicatures (and pragmatic inferences, in general) must meet:

Calculability Unlike *conventional implicatures*, *conversational implicatures* can, and in fact must, be calculable by the listener. I.e., the listener must be able to arrive at the intended meaning through a well-defined, motivated reasoning process, on the basis of the conversational maxims[2]. That is, one can construct a straightforward process by which, given a particular utterance, the comprehender will be able to straightforwardly deduce that this utterance may *implicate* a particular non-literal meaning, assuming the cooperative principle holds. To note, this does not imply that listeners explicitly follow such a reasoning process every time they infer a non-liberal meaning, but rather specifies a set of conditions that must hold for this to successfully occur.

Cancellability Critically, conversational implicatures are *cancellable*, which distinguishes them from the truth-conditional, semantic meaning of an utterance. For instance, 9 is a part of the truth-conditional meaning of 8. This meaning cannot be canceled without producing a contradiction, 10:

(8) John is a poodle.

(9) John is a dog.

(10) John is a poodle, but isn't a dog.

In contrast, an implicature can be canceled, by negating the utterance's implied meaning explicitly, without creating a contradiction:

(11) John got some of the poodles.

(12) John didn't get all of the poodles.

(13) John got some of the poodles – in fact, he got all of them.

[2]The degree to which conversational implicatures closely associated with specific lexical items or constructions must be calculated, as opposed to *presumed*, is currently under debate (Levinson, 2000), but outside the scope of this thesis.

Implicatures can further be canceled if they appear in a context where the normally-implicated meaning is not necessary for the utterance to follow the maxims of communication:

(14) Did anyone get any poodles?

(15) Bill got some of the poodles.

(16) # Bill didn't get all of the poodles.

These conditions for implied vs. conventional meaning become relevant in Section 4.2.1, where I ensure that the message to be inferred by listeners is strongly implied by the preceding context, but not entailed, and therefore cancellable when the context supports a different interpretation.

Generalized vs. Ad-hoc Implicatures

Conversational implicatures are typically split into two varieties: the *generalized* implicature, and the *ad-hoc*, or *particularized* implicature. This distinction concerns the degree to which the implicature is presumptively computed, independent of context. Generalized implicatures are thought to arise by default, often but not always associated with the use of specific lexical items, except in specific contexts which fail to license them.

Particularized or ad-hoc implicatures, in contrast, are typically one-off inferences which arise only on very specific linguistic and background contexts. In contrast to generalized implicatures, these have received very little attention in the theoretical and experimental literature. Although I will not tackle the nature or validity of this distinction in my thesis, the implicatures I use, in contrast to generalized implicatures, are not associated with the use of either particular lexical items, or otherwise unusual constructions which typically trigger inferences by default.

Generalized conversational implicatures typically include scalar implicatures and M-implicatures. These have received the most attention in the literature to date, due in part to their systematic association with specific lexical items, constructions, and unusual formulations of typical messages; and to how predictably they arise in a wide variety of contexts. The basic reasoning behind scalar implicatures is that the speaker uses a *less* informative expression in lieu of a *more* informative one, therefore implicating that the meaning of the more informative expression does not hold (e.g., *some* implicates *not all*, *warm* implicates *not hot*, and so forth). The reasoning behind M-implicatures is that the speaker uses an unusual and typically more costly way of expressing a particular message – for example, *I caused the car to start* in lieu of *I started the car* – therefore implicating that the car was not started in a normal manner.

Ad-hoc, or particularized implicatures typically include those inferences which are not associated with particular lexical items, constructions, or alternate formulations,

and which arise only in specific contexts, rather than by default. For example, in the following examples, which message 19 is implicated by 18 depends on context 17:

(17) Are you cold / Do you want to talk?

(18) Please close the window.

(19) I'm cold / I don't want anyone to overhear what we're saying.

As mentioned, to date most of the pragmatic literature has focused on generalized implicatures – and specifically, scalar implicatures. Non-scalar generalized implicatures are largely limited to M-implicatures, which may arise fairly systematically, but have sufficiently variable interpretations that they are difficult to test. Particularized implicatures are similarly difficult to study, due to their extreme context dependence, and the wide variety of implicatures any one utterance may trigger, particularly given relatively limited context, as is typical in experimental work.

It is therefore difficult to say much about particularized implicatures *in general*. This is, I argue, unfortunate, as the process of pragmatic inference arguably remains poorly understood, so long as what arguably constitutes the vast majority of pragmatic inferences undergoes no systematic or empirical study. Apart from the difficulty of studying them, particularized implicatures are also not as informative to the distinction between literal and non-literal meaning as are generalized implicatures, as their extreme context-dependence easily distinguishes them from the literal meaning of an utterance.

Implicit in the definition of a conversational implicature is that the meaning is *intentionally* communicated by the speaker, rather than being an accidental byproduct of context and background knowledge. The terms *implicature* and *inference* are often used interchangeably, and the same principles for distinguishing a truth-conditional vs. pragmatic meaning apply. However, some meanings are more likely to be intentionally communicated than others, and the specific modeling of pragmatic phenomenacrucially hinges on intentionality. I discuss this distinction further in the next subsection.

Non-implicature Inferences

Although the Gricean programme is primarily concerned with the transmission of intentionally implied pragmatic meaning, it is by no means limited to it. Pragmatic reasoning also plays a similar role in the generation of non-implicated non-literal meaning. For instance, if the implied meaning of a speaker utterance is pragmatically inconsistent with a comprehender's prior beliefs about the common ground, the comprehender may resolve this conflict by revising their common ground beliefs (cf. Degen et al., 2015). This does not necessitate that the speaker *intended* to communicate that the common ground is in some way different from what the comprehender assumed, but only that the speaker and listener's prior beliefs about the common ground diverge.

In the following example, taken from Degen et al. (2015), the literal meaning of 20, 21 is inconsistent with a naive comprehender's background beliefs, which is that marbles will (almost) invariably sink in a pool:

(20) John threw the marbles into the pool.

(21) Some of the marbles sank.

At this point, the listener has two options for how to resolve this inconsistency: they can either assume that the speaker is violating Quantity-I by using a less informative expression where a more informative one (*all*) would be more appropriate; or they can revise their background knowledge such that marbles (in this world) do *not* invariably sink into pools, which removes the apparent Quantity-I violation. What I term *common ground inferences* may arise where the speaker and listener's assumptions about the common ground differ – perhaps the speaker is unaware of what the listener believes, or erroneously believes that the listener shares their own beliefs.

This strategy of resolving a maxim violation has received very little attention in the pragmatic literature – possibly because until recently, there was relatively little notion of background speaker and comprehender beliefs, which could systematically be updated given evidence that contradicted them. Attention to the notion and importance of beliefs about the context and background knowledge seems to have increased with the advent of probabilistic and constraint-based models of pragmatic reasoning. The notion of belief change, however, mirrors the more well-known concept of presupposition *accommodation* (Karttunen, 1974; Stalnaker, 1974) – where a listener must accept as true any propositions entailed, but not explicitly asserted by, any given statement (e.g. '*The king of France is bald*' presupposes that there exists a *king of France*, in the first place).

In the realm of pragmatic inferences, the most well-known and systematic investigation of how background beliefs may be altered, in order to avoid assuming a maxim violation, is Degen et al. (2015). This work investigated comprehenders' willingness to revise their prior assumptions about the assumed common ground, in response to utterances whose default pragmatic implications would otherwise be inconsistent with it. They found that background assumptions about the world are surprisingly defeasible: comprehenders frequently accommodate the pragmatic meaning of utterances such as "***some*** of the marbles sank" (upon being thrown into a pool), by assuming that the utterance must signify a strange scenario where physics doesn't quite work as expected (perhaps the pool is filled with a viscous liquid, rather than water). The assumption, of course, is that not all of the marbles sank – had ***all*** of the marbles sunk into the pool, the speaker would have said so, instead.

Another example of how background beliefs may be altered, in response to what would otherwise constitute a maxim violation, is explored in Chapter 3. Next, I discuss two neo-Gricean programmes which aimed to expand on Grice's taxonomy of communicative norms, as well as to ground its existence in other principles.

The Hornian System

Although the programme proposed by Grice remains in heavy use, there have been multiple attempts since that original proposal to simplify and refine the taxonomy of principles and communicative norms introduced, as well as, in some cases, to ground them in more general cognitive principles. The most well-known of these are the Hornian and Levinsonian systems, as well as Wilson and Sperber's Relevance Theory, discussed in more detail below.

The Hornian System (Horn, 1984), like UID, has its roots in the Zipfian concept of the *Principle of Least Effort*, which in the context of language means balancing speaker economy over the accuracy and success of transmitting and intended message (Kasher, 1976, makes a similar proposal). Horn proposed two opposing principles: the Q- (Quantity) Principle, and the R- (Relation) Principle, which subsumed Grice's maxims:

a. The Q-principle: Make your contribution sufficient; Say as much as you can (given the R-principle).

b. The R-principle: Make your contribution necessary; Say no more than you must (given the Q-principle).

The Q-principle can be considered as analogous to the Zipfian or UID principle of saying as much as is necessary for the listener to correctly interpret the intended message (but no more, given the R-principle). The R-principle is intended to be analogous to the UID principle of saying *no more* than is necessary for transmission of the intended message, in the interest of speaker economy. This produces a pragmatic division of labor:

> The use of a marked (relatively complex and/or prolix) expression when a corresponding unmarked (simpler, less 'effortful') alternate expression is available tends to be interpreted as conveying a marked message (one which the unmarked alternative would not or could not have conveyed).
>
> (Horn, 1984)

Although Horn's proposal gives a first glance of what a theory of pragmatic reasoning, explicitly incorporating UID-like concepts and motivations, may look like, the theory proposed is not formally defined, and unlike UID makes no clear quantitative predictions. Further, when considering the question of information transmission, and what constituted excessive vs. insufficient information, Horn took *information* to refer both to the amount of signal transmitted, and the amount of new semantic information – arguably convoluting the two (Levinson, 2000).

The Levinsonian System

The Levinsonian System (Levinson, 2000) mainly aimed to provide a novel taxonomy and clarification of some of the principles underlying pragmatic inference. Unlike the Hornian system, it did not seek to ground these principles in any general cognitive mechanism. Further, unlike Horn, Levinson sought to separate the concepts of semantic informativeness, and how much signal is transmitted by the speaker. Levinson proposed the following three principles, each of which has a speaker's maxim, governing what the speaker out to say, and a comprehender's maxim, which governs what the comprehender should infer:

> The Q-principle (simplified): Speaker: Do not say less than is required (bearing the I-principle in mind). Addressee: What is not said is not the case.

The Q-principle concerns semantic informativeness, and dictates that given a set of varyingly strong semantic alternatives (e.g., <some, all>), the strongest or most informative one should be used whenever it is licensed. In this context, *all* is the most informative, or least informationally ambiguous (whereas *some* may mean *all*, or may mean *some, but not all*). Correspondingly, the listener should assume that if the speaker has not said *all*, then *all* must not be licensed in that context.

> The I-principle: Speaker: Do not say more than is required (bearing the Q-principle in mind). Addressee: What is generally said is stereotypically and specifically exemplified.

The I-principle effectively mirrors the Q-principle, and dictates that, if a speaker uses a semantically general, or vague, expression, this expression will be associated with the most likely or stereotypical situation that might be associated with such an expression.

> The M-principle: Speaker: Do not use a marked expression without reason. Addressee: What is said in a marked way is not unmarked.

The M-principle concerns cases where two different forms may be used to describe the same situation (e.g., *He started the car*, *He caused the car to start*). It dictates that, given a stereotypical situation, the least marked expression should be used. Correspondingly, if a marked (and typically, lengthier) expression *is* used, the listener should assume a non-stereotypical situation.

Interactions of the three principles produce the various implicatures. Although this system does not interface as cleanly with UID or Zipfian principles as the Hornian system does, it has the advantage of separately considering the influence of semantic informativeness and signal redundancy on production and interpretation.

2.2.2 Relevance Theory

Relevance Theory (Sperber & Wilson, 1986, 1995) differs from the aforementioned systems in that it seeks to explain everything through a refined supermaxim of *Relation*. To this end, they propose an overarching cognitive principle, which dictates that the speaker's foremost goal is the maximize the *relevance* of an input to a listener:

> Sperber & Wilson (1995)'s cognitive principle of relevance: Human cognition tends to be geared to the maximization of relevance.

Relevance is defined as the balanced maximization of *cognitive effects* (influence of the new information on existing beliefs), and the processing effort required to understand and situate the input. *Cognitive effects* are grouped into those which generate new beliefs by virtue of old information and novel input; those that strengthen existing beliefs; and those which contradict existing beliefs. For example, assume that someone is interested in whether *John* went bungee-jumping, or not. There might be three truth-conditionally-equivalent ways of expressing the answer, assuming that *John* did, in fact, go bungee-jumping:

(22) John went bungee jumping.

(23) Either John went bungee jumping, or the earth doesn't orbit the sun.

(24) John participated in a sporting event.

In this case, the first two provide the same information, but 23 requires much more work on the part of the listener to interpret the input, to no positive cognitive effect. 24, further, does not have as much *cognitive effect* as 22, while engendering a similar amount of processing difficulty. On the other hand, the following answer to the question, at face value, expends some amount of processing effort in exchange, to no positive cognitive effect, as what is stated at face value is a truism:

(25) The earth orbits the sun.

In this case, 25 may for example be interpreted as having positive *cognitive effect* by suggesting that *John's* having gone bungee-jumping is as self-evidently true as the earth orbiting the sun (in case this inference is consistent with background knowledge). In Relevance Theory, communication that comprehenders may draw inferences from is termed *ostensive-inferential communication*, and utterances which are specifically meant to signal to a comprehender a speaker's intent to inform them of something are considered *ostensive* stimuli. These utterances not only inform, but specifically signal that the speaker *wants* to inform the listener of something.

In other words, *ostensive* stimuli are specifically designed by the speaker to grab the comprehender's attention, and to make the comprehender aware that their attention is being purposefully grabbed, presumably for some specific purpose. The use of *ostensive* stimuli generates expectations of *optimal relevance* on the part of the comprehender:

> Presumption of optimal relevance: a. The ostensive stimulus is relevant enough to be worth the audience's processing effort. b. It is the most relevant one compatible with communicator's abilities and preferences.

As Wilson & Sperber (2004) note, by producing an *ostensive* stimulus, "the communicator therefore encourages her audience to presume that it is relevant enough to be worth processing." The more glaringly obvious the message, and the more purposeful energy the speaker expends in producing it, the more likely the comprehender is to pay attention, and to presume that the message is *relevant*.

In contrast to the relatively formal systems described above, Relevance theorists sought explicitly to create a *cognitive theory* of pragmatic reasoning, rather than simply refining the proposed taxonomy of, and relationships among, the various maxims of communication. While Relevance Theory suffers from making very few explicitly defined and empirically testable assumptions, it cleanly captures intuitions about phenomena of interest, Particularly relevant is the observation that more *ostensive* stimuli, as a function of being more likely to be noticed by comprehenders in the first place, and more likely to be interpreted as having particular communicative intent, should result in stronger or more reliable inferences. These will be explored more in Chapter 3.

2.2.3 Information-Theoretic vs. Gricean Rationality

Both information-theoretic models of utterance choice, and Gricean models of pragmatic reasoning, are fundamentally based in the idea that speakers are *rational*. However, what they term rationality is conceptualized somewhat differently. Rationality within the information-theoretic framework refers to the idea of the *ideal* or *rational* speaker, who behaves such that the (implicitly literal) message that they wish to get across is transmitted accurately, and with the least amount of effort necessary. There is no formal determination of what degree of *trade-off* of this sort is strictly *optimal*, but there is a clear prediction that the likelihood of producing a reduced linguistic element (or omitting it entirely), as empirically measured, should be inversely proportional to the likelihood of the semantic meaning or grammatical function of that form in context, which is formally defined, and can be estimated empirically through a variety of methods.

The Gricean conception of rationality comes down to the general idea of human action being *goal-directed*. In the context of communication, *goal-directedness* indicates that one behaves in a manner that facilitates, rather than hinders, communication. Further, this *goal-directedness* is assumed by comprehenders, who attempt to interpret everything a speaker utters in the light of this assumption. The account of what variety of behavior it is, precisely, that facilitates communication is largely descriptive, and not grounded in general cognitive principles, nor is it formally defined. Its aim is largely to make qualitative predictions about how speakers interpret utterances, given their assumptions of what constitutes rational behavior. Some of the

general principles appear to overlap: for instance, the Quantity maxim states that speakers should provide no more, and no less information than is needed for listeners to understand the intended message; and the Manner maxim states that speakers should not be excessively long-winded.

However, it is difficult to couch some of the remaining Gricean principles in UID terms. As discussed in Section 2.1, the speaker goal in models such as UID is to transmit an intended message with maximum fidelity. Within the Gricean model, the speaker goal is rather to engage in successful communication *more generally*, which includes ensuring that the listener comes away with accurate, truthful ideas about that which the utterance describes more generally, as well as any background beliefs about the common ground. As an example, consider the possibility that a speaker utters the following:

(26) John is outside.

The UID model explicitly predicts only that the speaker will choose that phrasing of this proposition that takes the least effort while reliably getting the proposition across to the listener. The Gricean model, however, for one dictates that this proposition should be *true* of the world, relevant to the communicative task at hand, and the most typical way of phrasing such a proposition. Optimally, one may want to reconcile the two, such that one has a formal model which makes reliable, quantitative predictions about speaker behavior, given a set of alternative ways of formulating a message, as well as taking into account factors such as the true state of the world, as well as the overall aim of the task.

One method for reconciling the two is to use a probabilistic framework, which incorporates both speaker and listener beliefs about the world and the general aim of the communicative act; and the effect of utterance cost/effort on production, given the likelihood of a particular meaning in context. Further, this framework should allow these beliefs to undergo revision, and make clear predictions of how listeners may interpret various utterances (as well as the state of the world), given their starting beliefs and belief revisions.

2.3 The Rational Speech Act Model

The Rational Speech Act (RSA) model originates in the attempt to use probabilistic Bayesian models, which had been successful in modeling rational, goal-directed behavior in other areas of cognition, for the purpose of describing how a rational speaker and listener may reason about each others' beliefs and intentions under uncertainty (Frank & Goodman, 2012). The core idea behind the Gricean programme, as discussed above, is that speakers are rational actors who pursue a specific communicative goal, and that comprehenders interpret speaker behavior in light of that assumptions. However, the Gricean programme gives no clear way of formalizing these assumptions,

or of making graded, quantitative predictions which interface cleanly with empirical data.

Although the RSA model was not intended to serve as a full reconciliation of Gricean principles and UID-like models of language production, it nevertheless accomplishes this aim in part. Speakers select utterances on the basis of their cost and utility in unambiguously communicating a certain meaning to a literal listener, and make the choice between alternatives close in meaning by considering their costs, relative to how unambiguously they communicate the meaning. Listeners attempt to interpret the utterance by taking into account utility and cost-based constraints on the speaker's choice, given the meaning(s) the speaker may have wanted to communicate, thus deriving the most likely intended meaning. Critically, however, the choice between meaning-equivalent alternatives in the standard RSA model is not modulated by the predictability of the information conveyed in context – but rather, only by the utterance cost. The solution to this issue is discussed further in Section 7.3.

One of the main shortcomings of the RSA model, compared to information-theoretic utterance choice models such as UID, is that there is no allowance for the possibility of an utterance being misperceived, as is critical for theories based on the notion of the *noisy channel*. This lack results in the RSA model failing to reflect a critical empirical fact: that less detailed and effortful means of expressing the same message are more likely to be chosen when the message meaning is less predictable. As I further demonstrate in 7.2.3, this also results in the RSA model failing to accurately predict the pragmatic inferences that may be drawn from a more effortful utterance being used where a less effortful one would have been sufficient. This shortcoming of what I term the clean-channel RSA model is mathematically proven in Bergen et al. (2016).

That said, in other ways, the RSA model presents a significant improvement over the UID model, in that the speaker is concerned not only with transmitting the literal meaning of a particular utterance, but with instilling particular *beliefs* about the world in their listener, which requires taking into account the pragmatic interpretation that a listener may assign to any particular utterance. The study of pragmatic reasoning makes it clear that simply recovering the utterance intended by a particular speaker, and its truth-conditional meaning, does not ensure an accurate understanding of the speaker's communicative intent.

In the following subsections, I present the basic RSA model, and its applications, and then the more complex joint reasoning model, and its applications. The noisy channel RSA model is discussed in detail in Chapter 8, and is presented as a straightforward solution to the failure of the clean-channel RSA model to account for those effects predicted by information-theoretic models of utterance choice.

2.3.1 Base RSA Model

Here, I will first outline how the base RSA model works. I will then talk of several successful applications of this model.

Take into consideration a scenario such as that presented in Frank & Goodman (2012), where the speaker must successfully refer to one of three colored geometric shapes, and the listener must infer which of the three geometric shapes the speaker is referring to. In this scenario, two different shapes share a color (e.g. *green square* and *green circle*), and a third shape that differs in color, but shares the shape of one of the preceding items (e.g., *blue square*). Each of the objects also has some independent prior likelihood of being referred to, perhaps due to its salience – for example, listeners may expect that the *blue square* is most likely to be mentioned, given its perceptual prominence.

In this task, the speaker is given the constraint of only using one word (color, or shape) to refer to the intended object, and must maximize the likelihood that the listener infers the intended object, given the utterance, and the listener's prior expectations that the object will be mentioned. Similarly, the listener must infer the object most likely referred to by the speaker, given the likelihood of the speaker having used that utterance to refer to any particular object, and the prior likelihood of the object being referred to.

At the first level of this model, a literal listener interprets the literal meaning of a given utterance, by reasoning about the truth-conditional meaning of the utterance, considering the prior likelihood of the world state the utterance refers to, where the utterance remains uninformative. For example, given a green *circle*, a green *square*, and a blue *square*, if the speaker simply says "*green*" to indicate one of three objects, the listener will use the semantic meaning of the utterance to limit themselves to the first two objects, and then take into account the prior probability of the speaker referring to either the *circle* or the *square*, to determine the most likely meaning.

At the second level, a pragmatic speaker selects an utterance, among several alternatives, based on two considerations. First, they consider the likelihood of the listener inferring the intended message, given the utterance – or, the expected utility of the utterance. Second, they balance this consideration against the cost of the utterance, typically measured by features such as the length of the utterance, the presence of additional articulatory effort (by way of unusual prosody), and so forth. For example, the speaker will consider the likelihood of the literal listener inferring reference to the *green square*, as in the scenario above, if they say either *green* or *square*, taking into account the utterance's literal meaning, the relative utility of the utterances in unambiguously identifying a given object, and the prior likelihood of the object being referred to. In this scenario, the speaker is unlikely to use *square* to refer to the *blue square*, as in that case the utterance *blue* would have higher utility in unambiguously identifying that object. Similarly, *green* is unlikely to be used for the *green circle*, and *circle* would identify that object unambiguously.

At the last level, the pragmatic listener takes a pragmatic speaker's utterance, and reasons about the likelihood of the speaker having chosen that particular utterance to refer to a given object, taking into account the independent prior likelihood of the object being referred to. In other words, given the word *green*, they reason about the likelihood that the speaker (taking into account the literal listener and utterance cost) would have chosen that word to indicate either the *green square* or *green circle*, again also taking into account whether there are alternate utterances the speaker would have more likely chosen, given a particular intended referent, and the prior likelihood of the object being referred to. In this case, the listener may reason that *green* is more likely to refer to the *square*, as the speaker could have used the word *circle* to refer to the circle unambiguously.

The base RSA model has been used to successfully predict speaker and listener behavior in reference, or signaling games such as the one described above. Frank & Goodman (2012) used a web-based paradigm to display a series of shapes similar to those above to participants, and either ask them which expression (a one-word description of color, shape, or texture) would be most likely to be used to refer to a given object, or to guess as to which shape a given one-word description referred to. The model predictions in this case showed a tight fit to empirical results. Qing & Franke (2015) replicated these results using a modified experiment configuration.

Most work, however, focuses on listener interpretations, rather than speaker behavior. Aside from reference games such as the above, RSA has been used to account for classic inferences such as the *some, but not all* interpretation of *some*. The model used by Goodman & Stuhlmüller (2013) predicted the pragmatic interpretation of terms such as *some* and *all*, as well as that of number words. By manipulating the extent of the speaker's knowledge, they were also able to predict the effect of speaker uncertainty on a listener's interpretation of whatever term they used. For example, perhaps the speaker only saw two out of three apples, and saw that those two were red, and then chose to announce that *some of the apples are red*. In this case, *some* does not indicate that not all of the apples are red, but rather indicates that the speaker does not have complete knowledge of how many of the apples are red (but only knows that at least some of them are).

2.3.2 Joint Reasoning Model

More complex inferences, which may include the listener reasoning in parallel about the speaker's background knowledge or beliefs about the world, the specific communicative goals, the word meanings, or the question under discussion (for example), can be represented by joint reasoning models of varying complexity. An uncertainty about speaker goals, and the context of the communicative act, may make for a more realistic model. This model type is discussed in more detail in Section 7.2.

The joint reasoning model has for instance been successful in predicting the interpretation of hyperbolic expressions (e.g., "*This kettle cost $1,000,000!*"), where the

listener reasons both about the price of the *kettle*, and about the speaker's more likely intent (to communicate the literal price, or to communicate their attitude towards the price) (Kao et al., 2014b). For instance, $1,000,000 is a very unlikely price for a kettle, and in this instance, the listener may reason that it is far more likely that the speaker is attempting to communicate an affective state. Kao & Goodman (2015) have used the same model to account for interpretation of verbal irony, and a similar model to account for interpretation of metaphor Kao et al. (2014a), where the listener reasons about the characteristics of the person being described, and how that informs interpretation of the metaphor.

Another class of models considers the possibility that there are *thresholds* for the use of certain expressions, such as *expensive*, or *tall*, which are inherently relative to the context or items under discussion. In these models, the (for example) price or height distribution of the relevant items is taken into account for the listener, who thus infers what the more likely threshold for referring to something/one as *expensive* or *tall* is (Lassiter & Goodman, 2013). Similar models have been used to account for the interpretation of quantifiers such as *few* and *many* (Schöller & Franke, 2017).

Further phenomena have been successfully described using joint reasoning RSA models, such as the use of embedded implicatures (e.g., Bergen et al., 2016; Potts et al., 2015), but they are beyond the scope of this background review.

2.3.3 Further Development of the RSA Model

Models which deviate from this general template may also take into account the possibility that:

1. Semantic meaning is underspecified; e.g., the meaning is 'fuzzy,' rather than fixed (Bergen et al., 2016).

2. The surface form of the utterance that the speaker attempts to transmit – given the basic notion of the *noisy channel* – may not be the one that the listener receives (Bergen & Goodman, 2015).

The latter possibility proves crucial towards accounting for those inferences that arise from more effortful utterances being used where less effortful ones may have sufficed. In other words, it reflects the Relevance-Theoretic idea that more *ostensive* stimuli will result in stronger and more reliable inferences, if only because the listener is more likely to notice them, in the first place, and to wonder why the speaker put additional effort into communicating a given meaning. I return to this model in Chapter 8.

The two models most relevant to my work are the joint reasoning model presented in Degen et al. (2015), and the noisy channel model presented in Bergen & Goodman (2015). I discuss both in more detail in Chapter 7, but provide a brief summary here. Degen et al. (2015) shows that it is possible to model pragmatic inferences

about common ground knowledge, where a listener reasons jointly about the current world state, and the speaker's background beliefs about the world (which inform the current world state, and vice versa). In their model, as well as the empirical data it accounts for, listeners revise their prior beliefs about the world, in light of utterances that would be pragmatically odd if they were to stick to their current assumptions.

Bergen & Goodman (2015) show that it is possible to model pragmatic inferences which arise from a listener assigning additional meaning to especially effortful or perceptually prominent input. This is accomplished by incorporating the notion of the noisy channel into a basic joint reasoning RSA model, which ensures that more effortful and perceptually prominent utterances are most likely to be perceived accurately, and to prompt the listener to reason as to why increased accuracy in message transmission was needed. In Chapter 7, I also argue that this mechanism can apply to higher-level reasoning about the likelihood of a given utterance grabbing a listener's attention and being correctly stored in memory, rather than simply the likelihood of it being perceived correctly. This helps capture the intuition that listeners assign stronger inferences to more attention-grabbing, or *ostensive* input, even when said input does not strictly provide any additional information to the listener (Wilson & Sperber, 2004)).

2.4 Summary

In summary, in the realm of human communication, one can currently find several major accounts for what it means for a speaker to be *rational*. One approach – UID, and similar information-theoretic models of utterance choice, argue simply that speakers aim to expend the least possible effort to correctly transmit a message and its underlying truth-conditional meaning to a listener. The benefits of this approach are that it defines rationality in clear, quantitative terms, and similarly makes clear empirical predictions. The shortcoming of this approach is that it does not account for the overall goal of the speaker to ensure that the listener's interpretation of the utterance is in fact the intended one – a somewhat difficult approach to defend, given the ubiquitous presence of pragmatic reasoning.

Another approach – a Gricean framework of maxims which must be followed by a rational speaker to ensure successful communication – primarily defines speaker rationality as the degree to which a speaker treats communication as a goal-directed activity. The benefits of it are that it takes into account the overall goal of communication as a means for speakers to transmit accurate beliefs about the world to their listeners, cleanly accounts for why comprehenders may interpret some utterances non-literally, and makes qualitative predictions about which utterance interpretations may arise in particular contexts. The shortcomings of this approach are that it is not formally defined, and therefore does not make clear quantitative predictions; it is for the most part descriptive, and not grounded in clearly defined underlying principles; and

finally, it does not always account for why certain apparently licensing conditions do not in fact trigger inferences, or why some conditions trigger weaker inferences than others.

The Rational Speech Act model is a partial reconciliation of both approaches – it predicts, given a world state, a set of alternative utterances, and the corresponding utterance costs, which communicative choices a speaker will make, and how listeners will interpret those choices, given the same. While the phenomena accounted for UID essentially fall out of the RSA framework – provided the model assumes a noisy channel – the RSA framework also accounts for general communicative goals (such as transmitting accurate beliefs about the world). Furthermore, by explicitly taking into account the possibility that listeners consider which underlying beliefs and world states are most likely to prompt certain utterance choices, it introduces the possibility that choosing one linguistic alternative over another will have different consequences for a listener's beliefs about the underlying world state – beyond simply their ability to correctly recover the underlying form and truth-conditional meaning of the utterance. This, for instance, places a clear limit on how redundant a speaker may be without distorting the intended meaning of an utterance.

Overall, the RSA framework clarifies the costs of excessive redundancy, on the part of the speaker (see Section 8.5), in a manner that simply isn't captured by UID, but which is empirically verifiable (see Chapter 4). It clearly formalizes the component pieces of message production and interpretation, including underlying beliefs about the world which guide the latter, giving it the ability to account for utterance choice beyond predicting which of two meaning-equivalent utterances is most likely to be used in context. It is able to explicitly reflect the factors which drive UID-postulated speaker behavior, and to make the same predictions regarding utterance choice that UID does, while going beyond it to demonstrate the additional costs, in terms of message distortion, of explicit redundancy.

Because the RSA framework does not necessarily assume perfect rationality on the part of speakers, it is also able to straightforwardly account for cases where speakers fail to exhibit rational behavior in less interactive or naturalistic contexts, but do so in more interactive or naturalistic contexts. In short, while the RSA framework can account for the same phenomena that UID does, it goes far beyond UID in accounting for a wide range of limits and influences on utterance choice.

Chapter 3

Informationally Redundant Utterances: Background

Efficiency-based theories of language production, such as the Uniform Information Density (UID) hypothesis, predict that utterance choice among meaning-equivalent alternatives is governed by a speaker's desire to conserve effort on the one hand, and to transmit the intended meaning of the utterance on the other. As discussed in Section 2.1.3, these theories do not inherently penalize redundancy from the point of view of communicative success, and assume that avoidance of redundancy is dependent solely on a speaker's desire to be maximally concise. Assuming that speaker redundancy may alter the interpretation of the intended message, these theories also have no mechanism for discussing or representing *how* redundancy may influence the success of the communicative act, as they concern themselves solely with whether the truth-conditional meaning of the intended message was successfully transmitted. However, here I address the possibility that excessive redundancy, as suggested by pragmatic theory (Grice, 1975), influences how a comprehender interprets the received utterance, potentially distorting the intended meaning of the utterance message. In the following chapter, I present empirical evidence for this. I finally argue that a comprehensive theory of utterance choice, concerned with communicative act success, must reflect the effect that utterance choices have on message transmission *beyond* the semantic meaning of the utterance.

Redundancy may appear at any level of production, and most pragmatic theories, as well as theories of language processing, typically include constraints against redundancy (beyond what is deemed useful to the listener). Below the discourse level, redundancy generally includes overt mention of, or increased articulatory effort towards producing material that is easily predictable or recoverable in context. In other words, more signal is provided than the comprehender requires to accurately recover the intended phonological, lexical, or syntactic form. Examples of redundancy avoidance below the discourse level, as discussed in Section 2.1.1, include vowel short-

ening (Aylett & Turk, 2004), use of shorter word variants (Mahowald et al., 2013), or omission of optional complementizers (Jaeger, 2010).

At the discourse level, one encounters what may be termed informationally redundant utterances (Walker, 1993) – utterances which contribute nothing new to the discourse, and whose content is either already present in the discourse, or easily inferred from other content that is present. Utterances such as 1 are at face value redundant, in that they overtly state that 'John' *paid the cashier*, which conventionally can be inferred simply on the basis of him having gone *shopping*:

(1) "John went grocery shopping. **He paid the cashier!**"

Once it has been established that *John* went grocery shopping, comprehenders' expectations of a world state where the *paying* action has occurred are very high (cf. Zwaan et al., 1995; Bower et al., 1979). A theoretic account of utterance choice which places a constraint on informational redundancy would predict that uttering the second sentence in this string would be marked, at best, and that it might possibly cause some comprehension difficulty. Further, a *pragmatic* account may predict that comprehenders should note this markedness, and that it should have consequences for either their view of the speaker (as somewhat odd), or their interpretation of the discourse. However, the degree to which such an utterance is, in fact, *marked* (beyond being, presumably, infrequent) is under some debate.

While it has generally been uncontested that speakers tend to avoid unnecessary redundancy, what has been contested is the degree to which redundancy violates communicative norms, or constitutes non-rational speaker behavior (in the sense that it impedes the speaker achieving their communicative goals). In the realm of pragmatics, Grice (1975) pointed out that providing excessive information hardly constitutes irrational communicative behavior, and questioned whether the second Quantity Maxim (which he defined as guarding against excessive redundancy) in fact held. For further discussion of the Quantity Maxim, and the concept of *overinformativeness* in pragmatics, see Section 3.1. In the realm of psycholinguistic and computational theories of language production, while redundancy is typically acknowledged as suboptimal from the speaker's perspective, it is not seen in any respect as something that *impairs* communication (cf. Aylett & Turk, 2004; Cohen, 1978; Jaeger, 2010).

In general, there is ample evidence that speakers are routinely overinformative at the informational level, and that speaker overinformativity is frequently tolerated by listeners (Baker et al., 2008; Engelhardt et al., 2006; Nadig & Sedivy, 2002; Walker, 1993). In this chapter and the next, I explore the question of whether there is empirical evidence for a constraint against redundant speech as a *communicatively irrational* act – meaning, one which negatively affects communication between interlocutors. If redundancy in fact impairs communication, then this is something that should be clearly accounted for by theories of efficiency-based language production, such as UID. However, as I discuss in Section 2.1.3, the UID framework by itself is critically

unequipped to represent the nature or effects of communicative impairment beyond the question of signal distortion.

3.1 Informational Redundancy

First, I will discuss in further detail how informational redundancy is conceptualized, in the context of the experiments I introduce in the following chapter. As efficiency-based theories of language production have by and large focused on redundancy at the sub-discourse level, I will not cover this literature here. Instead, the reader is referred to the discussion in Section 2.1.1.

3.1.1 Background World Knowledge

Utterances such as the one in 1 are redundant on the basis of background world knowledge. As background knowledge can be fairly unsystematic and comprehender-specific, and can be difficult to control for, in the Chapter 4 experiments I use *script*, or *schema* knowledge as a proxy for world knowledge. *Script* knowledge refers to people's implicit awareness of the typical event structures of various stereotyped activities, such as *going shopping*, or *going to a restaurant* (Minsky, 1975; Fillmore, 2006; Schank & Abelson, 1977). The former, for example, normally involves events such as *going to a store*, *selecting food items*, and *paying the cashier*. Comprehenders anticipate upcoming events once a script is 'invoked' (Zwaan et al., 1995), and when recalling stories based on scripts, have difficulty remembering which actions were actually mentioned, and which were unmentioned but only implied by the script (Bower et al., 1979). These findings suggest that events which are strongly associated with a script are almost part of its conventional meaning, and that explicitly mentioning their occurrence is therefore redundant[1].

Utterance 1 introduces a well-known script or event sequence (*grocery shopping*), followed by an informationally redundant event description (*he paid the cashier!*), which references a highly predictable sub-event from the script. In this example, the event described in the second sentence is already strongly implied to have occurred by the preceding invocation of the *grocery shopping* script – given the assumption, shared by most speakers and comprehenders, that people overwhelmingly pay cashiers when they go grocery shopping. Mentioning it explicitly, therefore, is redundant.

[1]Highly inferable events are occasionally used as temporal anchors (*After she entered the restaurant, she...*), and may be used to transition back from interruptions to the script (*She stopped to talk to Brad on the street. She then entered the restaurant...*). However, outside of these contexts, easily inferable script events are usually not mentioned overtly (Bower et al., 1979; Regneri et al., 2010).

3.1.2 Informational Redundancy

While most pragmatic theories do address cases where a speaker may be informationally redundant (Grice, 1975; Horn, 1984; Levinson, 2000, among many others), as mentioned, they often leave open the question of whether comprehenders do, in fact perceive (unjustified) redundancy as infelicitous, as well as how they interpret redundant utterances. Most accounts do argue that comprehenders expect speakers to behave rationally – namely, by communicating in a manner that is consistent with getting the intended message across. However, as Grice (1975) notes, it's unclear whether excessive redundancy comes into any real conflict with the goal of successful (truthful, sufficiently informative, relevant, etc.) communication – even though he floats the possibility that comprehenders may wonder what the 'point' of excessive information is, and attempt to rationalize unexpected 'dips' in informational utility by infusing them with additional pragmatic meaning.

Informationally redundant utterances do not clearly interfere with comprehension, as *underinformativeness* or underspecification does and may in fact aid comprehension in some cases (e.g., object identification; cf. Nadig & Sedivy, 2002; Rubio-Fernández, 2016)[2]. In this light, it is not straightforwardly clear whether overinformativeness constitutes non-rational speaker behavior. It is, however, possible, that comprehenders perceive excessive information as, at minimum, non-relevant to the discourse (Grice, 1975; Horn, 1984). The question then, is whether comprehenders make any particular note of redundancy, simply find it odd or infelicitous, or attempt to accommodate it.

If comprehenders do perceive redundant information as irrelevant, then rational speakers should avoid overtly stating conceptually redundant information, except in those cases where this information is intended to communicate a more informative non-literal meaning (or signal an unusual world state). Correspondingly, comprehenders where possible ought to interpret conceptually redundant utterances as either an attempt to convey some non-literal (relevant and informative) meaning, or as reflecting a background world state where the information conveyed can't be taken for granted, and is therefore informative. How comprehenders do in fact react to redundancy has to date only been empirically investigated within the relatively narrow scope of nominal modification, where the evidence, discussed further in Section 3.2.1, has largely been equivocal.

3.2 Literature Review

The work in the following chapter builds on two primary strains of research: interpretation and perception of informational redundancy on the one hand, and relatively new work on inferences bout background world states (vs. speaker intentions) on the

[2]This is not to say that comprehension is not in any way impaired by redundancy, and in fact it appears likely that it is - but at face value, there is nothing about receiving more information than needed that necessarily hinders one from arriving at the intended meaning of a message.

other. I also look at how implicit prosody affects pragmatic interpretation, in order to determine whether and to what extent any effect of informational redundancy on utterance interpretation is generalizable. To date, research on prosody and pragmatic inferences has largely been limited to the semantic effects of contrastive prosody. As a consequence of investigating the interpretation of informationally redundant utterances, I also look at the interpretation of *particularized*, or *ad-hoc* pragmatic inferences, which arise only in specific contexts, and/or on the basis of reasoning about world knowledge. These in general have not received a lot of attention in pragmatic theory and experimental work in pragmatics, partially due to their idiosyncratic nature, which makes them difficult to study systematically; and partially due to being seen as less relevant to a theory of pragmatic vs. semantic meaning than, for example, scalar implicatures Levinson (2000).

3.2.1 The Problem of *Overinformativeness*

First, I discuss a problem of terminology. In most experimental work, informational redundancy has been described as a problem of *overinformativeness*, *overspecification*, or *overdescription*, and as addressed by the second part of Grice's Quantity Maxim, which states that speakers should provide no more information than is necessary to get their message across. However, *overinformativeness* in the pragmatic literature has rather confusingly been used to refer to both informational redundancy (Engelhardt et al., 2006; Grice, 1975), as well as to the relative informativeness of terms on an implicational scale (e.g., the use of *some* when *all* is sufficient) (Horn, 1984; Levinson, 2000). The latter variety of *overinformativeness*, now ore typically associated with the Quantity Maxim, is more a problem of unjustified vagueness where a more *precise* description is available. Informational redundancy, in contrast, is a problem of *excessive* wordiness or precision, as in the case of overinformative nominal modification (such as using *the big red cup* or *the cup on the towel* to identify the only available cup in a given context), where speakers might choose to describe objects in more detail than is strictly necessary. In this thesis, I concern myself strictly with overinformativeness in the sense of informational redundancy, as originally described by Grice (1975), and in the literature on nominal overspecification.

While informationally redundant utterances are typically regarded as infelicitous in the linguistics literature, they have been observed to be surprisingly common in natural dialog. Baker et al. (2008) observed that such utterances are frequently used in response to signs of listener non-comprehension, when responding to listener questions, or when speaking to strangers. Walker (1993) also concludes that informationally redundant utterances are specifically used to address cognitive resource limitations (e.g., memory for preceding discourse, limited inference-making capacity), as well as to serve a narrative function. In the latter case, this may for example involve drawing attention to a particularly salient or relevant fact. In other words, many or most informationally redundant utterances are not in fact redundant, as the

apparent redundancy has communicative purpose. Literature on nominal overspecification has similarly found that speakers are extremely likely to attach 'redundant' color descriptions to nouns, even when doing so provides no new information regarding which object is being referred to. However, in this case as well, there is evidence that most *overinformative* nominal modification is not in fact *overinformative*, as 'overdescribing' an object can facilitate more rapid and efficient object identification. Here I will review some of the experimental work on informational redundancy, with a focus on interpretation of nominal overspecification.

Most experimental work on production and comprehension of informationally redundant utterances has focused on nominal modification in referent identification tasks, which typically instruct participants to look at or somehow engage with items such as: *the [red] apple*, *the [tall] boot* (Engelhardt et al., 2006; Nadig & Sedivy, 2002; Sedivy, 2003; Davies & Katsos, 2010, 2013; Pogue et al., 2016). The aim of these studies has been to determine some combination of the following: 1) whether overinformative descriptions are perceived as infelicitous by comprehenders (i.e., whether overinformativeness apparently violates some communicative norm); 2) whether overinformativeness helps, hinders, or has no effect on object identification; 3) whether comprehenders attempt to accommodate overinformative descriptions by making inferences which increase the informational utility of the descriptions; and 4) whether comprehenders alter their beliefs about the rationality of the speaker (or the baseline informativeness of their speech) following use of overinformative descriptions.

What has been found is that in interactive, spontaneous speech, speakers frequently modify nouns with adjectives that are not strictly necessary for referent identification (e.g., referring to a cup as *the red cup*, in a context where there are no other cups of any color) (Engelhardt et al., 2006; Nadig & Sedivy, 2002, 30% and 50% of nominal descriptions were overspecified in spontaneous speech, respectively). Further, comprehenders frequently do not find such utterances infelicitous: Engelhardt et al. (2006) showed that comprehenders judge overinformative descriptions as significantly more acceptable than underinformative descriptions, and that overinformative descriptions do not trigger additional (e.g., contrastive) inferences. Davies & Katsos (2010) find that overinformative expressions are more likely to be produced, and less likely to be judged infelicitous, than underinformative expressions, although they are still judged to be suboptimal[3]. Sedivy (2003) showed that when comprehenders hear an object described with a clearly overinformative and prototypical color adjective (e.g., "yellow banana"), they make contrastive inferences (e.g., rapidly infer that a non-yellow banana must also be present).

What seems to emerge is that overinformative descriptions are easily tolerated when they describe perceptually useful or non-canonical properties, which may speed

[3]However, Davies and Katsos purposefully use adjectives less likely to be produced spontaneously - color adjectives, by far the most likely to be used redundantly, are avoided, and the adjectives that they use are largely either inherently contrastive (e.g., 'tall,' 'big'); or describe a default, assumed state (e.g., 'unbroken egg,' 'fresh apple').

up object identification; and are more likely to be judged suboptimal, and/or trigger pragmatic inferences, when they don't. Indeed, Rubio-Fernández (2016) showed that experimentally increasing the perceptual usefulness of color adjectives causes them to be produced more frequently, as well as that color adjectives are more likely to be used for atypical than typical colors. In a related line of research, Pogue et al. (2016) found that after being exposed to a speaker repeatedly using overinformative (color or scalar) object descriptions, comprehenders are less likely to make generalizations about the speaker's rationality or informativity than when they use underinformative descriptions. This suggests that comprehenders are relatively insensitive to deviations from "optimal" informativity that do not interfere with basic utterance comprehension, or else perceive them as relatively commonplace and inconsequential.

Overall, this work has shown that some types of informational redundancy may be helpful to the comprehender, and that informational redundancy in general is tolerated by comprehenders. There is, however, also evidence that informationally redundant utterances which have no apparent (e.g., perceptual) utility are unlikely to be produced, are generally judged to be relatively infelicitous, and tend to generate inferences. More generally, there is still some difficulty in distinguishing what constitutes informational redundancy, which creates difficulty in determining the precise theoretical implications of previous work (e.g., perceptually helpful 'redundant' adjectives are questionably redundant in the first place, in the sense of having communicative utility).

Additionally, these studies are limited by the fact that they uniformly focus on a very particular, and relatively concise variety of informational redundancy, which is further bound to a specific class of lexical items, raising the question of to what degree it's possible to generalize from the results. What this points towards is a need to look at informational redundancy in the context of utterances and constructions that are both quite costly for speakers, and have no readily apparent utility to comprehenders - either in terms of perception or comprehension, or in terms of facilitating the completion of a task. Further, I would argue that it's important to investigate constructions that are less bound to a specific set of lexical items, and are more likely to be perceived as flouting of a conversational norm against redundancy - for example, complex and lengthy multi-word utterances such as those in Example 1.

3.2.2 Common Ground Beliefs

To date there has been relatively little work on the different strategies comprehenders might employ in making sense of an apparent violation of conversational norms. Most work has focused on the scenario where a comprehender detects an apparent maxim violation, assumes that the speaker is in fact being cooperative, and comes up with an additional, non-literal meaning that the speaker may have intended (which repairs the apparent violation). Another strategy is simply to assume that the speaker is being plainly uncooperative, or that there is an intended meaning but that the com-

prehender is not privy to it, if no plausible intended meaning can be computed. A third strategy, which has received little attention, is that of modifying background assumptions about the world in which events take place, if doing so would repair the apparent violation.

The lack of attention to this strategy is likely partially due to a focus on implicatures, or specifically intended meanings, in pragmatic theory. To my knowledge, the only work to look at this in depth is Degen et al. (2015), which investigated comprehenders' willingness to revise their assumptions about the assumed common ground, in response to utterances whose pragmatic meaning would otherwise be inconsistent with it. They found that background assumptions about the world are surprisingly defeasible: comprehenders frequently accommodate the pragmatic meaning of utterances such as '***some** of the marbles sank*' (upon being thrown into a pool), by assuming that the utterances signify a strange scenario where physics doesn't quite work as expected. Further, a pre-utterance belief that a scenario is strange significantly increases the strength of the '*some, but not all*' implicature that is then drawn by the comprehender.

In the case of informationally redundant utterances, if, as in Example 1, a speaker states that *John*, having gone shopping, *paid the cashier*, a comprehender might 'repair' the redundancy by assuming that *John* does not in fact habitually pay the cashier. While this may occur parallel to an assumption that the speaker *intended* to use this utterance to communicate that *John* is not a cashier-paying individual, the strategy of modifying background assumptions can well proceed without any assumptions about speaker intent. Perhaps the comprehender is a third party not privy to the background knowledge of the speaker and intended listener, or perhaps the speaker isn't aware that the listener isn't familiar with *John*'s usual paying habits. In fact, in the case of our example, it seems relatively unlikely that a speaker would choose to communicate information about *John's* paying habits in this particular manner, making this an inference, but not an implicature.

While most theoretical interest lies in implicatures, it's important to be able to model pre-utterance and changing post-utterance assumptions about the common ground, given that they have been demonstrated to have a marked effect on which inferences are drawn by comprehenders, as well as their strength (Degen et al., 2015; see also the literature on presupposition, e.g.: Stalnaker, 1973). Further, an exclusive focus on intended meanings, rather than changes in background assumptions, particularly in empirical work, may lead to erroneous conclusions that comprehenders are drawing no pragmatic inferences from a given utterance, even when this is not the case. In the studies described in the following chapter, I introduce a novel method for testing the shifting of background assumptions, collect data that can be used in the future to test formal models of pragmatic reasoning, and explore the willingness of comprehenders to shift background assumptions in different contexts.

3.2.3 Effect of Implicit Prosody on Pragmatic Interpretation

One of the questions I ask, relevant both to accurately detecting and modeling an effect of informational redundancy, is to what degree increased or decreased emphasis on the utterance (without changing the semantic content, or truth-conditional value) influences the interpretation of informational redundancy. Intuitively, an utterance with some amount of prosodic emphasis, such as "*John went grocery shopping.* ***He paid the cashier!***" is more consistent with what appears to be the most likely pragmatic inference: *John is not a habitual cashier-payer*, which is a fairly surprising bit of information to a naive comprehender. This raises the question of whether prosodic emphasis is necessary to obtain this interpretation. To look ahead, in Chapter 4, I show that it is not, but that some degree of prosodic emphasis, or other attention-drawing discourse marker, strengthens the interpretation. In Chapter 8, I present a formal model of *why* it is helpful in obtaining this interpretation.

Although it is generally accepted that prosodic emphasis may influence utterance interpretation, there is very little empirical evidence that prosodic changes which contribute little by way of conventional meaning have a substantial effect on the generation of pragmatic inferences[4]. One can however imagine that a redundant statement made loudly and confidently may lead a comprehender to believe that the speaker is very intentionally communicating that particular bit of information to them (cf. Wilson & Sperber, 2004), and that it should be taken seriously (signifying either that the speaker is being blatantly uncooperative by violating a communicative norm for no reason, or that there is an additional reason that the information was so purposefully transmitted). On the other hand, if a speaker vaguely mumbles an informationally redundant utterance under their breath, the comprehender might simply conclude that the speaker is reminding themselves of something, is unsure about what they really want to say, is mentally rehearsing a course of events, having some production difficulty, etc.. After all, if the speaker were truly concerned with the listener obtaining a rather unusual interpretation, perhaps they would ensure this by purposefully drawing the listener's attention to the utterance. To look ahead, in Chapter 8, I argue that these are exactly the considerations the comprehender makes, and present a formal model of the reasoning process by which a comprehender obtains a stronger inference from a more attentionally prominent utterance.

Along the same lines, Bergen & Goodman (2015) hypothesize, on the basis of formal probabilistic models of pragmatic reasoning, that rather than focal/contrastive stress carrying conventional semantic meaning, the contrastive or exhaustive interpretation ("***BOB*** *went to the movies*" -> **only** *Bob* went to the movies) arises due to the comprehender perceiving the speaker as having made extra effort to communicate exactly that particular bit of information to them. They argue that an

[4]An exception is the effect of contrastive prosody (e.g., Kurumada et al., 2012), which is generally thought to be semantic – however, it has also been suggested that the effect of contrastive prosody is a pragmatic inference, as discussed in the following paragraph.

utterance which is increased in volume or duration is more likely to be attended to or accurately perceived by the comprehender and that, correspondingly, speakers can intentionally exploit this to signal to comprehenders that this particular utterance is important, and specifically meant to not be confused with any alternative utterances. On the basis of this and similar work, I therefore experiment, in the following chapter, with having participants interpret an informationally redundant utterance both with implicit exclamatory prosody (ending with an exclamation mark), as well as without implicit prosody (ending simply with a full stop). To determine whether the critical distinction is one of the presence or absence of prosodic emphasis, I additionally present participants with an informationally redundant utterance preceded by an attention-drawing discourse marker, but without prosodic emphasis.

3.2.4 Context-dependent Implicatures

To date, most formal or experimental research on pragmatic inferences has focused on the production and interpretation of scalar implicatures (Horn, 1984, Levinson (2000)), such as the use of *some* to implicate *not all*, or *warm* to implicate *not hot*. Non-generalized *ad-hoc* inferences, which arise only in specific contexts, have not received much attention from pragmaticists, experimentally or otherwise. Traditionally, scalar implicatures have been regarded as a separate class of *conventionalized* inferences which rely minimally on context or general reasoning about speaker intentions (Levinson, 2000), and which arise from the use of specific lexical items (or classes of lexical items). In recent years this view has increasingly been challenged (Degen & Tanenhaus, 2016; Grodner et al., 2010), with evidence indicating that the distinction between *conventionalized* (*generalized*) inferences, and *particularized* (*ad-hoc*) inferences is in any case not categorical, although the nature of differences between the two classes remains difficult to determine, and the latter case has traditionally been understudied.

Research on conventionalized inferences has been critical to developing formal linguistic theory, due to the role they play in disambiguating pragmatic and semantic contributions to utterance meaning. However, context-dependent (*ad-hoc*) inferences, which occur far more frequently and ubiquitously, are similarly important to developing a more general theory of human communication. The body of experimental work teasing apart which properties of utterances trigger, alter, or modulate the strength of pragmatic inferences is still relatively small – however, having a more comprehensive model of cues which are taken into account by comprehenders, when interpreting utterances, is necessary both for building models of pragmatic reasoning, and for interpreting empirical results. In addition, there is a general need for further quantitative data on the specific conditions under which inferences are generated, in order to develop and test predictions of formal models of pragmatic reasoning, such as those discussed in Chapter 7.

3.3 How Might Speakers React to Informational Redundancy?

There are multiple ways in which comprehenders may react to informational redundancy. If redundancy causes comprehenders to reevaluate the meaning of the utterance, or their beliefs about the world, then this indicates that there is an upper limit on how redundant speakers may be without distorting the message that they intend to transmit. In this section, I briefly speculate how comprehenders may react to specific instances of informational redundancy, or *overinformativeness*. I distinguish 4 theoretical possibilities. Specifically, I consider what might happen when a comprehender encounters one of the following utterances (which are part of one of the experimental stimuli presented in the following chapter):

(2) John just came back from the grocery store. **He paid the cashier.**

(3) John just came back from the grocery store. **He paid the cashier!**

3.3.1 Hypothesis 1a: IRUs are not perceived as marked

The first possibility is that comprehenders do not find informational redundancy particularly marked, as it does not necessarily interfere with interpreting the intended message – or, at most, find redundant utterances slightly odd or suboptimal, as has been found in some studies (Davies & Katsos, 2010). It's both likely that comprehenders do not expect speakers to exhibit entirely rational communicative behavior at all times, and that conversational norms, if they have little or no effect on the success of the communicative act, are not consistently adhered to. In this case, comprehenders should interpret the informationally redundant utterance literally, and should not make any particular inferences about the speaker's rationality. This would be evidence that excessive redundancy does not impair communication, and that the only limit on how redundant a speaker may be is their own desire to conserve effort.

3.3.2 Hypothesis 1b: Markedness may be noted, but no pragmatic inference generated

The second, related, possibility is that comprehenders may note redundancy in speech, and find it marked, but not draw any inferences regarding the speaker's intended meaning, or the background world state implied. They might instead ascribe the redundancy to speaker error: perhaps the speaker is stalling for something else to say, having production difficulty, or is simply not communicating very well in that particular moment. Alternately, they may determine that the speaker is perhaps simply predisposed to making informationally redundant statements, and that future utterances should be interpreted in this light. In the case of the utterance 1, in this

scenario, one would again expect that comprehenders would interpret the utterance literally, and make no more of it than stated (i.e., they would simply take away the message that on some particular instance, *John* paid the cashier). However, they may take away the 'message' that the speaker is unusually prone to *overinformativeness*, and that they are relatively *irrational*, from the point of view of efficiency.

The experiments introduced in Chapter 4 are not designed to empirically distinguish between Hypotheses 1a and 1b – as I am primarily concerned with how redundancy may impair accurate message transmission in the moment, the question of how comprehenders perceive speaker competence is left to future work[5]. I will therefore subsequently refer to these two hypotheses jointly as Hypothesis 1. Hypothesis 1, then, predicts that comprehenders do not ascribe any special *meaning* to overinformativeness.

3.3.3 Hypothesis 2: Non-detachability from semantic content

If comprehenders do expect speaker utterances to always have a certain level of informational utility, then they may attempt to resolve the provision of excessive or unnecessary information by drawing pragmatic inferences, regarding what they believe the utterance may mean or signify from the speaker's perspective. These pragmatic inferences would increase the informational utility of the utterance, and allow comprehenders to maintain the belief that the speaker is being cooperative – since assigning an 'informative' pragmatic meaning to an apparently redundant utterance in effect cancels out the redundancy. In the case of utterance 1, comprehenders might conclude that *John's* cashier-paying is being announced due to its being unusual or unexpected, and that *John* can't therefore typically be counted on to pay the *cashier*.

This interpretation should take place as long as the background and linguistic context are basically consistent with such an interpretation. Critically, this interpretation should generalize across most contexts, and as in the case of most pragmatic inferences, it should be unaffected by changes to the utterance which do not alter its semantic content (generally referred to as *non-detachability*; Grice, 1975), such as prosodic and/or discourse markers which do not change the truth-conditional meaning of the utterance. In other words, the inference should be attached to the semantic content of the words, in their context, and not the specific form of the utterance – it should not be tied to any specific mode of expression.

3.3.4 Hypothesis 3: Sensitivity to form of expression

The third possibility is that, as in Hypothesis 2 (Non-detachability), comprehenders react to a statement of *John's* having paid the cashier by inferring (for example) that *John* must be a habitual non-payer. However, as the inferences I am concerned with

[5]However, it is worth considering that if the listener begins to perceive the speaker as less *rational*, then this may alter their interpretation of future utterances.

here are *ad hoc* inferences heavily reliant on contextual support, and are based on a comprehender's reasoning about the possible causes for a speaker's redundancy, it is likely that any inferences generated may be relatively sensitive to how exactly the utterance is expressed. In particular, it is likely that expending extra articulatory effort on expressing an already redundant utterance would increase the strength of any pragmatic inferences drawn, or even cause inferences to be drawn where none would be otherwise. Fundamentally, a greater show of *intentionality*, in apparently violating communicative norms, provides more evidence that this norm-violation is not simply due to a speech error, or difficulty in utterance planning.

This echoes Wilson & Sperber (2004)'s stated basis for their Communicative Principle of Relevance: "by producing an ostensive stimulus, the communicator therefore encourages her audience to presume that it is relevant enough to be worth processing" (an *ostensive stimulus* being one that is "designed to attract an audience's attention and focus it on the communicator's meaning"). In the case of utterances 2 and 3, what would be predicted in this case is that the more obvious effort is expended on producing the utterance (whether in the form of prosodic emphasis, or another attention-drawing signal of relevance and intentionality), the stronger the inference. To note, some about of sensitivity to the form of expression is not necessarily incompatible with the second hypothesis (that of *non-detachability* from the semantic content), but the complete absence of an inference would be.

Evidence in favor of either Hypothesis 2 or Hypothesis 3 would indicate that redundancy influences message interpretation. However, evidence that comprehenders are sensitive to how this redundancy is *framed* would indicate that either the effect of redundancy on interpretation is not robust (if it is dependent on a very particular context), or that listeners engage in fairly sophisticated pragmatic reasoning in determining what, exactly, unexpected redundancy is evidence *of*. In the case of Hypothesis 3, evidence that increased articulatory effort, specifically, influences interpretation of redundant utterances would be a particularly strong indicator that speakers are limited in how much effort they may expend on producing an utterance before it begins to distort the intended message.

3.4 Experimental Setup

In the following chapter, I test the hypothesis that excessive redundancy distorts the message received by the speaker. To this end, I present three experiments which test whether informationally redundant event descriptions substantively alter a comprehender's interpretation of the discourse common ground. I expect that redundant event descriptions will lead comprehenders to alter their initial beliefs about how predictable, or habitual, the event in question is, on the premise that less habitual events are more likely to be mentioned. Specifically, I predict, consistent with the second and third scenario outlined above, that informationally redundant event descriptions

should generate *habituality inferences* – wherein comprehenders resolve the apparent dip in informational utility by concluding that the usually predictable and habitual event described is, in fact, *non-habitual*, as this would justify its mention. If comprehenders accommodate informational redundancy by altering their beliefs about the common ground, then this indicates that excessive redundancy in speech distorts the speaker's intended message, which is not accounted for by the UID hypothesis, although not inconsistent with it. Further, it would suggest that UID as a theory of language production is somewhat underspecified, as it has no mechanism to represent this manner of message distortion (which occurs about the level of recovering the intended signal, or utterance semantics), and no theoretical interface with the pragmatic interpretation of an utterance.

Chapter 4

Informationally Redundant Utterances: Results

In this chapter, I present three experiments which test whether redundancy influences message interpretation. The basic utterance I consider is the following:

(1) John went shopping. ***He paid the cashier.***

In the given scenario involving *shopping* and *paying the cashier*, a likely inference is that *John* does not habitually pay (e.g., he has someone else pay for him, is a habitual shoplifter, or gets free groceries). I first look at these utterances in discourse contexts which implicitly support a *habituality* inference – ones where additional prosodic emphasis is put on the utterance, or where the utterance is framed by an attention-drawing discourse marker. As noted above, the more ostensive the stimulus, the more likely the comprehender should be to attempt to interpret its meaning. Further, if the speaker is attempting to communicate something *unusual* (*John* doesn't usually pay, and yet *this time* he paid), then it intuitively seems more likely that they will attempt to draw the listener's attention to this face. I therefore consider the presence of some about of prosodic or discourse emphasis to be the 'default' case.

The first experiment uses implicit exclamatory prosody (the marker '***!***') at the end to signal that the utterance is an intentionally conveyed, important, and relevant piece of information. The second experiment uses the discourse marker '***oh yeah, and. . .***' to do the same, while avoiding the surprise conventionally implied by the exclamation mark. In the third experiment, there are no specific markers of relevance or 'attention-worthiness,' as in 1 above. In this case, I predict that informational redundancy by itself, in the presence of prosodic or discourse cues as to relevance and intentionality, triggers weaker *habituality* inferences, consistent with the hypothesis of *form sensitivity* (see Section 3.3.4). In general, evidence that increased speaker effort – whether in terms of making the obvious explicit, or in terms of putting increased emphasis on an utterance – alters a comprehender's interpretation of the intended

message (or of the common ground) would indicate that there is an upper limit on how redundant speakers may be without impairing communication.

4.1 Experimental procedure

The three experiments presented in this chapter are conceptual replications of 3 previously run studies, an account of which can be found in the Appendix C. The studies described here have an increased sample size, and were conducted concurrently on the same general population, to ensure that their results could be compared directly. The stimuli were also redesigned to read more naturally, and filler stimuli were included to ensure replicability[1].

The following experiments were run using the same interface, and on the same population of Amazon Mechanical Turk workers, in small rotating batches (of 9, or less): a batch of 9 participants completed the first experiment, after which the second experiment was scripted to go live until it was completed by 9 participants, and so forth. The only difference between the 3 experiments was the manipulation of prosody or discourse markers. Running them concurrently and on the same population therefore makes it possible to directly compare their results. All workers who participated in an experiment were automatically disqualified from participating in any future batches; i.e., no participant took part in more than one experiment or batch.

The number of eligible participants (n=2100) was predetermined through a simulation power analysis (adapted from Arnold et al., 2011): all predicted higher-order interactions, assuming effect sizes determined by the results of the experiments I am replicating, were detectable at $> .80$. The `R` code and a plot for the power analysis can be found in Appendix D.

4.2 Experiment 1: Implicit intent signaled by prosody

I first test whether informationally redundant event descriptions trigger *habituality* inferences when the utterance is apparently effortful, intentional, and attentionally prominent – here signaled by an exclamation mark at the end of the utterance (which would be inconsistent with the "no inference" hypothesis). Intuitively, exclamatory

[1]To note, in the original studies, participants saw each condition no more than once. It appears likely that in crowdsourced studies, when participants can be exposed to a very small number of stimuli without repeating any experimental conditions, fillers may at times be unnecessary, provided there is sufficient difference between experimental conditions – and in fact, possibly detrimental to performance, if an increased number of items, and/or longer experiment duration, causes participants to read less closely for meaning.

intonation is a natural way of introducing information that may be noteworthy or unusual (Rett, 2011), without otherwise altering the semantic content of the utterance. When utterances are context-dependent and even if they are not; see Degen & Tanenhaus (2015)], speakers tend to provide multiple signals of their intended meaning, in order to make inferences easier for the comprehender to compute. One would expect for this to particularly be the case when the meaning of the utterance substantially violates expectations or previously held beliefs, as opposed to simply providing new but marginally expected (or at least unsurprising) information.

I present naive participants with a limited number of brief 'narratives,' which set up the common ground context, relationships between discourse participants, and some typical or atypical properties of their usual behavior (where relevant). Some of the narratives include brief dialogue between two discourse participants at the end (which may include informationally redundant or non-redundant event descriptions). After reading the narratives, participants rate how *habitual* they believe certain behaviors in the story to be. I expect that participants who read informationally redundant event descriptions will infer that the utterance in fact signals that the event is relatively unexpected, or non-habitual (on the assumption that only relatively unexpected events warrant explicit mention). In contrast, those participants who read non-redundant event descriptions should draw no such inferences.

4.2.1 Methods

Participants

700 eligible participants (760 total; median age bracket 26-35; 50% female), were recruited on Amazon Mechanical Turk. The task was open only to workers located in the US, and with an approval rating of $\geq$ 95%. All workers were asked to state their native childhood language (with no penalty for stating a language other than English, to encourage accurate reporting), age bracket (under 18, 18-25, 26-25, and up, in intervals of 10), and gender. Those who did not indicate English, or listed their age as outside the interval of 18-65, were excluded from all analysis (60; 7.89%), with additional participants recruited to replace them.

Those who did not provide accurate or plausible responses to the trial questions, all of which had a range of 'valid' and 'invalid' responses, were unable to proceed to the main task, and their data as a result was not recorded by the platform (e.g., those who rated the likelihood of 50% heads on multiple fair coin flips as low, compared to other possible outcomes). Participants were likewise unable to proceed in the study, or submit their results, without having answered all questions.

Design

The primary question of interest is whether informationally redundant utterances (in this case, descriptions of highly *habitual* activities) trigger pragmatic inferences.

These inferences should lead to the revision of common-ground beliefs about the *habituality* of said activities (and so 'repair' the violation, or dip in informational utility):

(2) "John just came back from the grocery store. **He paid the cashier!**"

The bolded utterance here, given a 'default' or *ordinary* common ground, is ***informationally redundant***. The hypothesis is that readers will infer that *John* does *not* habitually pay the cashier, as such a scenario would justify overt mention of *John's* cashier-paying. The informational redundancy arises due to the high *conceptual* (or *event*) *predictability* of *paying the cashier*, and is resolved if one assumes that this activity is not as habitual, or predictable as initially assumed.

A further goal was to see whether the inference (that an activity is less habitual than would otherwise be expected) could be canceled by manipulating the common ground.

Common ground manipulation The activity described becomes 'non-habitual' given a *wonky* common ground[2] such as in 3, where the context suggests that typical assumptions (e.g., that some given individual would *pay the cashier* when they *go to the grocery store*) may not hold. At that point, the activity description ceases to be informationally redundant, and the inference should therefore not arise. This control condition keeps the description itself constant and manipulates only the common ground. It thus ensures that any effect measured is in fact due to the presence of informational redundancy, and verifies that comprehenders are sensitive to discourse context.

(3) Common Ground Context: John habitually doesn't pay.
"John just came back from the grocery store. **He paid the cashier!**"

Finally, I wanted to provide a baseline for 'typical' interpretation of non-redundant event descriptions; and to confirm that similarly structured descriptions of conventionally *non-habitual* activities, as in 4, do not provoke similar inferences (which would suggest a problem with the stimulus design or response measure). In 4, the utterance is not informationally redundant, and is not expected to generate any specific inferences. I also wanted to confirm that the *wonky* common ground in the previous example does not significantly affect the interpretation of conventionally *non-habitual* event mentions (which would suggest that there is an unexpected effect of context manipulation on stimulus interpretation, in general):

[2]I borrow the term *wonky* from Degen & Tanenhaus (2015), where it is similarly used to describe non-default common grounds, in which typical rules as to how things proceed are expected to not hold, and which comprehenders may assume when encountering otherwise pragmatically infelicitous utterances.

(4) CONTEXT: *Ordinary* **or** John habitually doesn't pay.
"John just came back from the grocery store. **He got some apples!**"

As in 3, participants should draw no habituality inferences here, as the event described is not (typically) overly habitual. These conditions therefore provide a secondary control measure.

Materials

24 stimuli were constructed as brief stories/narratives, based on distinct stereotyped scripts or events. Each story had one of 2 context types (*ordinary* vs. *wonky* common ground, relative to the *conventionally habitual* script activity). In all stories, declarative utterances, spoken by one of the discourse participants, described one of 2 types of script activities (*conventionally habitual* vs. *non-habitual*), making a total of 4 conditions[3].

Conventionally habitual activities 5 can normally be inferred simply from the 'speaker' having invoked the script, while *non-habitual* activities 6 can not be inferred automatically, as they may only occasionally occur as part of the script activity. To clarify, I am using the term *conventionally habitual* to specify that the event almost invariably occurs as part of the event script (under normal conditions, and for typical individuals). Initial common ground was either *ordinary* ([1a] below) with respect to the script, or *wonky*, in that it implied the *conventionally habitual* event was in fact unusual for the event participant ([1b] below):

(5) CONVENTIONALLY HABITUAL EVENT

[1a] John **often goes to the grocery store around the corner from his apartment$_{\textbf{ordinary}}$**	[1b] John **is typically broke, and doesn't usually pay when he goes to the grocery store$_{\textbf{wonky}}$**
[2] Recently, he came home from the store with groceries. When he came in, he saw his roommate Susan in the hallway, and started talking to her about his trip to the store. As he went to the kitchen to put his groceries away, Susan went to the living room, where their roommate Peter was watching TV.	
[3] Susan said to Peter: "John just came back from the grocery store. [4] He **paid the cashier$_{\textbf{habitual}}$!**"	

The context/common ground manipulation in [1b] was used in order to render the *conventionally habitual* event unusual, or at least not habitual. Conventionally *non-habitual* activities could not be automatically inferred from the script having been invoked:

[3]The complete list of stimuli can be found in Appendix A

(6) Non-habitual Event

[1a] John **often goes to the grocery store around the corner from his apartment**$_{\textbf{ordinary}}$	[1b] John **is typically broke, and doesn't usually pay when he goes to the grocery store**$_{\textbf{wonky}}$
[2] Recently, he came home from the store with groceries. When he came in, he saw his roommate Susan in the hallway, and started talking to her about his trip to the store. As he went to the kitchen to put his groceries away, Susan went to the living room, where their roommate Peter was watching TV.	
[3] Susan said to Peter: "John just came back from the grocery store. [4] He **got some apples**$_{\textbf{non-habitual}}$!"	

Participants saw either only the common ground *context* [1] and *discourse setup* [2] (without numbering or special formatting), which made it possible to collect estimates of how habitual activities are believed to be, based on the context alone (*pre-utterance beliefs*); or the entire text, made it possible to collect estimates of how habitual activities are believed to be, based on both the context *and* the event description [4] (*post-utterance beliefs*).

Following each passage, participants were queried as to how habitual they believed the *conventionally habitual* and *non-habitual* activities (as well as 2 other scenario-relevant distractor activities) were, for the person who was the subject of the discourse (the individual mentioned in the context [1] and event description [4]):

1. How often do you think John usually *pays the cashier*, when going shopping?
2. How often do you think John usually *gets apples*, when going shopping?
3. How often do you think John usually goes to the grocery store?
4. How often do you think Susan and Peter usually talk to each other?

Each question could be responded to on a continuous sliding scale of 'Never' to 'Always' (see Fig. 4.1). The slider itself was not visible until the participant clicked on the point on the scale that they thought was most appropriate, to avoid having people default towards a particular value. After they responded to all questions, participants could submit their answers. Once they did, the next passage was displayed on a new screen.

12 of the stimuli included 3 discourse participants – one of whom engaged in the script activity (*John*), the second who learned from that participant that they engaged in it (*Susan*), and the third to whom the second communicated this fact (*Peter*). The other 12 only included two – the subject of the discourse, who engaged in the activity (*John*), and the second participant to whom they communicated this fact (*Susan*). Compared to the example above, for instance, *John* might instead be communicating directly to *Susan*: "*I just got back from the grocery store. I paid the cashier!*".

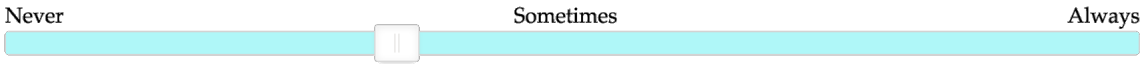

Figure 4.1: This is a slider, as used by experiment participants.

The construction of these stimuli was constrained in several ways. The scripts (e.g., *going shopping*) needed to be sufficiently complex to include multiple subactivities or subroutines, and there needed to be habitual as well as non-habitual subactivities (*paying the cashier*, *getting apples*). It needed to be possible for the script to play out without the habitual activity having taken place – otherwise, the discourse would be incoherent, or the inference would not be drawn. For example, one arguably cannot play *tennis* at all, without using a *racket*. There was also established common ground between all discourse participants, so that all were plausibly (from the point of view of the reader) aware of the typical habits of the discourse subject, particularly with regard to the activity described. Finally, the activities needed to be sufficiently stereotyped and (relatively) culturally invariant, so that participants could be expected to agree on what a script entailed, which activities were or weren't obligatory to the script sequence, etc..

All stimuli were normed on 3 qualities (in separate tasks): whether the activity fell into the *habitual* or *non-habitual* activity bin; whether the common ground manipulation was effective; and whether participants found it plausible that the script could be engaged in without the *habitual* activity. For activity predictability norming, participants were asked to rate the habituality of the activity (on a 0-100 scale), with an arbitrary cutoff of 70 between activity types. *Non-habitual* activities were on average rated 48.0 (25.1-68.1), and *habitual* activities were rated 87.8 (78.1-95.2). For common ground norming, participants rated *habitual* activities in *ordinary* (mean 83.4 [72.2-96.9]) or *wonky* common grounds (mean 39.2 [20.7-62.0]), with a within-item difference between the two of at least 15 points (mean difference 44.2; [19.8-72.9]); *non-habitual* activities had to score below 70 regardless of common ground (mean 45.2; on average 10.7 points higher in the *ordinary* common ground). For plausibility norming, a statement in the form of '*John* ***went shopping****, but didn't pay the cashier*' was rated as either *coherent* (plausible) or *incoherent* (implausible), with criteria being a majority of participants finding the statement coherent (*habitual*: 91% [67%-100%]; *non-habitual*: 94% [80%-100%]).

Measures

To measure comprehender beliefs regarding activity habituality, each story that was presented was followed by 4 questions presented in random order, regarding activities mentioned in the story (including both conventionally *habitual* and *non-habitual* activities associated with the stimulus item). The questions were accompanied by sliding scales which ranged from *Never* to *Always*, where participants could select any point along the scale, as seen in Fig. 4.1.

Prior to seeing any experimental items, participants were given several practice questions, unrelated to the experimental stimuli, which also used continuous sliding scales ranging from *Never* to *Always* (or similar). Unlike the experimental stimuli, these questions had 'correct' answers – such as *How likely is a fair coin to come up heads twice, if flipped 10 times? (very unlikely–very likely)*. If participants provided responses that could not be judged reasonably accurate, they were asked to re-read the instructions, and respond again, before they were able to proceed. This ensured that they were able to follow instructions, and were less likely to guess randomly throughout the experiment. There were no 'accurate' answers in the experiment itself. All points on the response scale were associated with a number ranging from 0 (*Never*) to 100 (*Always*).

Pre-utterance beliefs, or baseline beliefs regarding activity habituality, were estimated from responses to stimuli presented without the activity description (see the next section for a more detailed explanation). The responses, aside from setting baseline measures (*pre-utterance beliefs*) of activity habituality, also provide an additional norming measure for how likely it is that a particular activity would be engaged in, in the context of a given script. Thus, activities which are more or less habitual, within a given class, can be compared against one another.

Post-utterance beliefs regarding activity habituality were estimated from responses to stimuli which included the redundant or non-redundant utterance (activity description), or where the activity description/utterance was visible.

Belief change due to reading the activity description was determined by modeling the magnitude and direction of difference between *pre-utterance beliefs* and *post-utterance beliefs*.

Procedure

Participants were asked to read 6 experimental stimuli randomly selected out of the total of 24, as well as 4 filler items[4]. Each condition was only presented once, as follows. 2 of the stories were presented without the dialogue and event description (context and setting up of common ground only), and 4 stories were presented in their entirety (context, setting up of common ground, and the dialogue/event description). The 2 partial stories made it possible to collect measures of *pre-utterance beliefs* regarding activity habituality, and the 4 full stories provided measures of *post-utterance beliefs* conditioned on the event description.

SUBJECT 1: *pre-utterance* belief	SUBJECT 2 *post-utterance* belief
<context>	<context>
<setting up of common ground>	<setting up of common ground>

[4]To note, this means that each participant saw each manipulation only once, and the number of fillers was equal to the number of stimuli presented with dialogue.

	<dialogue>
#. <**habituality question**>	#. <**habituality question**>

The experiment thus employed a between-subject design for belief measures, where *pre-utterance* and *post-utterance* belief estimates for any given item were provided by different participants, to eliminate the possibility of participants conditioning their *post-utterance* estimates not only on inferences made from the text, but also on their own *pre-utterance* estimates[5]. The 4 filler stimuli had the same structure as above, but with the dialogue portion replaced by script-neutral utterances: *"You know, I'm really tired."*, "*Hey, do you know what time it is?*", "*So, what are you up to?*", or "*Have you heard the news today yet?*".

4.2.2 Results

For the purposes of determining whether participants made any inferences regarding activity habituality, I modeled *belief change*, i.e. the difference between *pre-utterance* and *post-utterance* beliefs, or activity habituality estimates made with and without seeing the activity description. *Conventionally habitual* and *non-habitual* activities were modeled separately, as the conventionally *non-habitual* activity was used primarily as a control, and manipulations of common ground context did not otherwise target it. All factors were effect/sum coded.

***Conventionally habitual* activities ('Paid the cashier')**

Pre-utterance belief ratings (obtained from participants who did not see the activity descriptions) showed that *ordinary* context activities are perceived as highly habitual (85.79 on a 0-100 scale). As predicted, *post-utterance belief* ratings (obtained from participants who saw the here, redundant, event descriptions) show lower habituality for the *ordinary context* activities (72.37) than *pre-utterance belief* ratings.

Wonky context activities (i.e., the condition where the *conventionally habitual* activity was made non-habitual by the common ground context) are perceived as relatively non-habitual a priori (48), and there was little change in participants' ratings (45.71 for *post-utterance beliefs*). The results are illustrated in Fig. 4.2, using violin plots.

A linear mixed effects regression analysis, the results of which are summarized in Table 4.1, showed that the interaction between context and belief measure is statistically reliable (β=-10.77, p<.001). This interaction is driven by lowered activity habituality ratings when the readers see the utterance in a *ordinary* context (β=-13.21, p<.001).

[5]However, the results below largely mirror the results of a within-subjects version of the study reported in Kravtchenko & Demberg (2015).

In this experiment as well as the two following experiments, I use linear mixed effects models with the maximal random effects structure that was justified by the design. This means that I included by-subject random intercepts and slopes for common ground context (*ordinary* / *wonky*) and belief measure (*pre-utterance* / *post-utterance*), as well as by-item random intercepts and slopes for both factors and their interaction (Barr et al., 2013). By-subject random slopes for the interaction were not included in the model, because there were no repeated measures for the interaction (each subject saw each condition only once). *P*-values were obtained using the Satterthwaite approximation for degrees of freedom, as implemented in the lmerTest package (Kuznetsova et al., 2017).

Table 4.1: Experiment 1: *conventionally habitual* (*cashier-paying*) activity analysis. This table shows the beta coefficients associated with each main effect in the model, as well as corresponding standard errors, *t*-values, and significance levels.

	β	SE(β)	**t**	**p**
Intercept	63.03	1.84	34.32	**<.001**
Common Ground: Ordinary	32.38	3.33	9.72	**<.001**
Belief: Post-utterance	-7.83	1.71	-4.58	**<.001**
Common Ground * Belief	-10.77	2.40	-4.50	**<.001**

These results show that, as predicted, when a *conventionally habitual* activity is explicitly described in a *ordinary* common ground context (i.e. a context in which the activity can be automatically inferred), many readers infer that the *conventionally habitual* activity must in fact be *non-habitual*; i.e., unusual for the individual who is the subject of the story, and therefore worth mentioning explicitly.

Conventionally *non-habitual* activities ('Bought some apples')

There was little change in participants' ratings of conventionally *non-habitual* activities from *pre-utterance beliefs* to *post-utterance beliefs* (*ordinary*: 40.8 to 42.47; *wonky*: 38.49 to 39.56), see Fig. 4.3.

A linear mixed effects regression analysis showed that estimates of activity habituality do not vary with the common ground context, nor are they conditioned on the utterance describing the activity (see Table 4.2). This is also consistent with predictions, and indicates both that the context alteration does not inherently cause a change in activity habituality estimates (regardless of how script-central the activity is), and that conventionally *non-habitual* activities, given the *ordinary* context, are not interpreted as less habitual when mentioned.

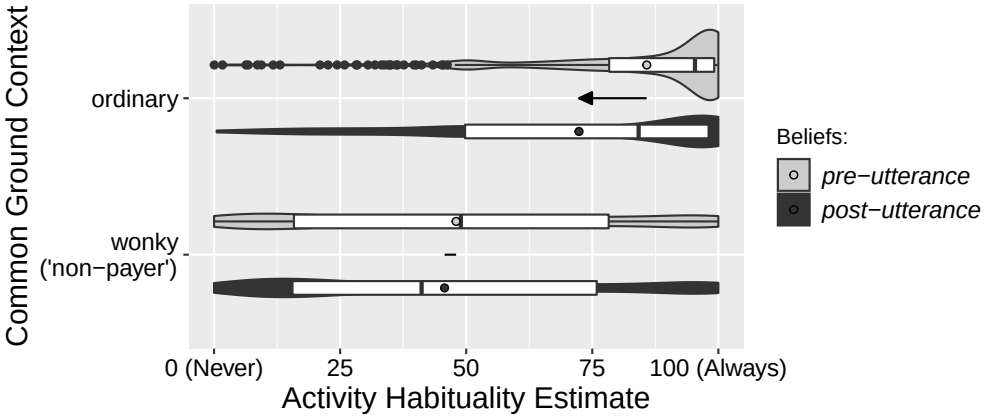

Figure 4.2: Experiment 1: *conventionally habitual* (*cashier-paying*) activity analysis. This plot shows changes in activity habituality estimates depending on whether the utterance is seen, as well as whether the context causes the utterance activity to be perceived as non-habitual. Violin plots, overlaid with box plots, show the distribution of estimates. A violin plot is simply a smoothed and mirrored histogram: the fatter the distribution at a given point, the more instances there are of that particular activity habituality estimate. Circles represent mean values. Arrows show statistically significant differences between *before/pre-utterance* and *after/post-utterance* ratings.

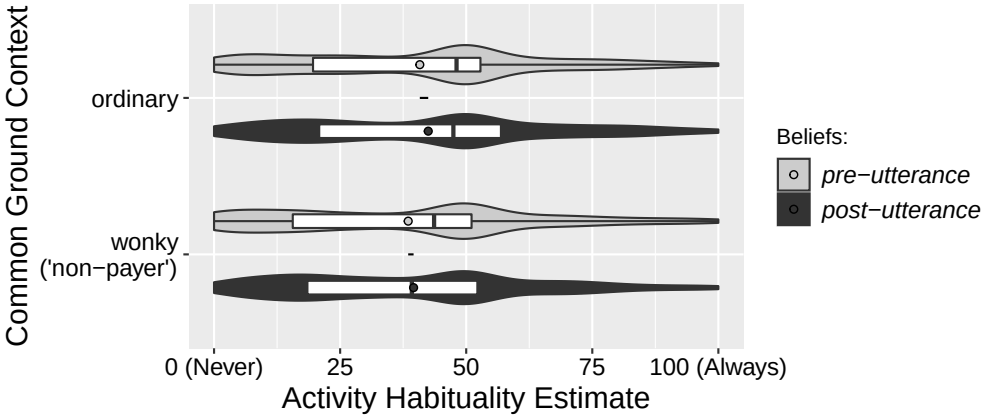

Figure 4.3: Experiment 1: conventionally *non-habitual* (*apple-buying*) activity analysis.

Table 4.2: Experiment 1: conventionally *non-habitual* (*apple-buying*) activity analysis.

	β	SE(β)	**t**	**p**
Intercept	40.29	1.86	21.69	**<.001**
Common Ground: Ordinary	2.88	2.07	1.39	0.2
Belief: Post-utterance	1.34	1.85	0.73	0.5
Common Ground * Belief	0.01	2.14	0.00	1

4.2.3 Discussion

The results of the first experiment indicate that comprehenders do, in fact, take note of informational redundancy, in the form of explicit mention of overly habitual activities. They appear to perceive these utterances as potentially violating conversational norms, at face value, and resolve this apparent violation by reinterpreting the activities described as *non-habitual*. On average, participants rate conceptually predictable activities as less habitual if they see them mentioned overtly, in contrast to all other activities. In other words, comprehenders react to redundancy as they typically do to other apparent violations of conversational norms – by assuming, where possible, an implied non-literal meaning, or alternate background world state, that resolves the apparent violation. This partially contradicts Grice (1975)'s ambivalence over the existence of a constraint on redundancy, and relatively equivocal evidence from studies of informationally redundant nominal modification.

These results rule out the "no inference" hypothesis outlined in Section 3.3.2, and raise two questions that are addressed in the following experiments, regarding the importance of (implicit) prosody, and of the speaker signaling the intentionality of the activity description. First, exclamatory prosody may serve multiple purposes: it may signal surprise as to the course of described events, a speaker's intentionality in communicating a piece of information[6], the importance and relevance of the information conveyed to the general discourse and comprehender's interests, and that the information thus emphasized constitutes an "encapsulated" message in its own right (rather than serving as a temporal or causal anchor[7]). Although it could be argued that the exclamation point, as a signal of surprise, forces a relative 'non-habitual activity' interpretation, independently of utterance informativity, this is not a likely explanation, as no signs of a similar effect are present in any of the other conditions.

Therefore, the first question is: how generalizable is the effect, and does the inference arise in contexts that do not implicitly signal the unexpectedness of the information conveyed (beyond the point that it is mentioned at all)? There is relatively

[6]I.e., the speaker displays clear and conscious intent to draw to the comprehender's attention the fact that a given event occurred – as opposed to stalling for time, thinking of something to say, aborting a previously planned utterance, simply being uncooperative, and so forth.

[7]For example: *He paid the cashier. Then he noticed it was his classmate.*

little work on the question of which contextual cues specifically comprehenders employ in computing context-dependent inferences, as well as how these cues influence final interpretation. To test this, in Experiment 2 I use a discourse marker ("*Oh yeah, and...*") which does not clearly signal surprise – but does frame the event description as intentionally conveyed, as important/relevant to the topic at hand, and as an "encapsulated message."

A secondary question raised is whether informational redundancy itself, in general, is sufficient to trigger such an inference. As mentioned previously, I start from the premise that rational speakers mention only that which cannot be automatically inferred by the comprehender. A charitable comprehender may be expected to expend considerable effort on rescuing the assumption of a cooperative or rational speaker (Davidson, 1974). If only activities under a certain threshold of habituality deserve mention, and if comprehenders are highly averse to concluding that the speaker is *irrational*, then comprehenders could conclude that the activity mentioned is relatively unusual, independently of any specific emphasis on the utterance. In general, most types of inferences, if they occur, should occur as long as the semantic content of the utterance remains constant (cf. the "non-detachability" hypothesis).

On the other hand, pragmatic inferences must be calculable (Levinson, 2000) – presumably not only in principle, but without excessive effort. Utterances must also be attended to closely enough in the first place, before they may trigger any inferences (Wilson & Sperber, 2004). That is, particularly for non-generalized (context-sensitive) inferences, the context should offer sufficient support that the reader can infer the speaker's intent, or a plausible background state, with reasonable certainty. It's not clear, in the case of the utterances described here, if the blatant redundancy in itself constitutes sufficient support. Likewise, while rational and efficient speakers may only mention activities that are not easily inferable, forcing a comprehender to expend significant effort on recovering an utterance's intended meaning or broader significance is not particularly rational behavior. The degree of "intentionality" on the part of the speaker (also signaled in the stimuli by the exclamation mark) may affect comprehenders' willingness and effort in guessing any implied meaning, as an utterance that may be a stray thought, uttered without any specific intent, may not be worth much effort to attempt to decipher (cf. the "form sensitivity" hypothesis). To test whether informational redundancy in itself is sufficient to trigger a *habituality* inference, or whether some amount of discourse or prosodic emphasis is necessary for its generation, in Experiment 3 I present readers with the same task and stimuli, but strip the event description of prosodic or discourse cues signaling speaker intentionality.

4.3 Experiment 2: Implicit intent signaled by discourse markers

The second experiment tests whether the effect of interest, of informationally redundant event descriptions being interpreted by readers as signaling activity non-*habituality*, is generalizable. I replace the exclamation point with a non-prosodic discourse marker that signals speaker intentionality and utterance relevance (but crucially, not surprise). In this experiment, I frame the informationally redundant event description as an apparent recalling of information specifically intended to be mentioned to the comprehender, and implicitly relevant to the material just discussed: "*Oh yeah, and [he paid the cashier].*" This discourse marker does not clearly signal surprise at the activity having been engaged in, nor does it explicitly support the intended inference otherwise – and in contrast to the exclamation mark in Experiment 1, is a non-prosodic manipulation of the event description. I therefore consider it a good test of whether the effect generalizes beyond the specific context used in the first experiment.

4.3.1 Methods

Participants

700 eligible participants (787 total; median age bracket 26-35; 51.3% female) were recruited on Amazon Mechanical Turk. 87 participants were excluded from analysis (11.05%), following the same exclusion criteria as applied in Experiment 1.

Design

The design of this experiment was motivated by the same considerations as Experiment 1 – with the exception of how the event description was framed. Instead of marking the target utterance with an exclamation mark, the same utterance was framed as a piece of information the speaker had just recalled, apparently having previously intended to mention it to the comprehender:

(7) "John just came back from the grocery store. **Oh yeah, and he paid the cashier.**"

The *oh yeah...* discourse marker does not conventionally signal surprise, and therefore does not potentially signal the specific inference that I am testing for. It does, however, imply speaker intent behind conveying precisely this message, the importance and relevance of the message to the current discourse and comprehender – as well as that the message stands alone, and is not intended to simply serve as causal or temporal scaffolding for a further message/event.

Materials

The same 24 stimuli were used as in Exp. 1. In this case, the critical utterance was prepended by "*Oh yeah, and...*":

(8) Ordinary context

[1] John often goes to the grocery store around the corner from his apartment.
[2] Recently, he came home from the store with groceries. When he came in, he saw his roommate Susan in the hallway, and started talking to her about his trip to the store. As he went to the kitchen to put his groceries away, Susan went to the living room, where their roommate Peter was watching TV.
[3] Susan said to Peter: "John just came back from the grocery store.
[4a] **Oh yeah, and** he *paid the cashier*$_{habitual}$." | [4b] **Oh yeah, and** he *got some apples*$_{non\text{-}habitual}$."

Procedure

The procedure was identical to that of Exp. 1.

Measures

The same response measures as in Exp. 1 were used to estimate *pre-utterance beliefs* and *post-utterance beliefs*.

4.3.2 Results

As in Experiment 1, to determine whether participants made inferences regarding activity habituality, I modeled *belief change* - the difference between *pre-utterance* and *post-utterance* beliefs. *Conventionally habitual* and conventionally *non-habitual* activities were again modeled separately. All factors were effect/sum coded.

Conventionally habitual activities

As predicted, *pre-utterance belief* ratings for *ordinary context* activities showed that these activities are judged to be highly habitual (84.71). As in Experiment 1, *post-utterance beliefs* about the habituality of *ordinary context* activities were significantly lower (73.84), and *wonky* common ground estimates remained stable (47.45 *pre-utterance* to 47.47 *post-utterance*).

A linear mixed effects regression analysis, the results of which are summarized in Table 4.3, showed an interaction between context and belief measure (β=-11.71, p<.001), which is driven by lowered activity habituality ratings when the readers see the utterance in an ordinary context (β=-11.11, p<.001). All model specifications

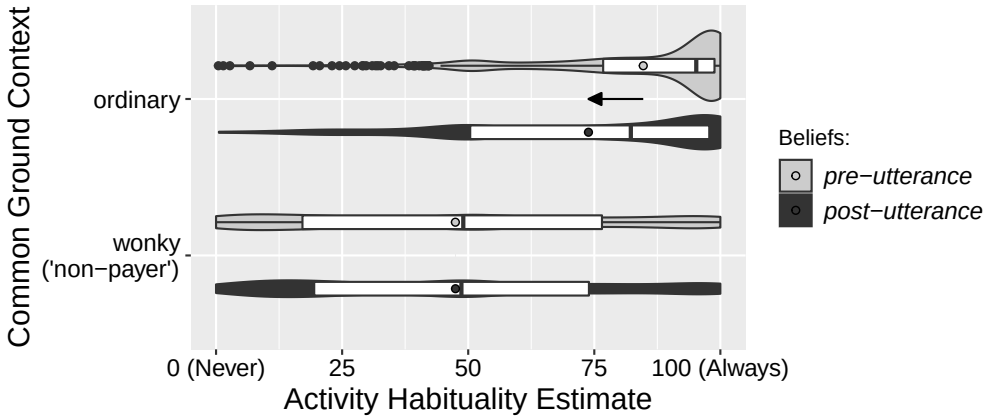

Figure 4.4: Experiment 2: *conventionally habitual* (*cashier-paying*) activity analysis.

are as described in Exp. 1. A plot illustrating the interaction can be seen in Fig. 4.4, which shows a pattern of results that is remarkably quantitatively and qualitatively similar to that of Exp. 1. Exp. 1 and 2 are compared directly, and to Exp. 3, in Section 4.5.

Table 4.3: Experiment 2: *conventionally habitual* (*cashier-paying*) activity analysis.

	β	SE(β)	**t**	**p**
Intercept	63.58	1.85	34.33	**<.001**
Common Ground: Ordinary	31.60	3.35	9.43	**<.001**
Belief: Post-utterance	-5.31	1.38	-3.83	**<.001**
Common Ground * Belief	-11.71	2.03	-5.76	**<.001**

These results support the prediction that readers perceive informationally redundant utterances as marked, and make pragmatic inferences (regarding activity *habituality*), regardless of whether implicit prosody or other markers conventionally associated with surprisal are present.

Conventionally *non-habitual* activities

In contrast to Experiment 1, there was some increase in participants' ratings of conventionally *non-habitual* activities from *pre-utterance beliefs* (*ordinary*: 40.3 to 43.22; *wonky*: 37.74 to 43.05), see Fig. 4.5.

A linear mixed effects regression analysis showed that estimates of activity habituality increase slightly when the utterance describing the conventionally *non-habitual* activity (see Table 4.4) is visible (β=5.09, p<.01).

While not identical to the results of the first experiment (which showed a slight numerical increase in rating only), this is consistent with a peripheral prediction

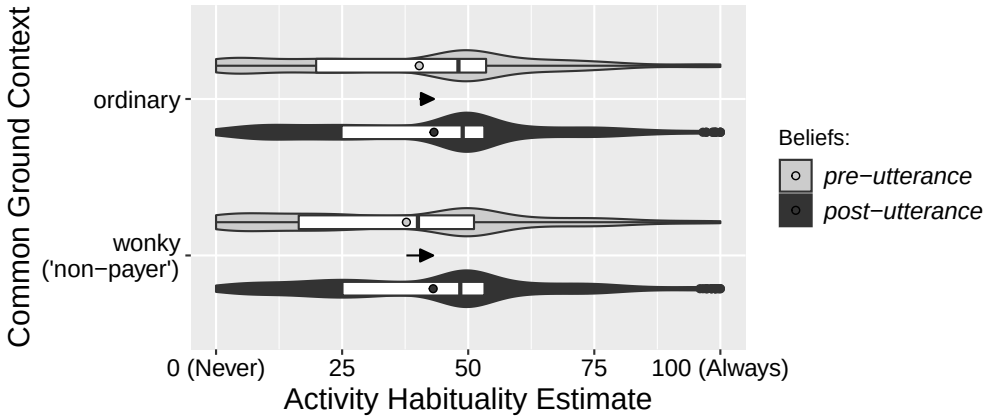

Figure 4.5: Experiment 2: conventionally *non-habitual* (*apple-buying*) activity analysis.

made prior to running the experiments: simply mentioning a non-habitual, or non-redundant activity may increase the perception of its habituality, by providing some evidence that, e.g., *John* is at least an occasional *apple purchaser*. As the direction of this effect does not change the interpretation of these results, I leave it aside for future exploration.

Table 4.4: Experiment 2: conventionally *non-habitual* (*apple-buying*) activity analysis.

	β	SE(β)	**t**	**p**
Intercept	40.99	1.85	22.14	**<.001**
Common Ground: Ordinary	0.95	1.83	0.52	0.6
Belief: Post-utterance	5.09	1.78	2.86	**<.01**
Common Ground * Belief	-1.22	1.55	-0.79	0.4

4.3.3 Discussion

Together with Experiment 1, these results show that comprehenders take note of informational redundancy, and make pragmatic inferences to reconcile apparent redundancy with their expectations of utterance utility. This is further evidence against the "no inference" hypothesis, and indicates that the effect is generalizable, and not dependent on conventional indicators of activity *non-habituality*, such as implicit exclamatory intonation. The results of Experiments 1 and 2, however, do not make it possible to distinguish between the 2nd and 3rd hypotheses ("non-detachability" vs. "form sensitivity"), as they leave open the question of whether the *habituality* inference is dependent on some degree of intentionality-signaling, or applies independently of discourse context. Experiment 2 provides some support for the "non-detachability"

hypothesis, as the magnitude of the inference remains entirely stable, even as the form of intention or relevance signaling is substantially changed.

If the effect is dependent on some degree of relevance or intentionality signaling, this would support the "form sensitivity" hypothesis over the "non-detachability" hypothesis, by suggesting one of the following. Comprehenders may be relatively unwilling to expend substantial effort on decoding a merely plausible inference in the absence of evidence that doing so is worth it, and that the utterance has some amount of import. Similarly, they may stop short in their efforts, on the assumption that it is more likely that speakers would occasionally violate this particular conversational norm, than that they would provide insufficient evidence that the utterance communicates something of note. Finally, they may simply be generally tolerant of informational redundancy, unless context suggests that the redundancy has a "point." Experiment 3 presents the same task and materials to participants, but removes the prosody or discourse markers that signal relevance and speaker intent.

4.4 Experiment 3: Removing evidence of speaker intent

To investigate whether explicitly signaling speaker intent has an influence on the strength of *habituality* inferences, I designed a third experiment which differs only in the absence of specific prosodic or discourse markers, or evidence for the relevance/informativity of the activity description. The prediction is that while the effect may be attenuated somewhat, comprehenders should nevertheless make a measurable attempt to compensate for a violation in expected informational utility (i.e., while there may be some degree of "form sensitivity," the inference should nevertheless arise).

4.4.1 Methods

Participants

700 eligible participants (759 total; median age bracket 26-35; 51.6% female) were recruited on Amazon Mechanical Turk. 59 participants were excluded from analysis (7.77%), following the same exclusion criteria as applied as in previous experiments.

Design

The design was motivated by the same factors as Experiments 1 and 2, but all markers of relevance were removed from the activity description:

(9) "John just came back from the grocery store. **He paid the cashier.**"

In this case, there is no clear signal indicating the relevance or informativity of the utterance. One could plausibly imagine the event description, in this case, to be 'filler material,' only semi-intentionally uttered while the speaker is planning what to say next, or as (planned, but then possibly abandoned) temporal or causal scaffolding for a more important event to be described, such as in:

(10) "John just came back from the grocery store. He paid the cashier. *He then realized he'd forgotten his driver's license!*"

Materials

The same 24 stimuli were used as in the previous experiments. The only alteration from Experiment 1 was the substitution of the exclamation point with a period:

(11) Ordinary context

[1] John often goes to the grocery store around the corner from his apartment.	
[2] Recently, he came home from the store with groceries. When he came in, he saw his roommate Susan in the hallway, and started talking to her about his trip to the store. As he went to the kitchen to put his groceries away, Susan went to the living room, where their roommate Peter was watching TV.	
[3] Susan said to Peter: "John just came back from the grocery store.	
[4a] He *paid the cashier*$_{habitual}$."	[4b] He *got some apples*$_{non\text{-}habitual}$."

Procedure

The procedure was identical to that of previous experiments.

Measures

The same response measures as in the previous experiments were used to estimate *pre-utterance beliefs* and *post-utterance beliefs*.

4.4.2 Results

As in previous experiments, I modeled the difference between *pre-utterance* and *post-utterance* beliefs. *Conventionally habitual* and conventionally *non-habitual* activities were modeled separately. All factors were effect/sum coded.

Conventionally habitual activities

As in the previous experiments, *pre-utterance belief* ratings showed *ordinary context* activities to be highly habitual (85.59), and *wonky context* activities to be less habitual

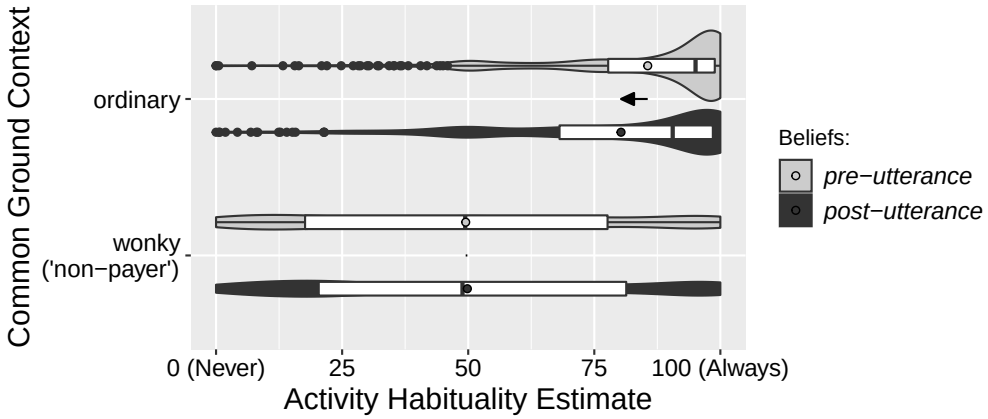

Figure 4.6: Experiment 3: *conventionally habitual* (*cashier-paying*) activity analysis.

(49.5). Consistent with predictions, *post-utterance beliefs* are significantly lower in the *ordinary context condition* (80.3), but less so than in the previous two experiments. Exp. 3 is compared directly to Exp. 1 and 2 in Section 4.5.

A linear mixed effects regression analysis, the results of which are summarized in Table 4.5, showed an interaction between context and belief measure (β=-5.4, p<.01), which is driven by lowered activity habituality ratings when the readers see the utterance in an ordinary context (β=-4.87, p<.001). All model specifications are as described in Exp. 1 and 2. A plot illustrating the interaction can be seen in Fig. 4.6.

Table 4.5: Experiment 3: *conventionally habitual* (*cashier-paying*) activity analysis.

	β	SE(β)	**t**	**p**
Intercept	66.38	1.87	35.40	**<.001**
Common Ground: Ordinary	33.21	3.40	9.77	**<.001**
Belief: Post-utterance	-2.20	0.93	-2.36	**<.05**
Common Ground * Belief	-5.40	1.75	-3.10	**<.01**

These results indicate that, consistent with predictions and the results of Exp. 1 and 2, when an easily inferable activity is overtly mentioned in a *ordinary* common ground context, comprehenders do infer some degree of activity non-habituality, even without implicit prosody or discourse markers putting additional emphasis on the utterance.

Conventionally *non-habitual* activities

In contrast to Experiment 1 and similar to Experiment 2, there was some increase in participants' ratings of conventionally *non-habitual* activities from *pre-utterance* to *post-utterance* beliefs (*ordinary*: 41.08 to 46.46; *wonky*: 37.61 to 44.42), see Fig. 4.7.

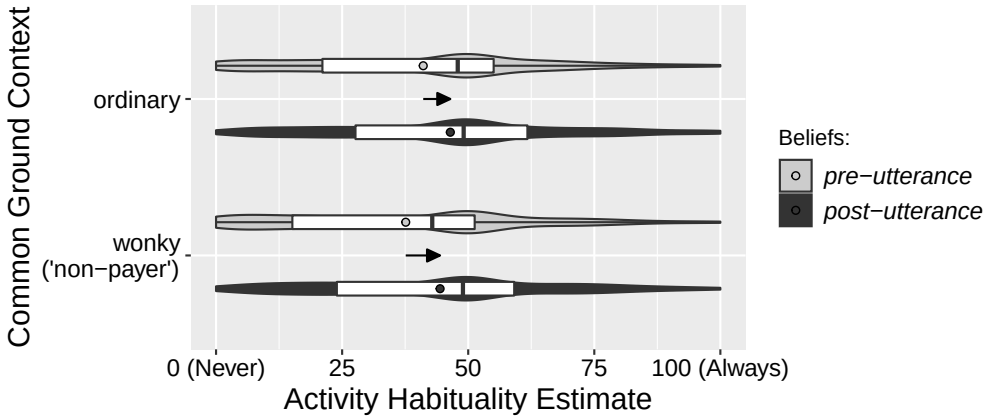

Figure 4.7: Experiment 3: conventionally *non-habitual* (*apple-buying*) activity analysis.

A linear mixed effects regression analysis showed that estimates of activity habituality do not vary with changes in the common ground context (or common ground *wonkiness*), but do increase slightly when the utterance describing the conventionally *non-habitual* activity (see Table 4.6) is visible (β=6.88, p<.001). As in the case of Exp. 2, I suspect that explicitly mentioning a relatively unusual activity leads participants to believe that activity to be slightly more habitual than they would otherwise assume.

Table 4.6: Experiment 3: conventionally *non-habitual* (*apple-buying*) activity analysis.

	β	SE(β)	**t**	**p**
Intercept	42.12	2.12	19.84	**<.001**
Common Ground: Ordinary	2.29	2.41	0.95	0.4
Belief: Post-utterance	6.88	1.77	3.88	**<.001**
Common Ground * Belief	-1.39	1.72	-0.81	0.4

4.4.3 Discussion

In contrast to the results of the first two experiments, these results suggest that, when informationally redundant utterances are presented without a signal of speaker intent and utterance relevance comprehenders are relatively unlikely to draw *habituality* inferences. This is consistent with the "form sensitivity" hypothesis described in Section 3.3.4. It is also consistent with the premise that, while rational speakers may typically avoid making utterances that have no apparent informational utility, and while such utterances may prompt pragmatic inferences on the part of comprehenders, such inferences are dependent on the degree to which the utterances are perceived as intentional. Further, while the results are not consistent with a strong form of the

"non-detachability" hypothesis, they do broadly suggest that redundancy generates inferences regardless of form of delivery.

It should be noted, however, that this is not what was found in the experiments that are being replicated here – where the inference disappeared entirely without prosodic or discourse emphasis (strongly supporting the "form sensitivity" hypothesis, and at odds with the "non-detachability" hypothesis). Although the replicated experiments were not as highly powered, the difference may be due to stimulus redesign – in Appendix C, I speculate as to why this might be the case.

4.5 Cross-experiment analysis and gradience of the non-habituality effect

In this section, I directly compare the results of the three experiments. I predict that informationally redundant utterances can trigger *habituality* inferences of similar magnitude independently of whether one uses an explicit marker of surprisal: in other words, that the effect is generalizable. However, I also predict that the effect is significantly attenuated in absence of a prosodic or discourse marker which signals relevance, speaker intent, and a desire to draw the listener's attention.

4.5.1 *Conventionally habitual* activities

To directly compare the three experiments, I run a $3 \times 2 \times 2$ linear mixed effects regression analysis of *conventionally habitual* activities. I model *belief change* (*pre-utterance* vs. *post-utterance* beliefs), as a function of common ground (*ordinary* vs. *wonky*), as well as the between-subject discourse marker manipulation ('*!*' vs. '*Oh yeah, and*' vs. '.'). The first two factors were effect/sum coded. I used Helmert coding for the 3-level experiment factor, as this made it possible to make the comparisons of theoretical interest: Exp. 1 vs. Exp. 2 ('*!*' vs. '*Oh yeah, and*'), and then Exp. 3 vs. Exp. 1 and 2 grouped together ('.' vs. the relevance markers).

The regression analysis showed a significant three-way interaction between relevance marker presence, common ground context, and belief measure: there was a significantly smaller *habituality* effect in Exp. 3 than in Experiments 1 and 2 (β=5.69, p<.01), and no significant difference between Experiments 1 and 2 (β=-0.6, p=.80).

I used the maximal converging model, with by-subject random intercepts and slopes for common ground context (*ordinary / wonky*) and belief measure (*pre-utterance / post-utterance*), by-item random intercepts and slopes for both factors and their interaction, and a by-item random slope for experiment. By-subject random slopes for the interaction were not included in the model due to lack of within-subject repeated measures. The random slope for the full (by-item) experiment by common ground by belief measure interaction was not included due to non-convergence. A plot illustrating the comparison can be seen in Fig. 4.8.

Table 4.7: Experiments 1-3: *conventionally habitual* (*cashier-paying*) activity analysis.

	β	SE(β)	**t**	**p**
Intercept	64.34	1.78	36.18	**<.001**
'!' vs. 'Oh yeah...'	0.48	0.85	0.56	0.6
'.' vs. Relevance Markers	3.08	0.78	3.96	**<.001**
Common Ground: Ordinary	32.39	3.22	10.05	**<.001**
Belief: Post-utterance	-5.06	1.04	-4.86	**<.001**
'!' vs. 'Oh yeah' * Common Ground	-0.61	1.43	-0.43	0.7
'.' vs. Relevance Markers * Common Ground	1.24	1.24	1.00	0.3
'!' vs. 'Oh yeah' * Belief	2.50	1.30	1.92	0.1
'.' vs. Relevance Markers * Belief	4.50	1.13	3.99	**<.001**
Common Ground * Belief	-9.34	1.11	-8.41	**<.001**
'!' vs. 'Oh yeah' * CG * Belief	-0.60	2.35	-0.26	0.8
'.' vs. Relevance Markers * CG * Belief	5.69	2.03	2.80	**<.01**

The results are summarized in Table 4.7. As predicted, the effect holds regardless of which relevance marker is used, and in fact there is no statistically significant difference between the two markers. Further, the effect size of the common ground by belief measure interaction is significantly smaller in the absence of the markers; in other words, participants are significantly less likely to make a *habituality* inference in the absence of a prosodic or discourse marker signaling relevance or intentionality.

The effect direction is consistent across experimental items, with by-item common ground by belief measure interaction effect sizes ranging from -5.35 to -12.89. I again note here that this set of 3 experiments is a replication of a previously run set with somewhat less naturalistic stimuli, a full description of which can be found in the supplementary materials linked to in the previous paragraph. In addition, the 'exclamation' experiment in that set is a further replication of a within-subjects version (same stimuli), previously published as Kravtchenko & Demberg (2015), where participants updated their own ratings after seeing the utterance. I therefore argue that this is overall a robust and replicable effect.

This result clearly favors the "form sensitivity" hypothesis described in Section 3.3.4 over a strong version of the "non-detachability" hypothesis (which might predict an effect of the same magnitude for all experiments). I conclude that in the absence of a clear signal of utterance relevance or speaker intentionality, comprehenders are either less likely to attempt to resolve the violation, resolve it in a manner that is not reflected in the response measures, or do not detect the violation in the first place. The first possibility is supported by observations that comprehenders approach speaker utterances *charitably*, and may expend significant effort on interpreting them in a manner that is consistent with the speaker making cooperative conversational

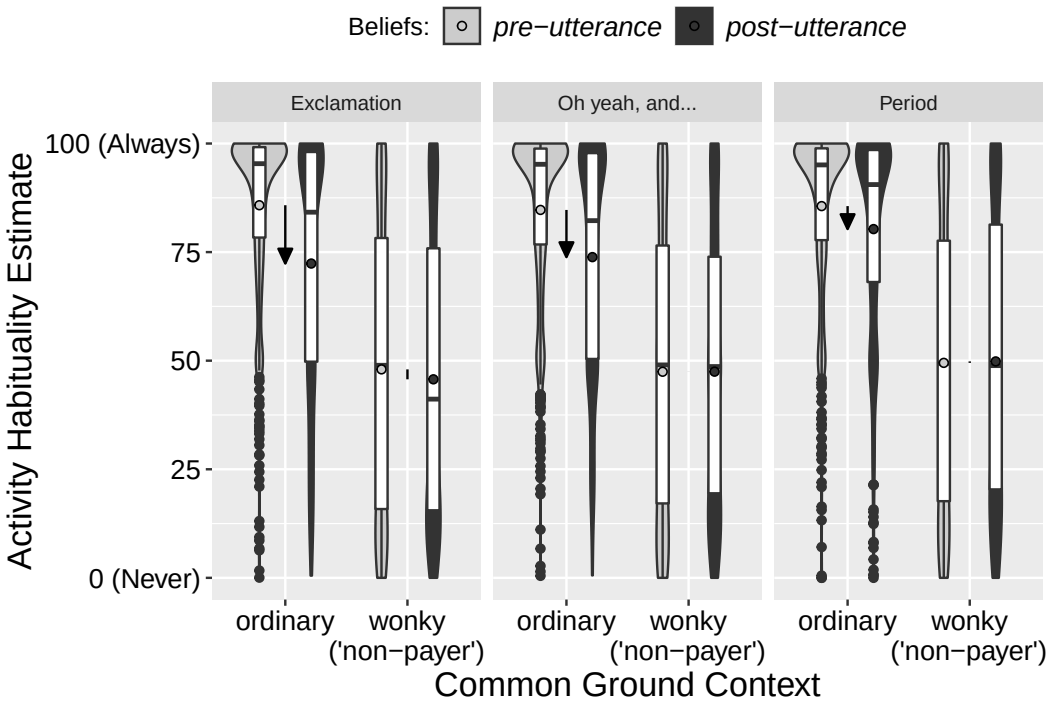

Figure 4.8: Experiments 1-3: *conventionally habitual* (*cashier-paying*) activity analysis.

choices (Davidson, 1974). However, it is also possible that comprehenders are less 'charitable' in general when presented with oddly phrased psycholinguistic stimuli in an artificial setting – as well as less motivated to expend cognitive effort on calculating a non-obvious inference in a non-interactive environment, on the basis of an utterance that their attention is not otherwise drawn to.

Less charitable comprehenders, who may detect the redundancy but fail to in some way resolve it, may assume that the speaker is odd or not a particularly cooperative speaker, or perhaps that they are having production difficulties. Another possibility is that they assume the speaker is in the process of planning a more informative utterance (where, for example, the description might serve as a temporal/causal anchor; see Example 7). Determining which strategies comprehenders do in fact resort to, and in which contexts, is left to future work. Finally, there is the possibility that, given the non-interactive experimental setting, comprehenders are processing the utterances at a relatively shallow level, and absent some (prosodic, discourse) indication that an utterance is somehow important, they do not expend effort on it (Sanford et al., 2006). Along the same vein, comprehenders may be assuming that the utterances are *intended* to be processed at a relatively shallow level, or not given any particular attention, or else the speaker would have drawn more attention to the utterance. To note, it has frequently been observed that comprehenders often do not make expected inferences in behavioral studies, for reasons that are not yet fully known (cf., Noveck & Posada, 2003). Determining whether this plays a role in these studies is left to future work, as is the question of whether similar or stronger effects

can be observed in less artificial, and/or more interactive settings.

4.5.2 Is the effect of habituality on pragmatic inferences gradient?

Fig. 4.9 plots the measured average activity habituality, with and without seeing the target utterance, for each item in each condition, for all three experiments. The diagonal dashed line demonstrates what the "no inference" hypothesis would predict: i.e., no effect of the utterance on belief change (*pre-utterance* ratings mapping straightforwardly onto *post-utterance* ratings). Points found above the line indicate that for those items, participants were more certain, for example, that *John usually buys apples* when the story mentioned that "he got some apples." Points below the line indicate a *habituality inference*: e.g., mentioning that "he paid the cashier" causes people to believe that *John does not usually pay the cashier*.

In Experiment 1 (exclamation mark), it can be seen that for *ordinary* common grounds, and *conventionally-habitual* activities (e.g., *paying the cashier* given an ordinary common ground), most data points fall below the line, indicating a habituality inference. Interestingly, one can also see a gradual 'trend' towards *non-habituality* in the other three (non-redundant) conditions: items that are similar to *ordinary habitual* items, in terms of pre-utterance habituality estimates, are more likely to trigger habituality inferences. In contrast, items with low pre-utterance habituality estimates show the opposite effect: i.e., if it's mentioned that an individual engaged in a particularly non-habitual activity, it leads comprehenders to believe that the individual is more likely to engage in that activity habitually. The same observations also hold for Experiment 2.

In Experiment 3 (period), one can again see a gradual effect of *pre-utterance* beliefs regarding activity habituality on the likelihood of a *habituality inference*, but this time the slope of the regression line is shifted upwards (Exp. 1: β=0.64; Exp. 2: β=0.64; Exp. 3: β=0.76). It can still be seen, however, that there is a gradient difference between highly expected vs. relatively expected events, in terms of likelihood of a habituality inference occurring.

Taken together, one can see in these figures that the exclamation mark and the '*oh yeah...*' discourse marker, as signals of speaker effort and intentionality, make it more likely that habituality inferences will arise for *ordinary* common ground, *habitual* activity mentions. Furthermore, one can see that the effect of pre-utterance beliefs on habituality inferences is clearly gradient rather than binary: relatively more habitual activities, in all conditions, generally elicit larger *habituality* inferences.

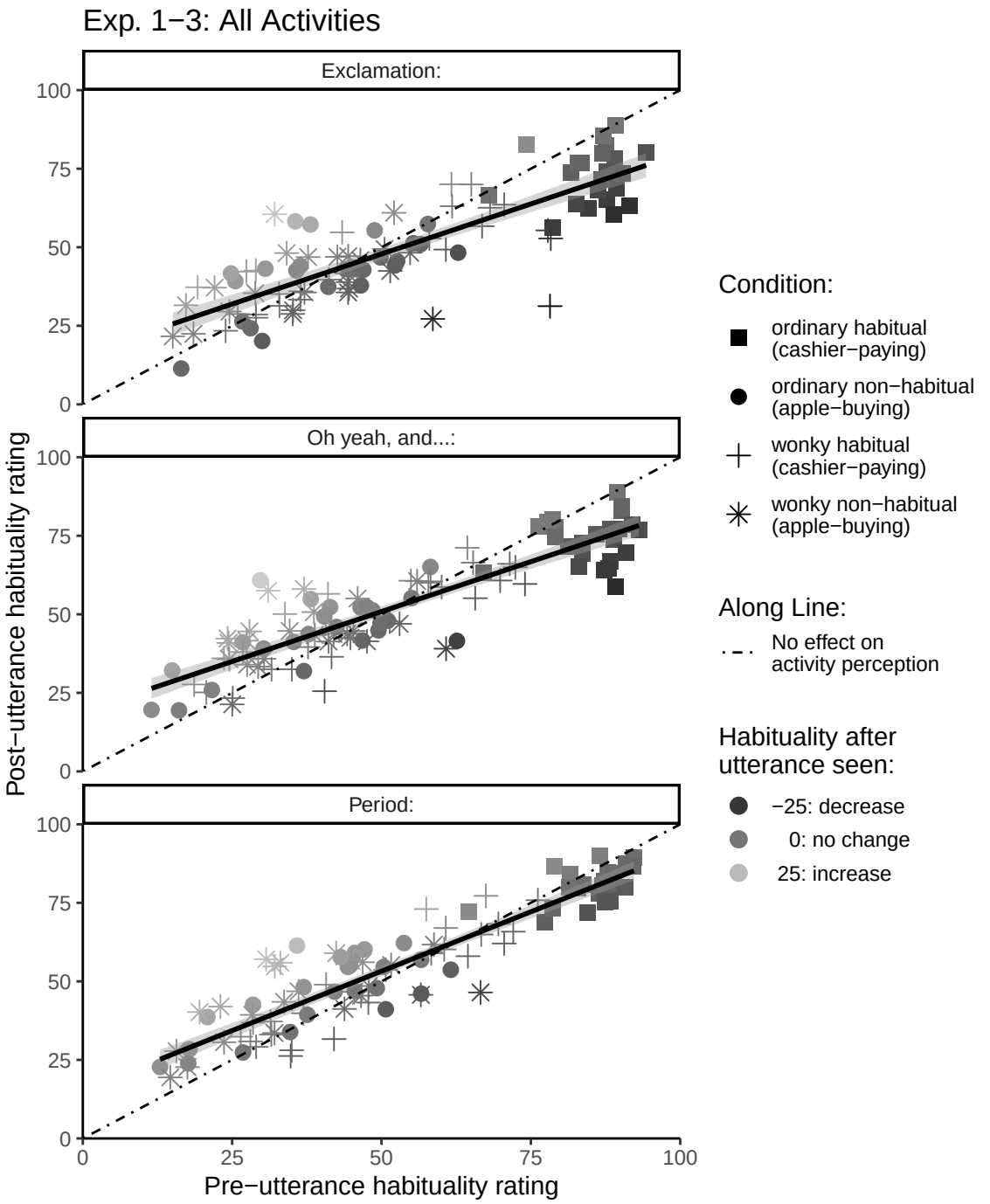

Figure 4.9: These plots show by-item belief change for all conditions of my three experiments. The dotted diagonal line represents the "no inference" hypothesis; i.e., what one would expect the data to look like if the critical utterance had no effect on habituality beliefs. The solid black line is a regression line with 95% CIs across all conditions. The shading of the points represents the degree and direction of *belief change*: negative/black indicates a *non-habituality* inference; positive/light gray indicates a perception of increased habituality.

4.6 General discussion

Taken together, this series of experiments shows that comprehenders react to informationally redundant utterances by shifting their beliefs about the common ground, such that the utterances are more 'informative' in context, thus increasing their utility. In other words, comprehenders expect for speaker utterances to have a certain level of informational utility, and they adjust their beliefs about the world and/or utterance meaning when such expectations are violated. This constitutes clear evidence that there is an upper bound on how (over-)informative speakers may be without altering the meaning of their utterances, or potentially distorting the intended meaning. This restriction on speaker redundancy is not straightforwardly predicted by the UID hypothesis, which does not concern itself with non-literal utterance interpretation, and therefore arguably misses a critical aspect of what it means for a comprehender to recover the intended message. This comprehender behavior, however, is consistent with accounts in pragmatic theory of what constitutes 'cooperative' communicative behavior (Grice, 1975). In Chapters 7-9 I explore an alternative account of utterance production, which takes into account the possibility of non-literal utterance interpretations in determining what constitutes a successful, efficient communicative act.

Further, as the third experiment shows, the *habituality* effect is significantly modulated by how the utterance is framed in the discourse, supporting the hypothesis that inference strength is sensitive to utterance form. This particular effect is broadly consistent with a UID-based consideration of more effortful utterances being more likely to be perceived accurately by listeners – which in turn leads comprehenders to infer that the utterance is more likely to have been intentionally transmitted by the speaker (with the latter effect not taken into account by the UID hypothesis). Overall, these results provide robust evidence that comprehenders are sensitive to utterance redundancy, and that excessive redundancy alters their interpretation of the utterance's intended message, or of the background world state. Comprehenders overall appear quite willing to alter their situation models to accommodate unexpected redundancy, but particularly so when there's clear evidence of a specific speaker intent to transmit this particular utterance to the listener. In Chapter 8, I present a model of how listeners derive *habituality* inferences, and of how UID- and Relevance Theory-inspired considerations (of how increased speaker effort increases the likelihood of literal message recovery) make it possible for such a model to represent the influence of increased speaker effort on inference strength.

Further, with respect to pragmatic theory, while redundancy does not obviously impair comprehension of the literal message (unlike underinformativeness), as discussed in Section 3.2.1, comprehenders may nevertheless prefer or expect that speakers be relatively concise, as it allows them to receive more information in a shorter span of time. Excessive redundancy may make it more difficult to follow the point of a conversation, or to reliably distinguish important from unimportant information. In this context, it would be expected that comprehenders should perceive highly and

unnecessarily redundant utterances (e.g., '*yellow banana*,' '*he went shopping and paid the cashier!*') as such. Correspondingly, they should infer that the speaker is either not behaving rationally[8], or is conveying a message or background world state that is unusual from the comprehender's perspective. 'Moderately' redundant utterances, such as when a speaker points out '*the **long** fork*,' in the absence of another fork to compare, may be perceptually useful in many tasks (see Section 3.2.1), and at least provide information that can't otherwise be inferred from the rest of the utterance, while not requiring much additional articulatory effort on the part of the speaker. In contrast, it would be expected that the clearest evidence for how excessive redundancy is perceived should come from relatively costly utterances, such as those investigated here – what might be termed *highly redundant utterances*.

Another area in which these results contribute is that they illustrate a case in which comprehenders are willing to revise the assumed common ground of the discourse, in order to accommodate a perceived violation in the informational utility of an utterance. The redundant utterance violates conversational norms, or comprehender expectations of utterance utility, but only under the default assumed world state (one in which, for example, *shoppers* may be assumed to *pay cashiers*). In contrast, in an alternate *wonky* world state (e.g., one in which the shopper is a habitual non-payer), the utterance is no longer redundant, as it communicates relatively unexpected information. Hearing such an utterance, under an initially assumed default world state, therefore biases comprehenders towards assuming that the alternative, or *wonky*, world state holds. Unlike the strategy of shifting assumptions about intended utterance meaning, this is a strategy of accommodating potential violations of conversational norms that has not received much attention to date, with the notable exception of Degen & Tanenhaus (2015). The shifting of common ground assumptions appears to be an important, and surprisingly understudied strategy for interpreting utterances which, at face value, violate conversational norms. Neglecting it as a possibility in practice likely results in misinterpretation of online effects, and under-detection of pragmatic inferences in experimental work.

Finally, these results show that utterance features which do not contribute to the truth-conditional meaning of an utterance, in the form of implicit prosody or discourse markers, significantly influence the extent to which comprehenders are willing to draw an inference predicted by pragmatic theories of rational speaker behavior. Aside from the case of contrastive prosody (Bergen & Goodman, 2015; Kurumada et al., 2012; Ward & Hirschberg, 1985), this has not to date been systematically investigated in formal or experimental literature, and most likely also extends to other pragmatic phenomena. Such features are easily harnessed to increase the *ostensiveness* of utterances, and would therefore be predicted by Relevance Theory to minimally increase the strength of inferences drawn [Wilson2004]. In the present case, I argue that comprehenders are likely carefully weighing and evaluating multiple cues of how likely it is that a speaker intended to communicate a particular meaning – or that a de-

[8]See the discussion of Gricean vs. Bayesian rationality in Section 4.6.1 below.

viation from expected utterance form and/or meaning signifies a common ground or background state that is substantially different from what was initially assumed.

4.6.1 Processing Difficulty and Surprisal

In this subsection, I discuss the potential implications of this work for formal models of language processing, such as Surprisal Theory (Hale, 2001; Levy, 2008). A question that is raised for future research is whether encountering informationally redundant utterances results in measurable processing difficulty on the part of comprehenders. I further argue that this has significant implications for current models of language processing.

As Walker (1993) points out, informationally redundant utterances are common in natural dialog – they therefore cannot be regarded as edge cases, and must be integrated into any predictive models of language. I argue, however, that they pose several unique challenges for formal models of language processing. There are several potential sources of processing difficulty associated with such utterances, resulting on the one hand from processing the *surface form* (the particular string of words that comprises the utterance), and on the other hand from computing the pragmatic inference itself[9]. First, there could be processing difficulty associated with the (un)predictability of the *surface form* of the utterance: *John paid the cashier!* is an utterance one would not expect to hear in an ordinary context, as paying the cashier is normal, and reading unpredictable utterances such as this should cause some difficulty (Smith & Levy, 2013)[10]. Second, I work on the assumption that context-dependent implicatures incur processing cost (Levinson, 2000; Sedivy, 2007), although there is evidence that processing may be relatively rapid, provided the context adequately supports the inference (Degen & Tanenhaus, 2015; Grodner et al., 2010).

Speaker Rationality

First, I look at the link between Gricean notions of rationality, and an information-theoretic or Bayesian approach to rational speaker behavior. Rationality in the Gricean sense concerns whether speakers are constructing their utterances in a manner that is consistent with their goals, which is accurately communicating their message to a comprehender (Grice, 1975). To this end, *underinformativeness* (saying less

[9]Although there is debate currently over just how rapidly or efficiently comprehenders are able to make pragmatic inferences, much of the evidence converges on there frequently being some cost, even for relatively conventionalized inferences (Degen & Tanenhaus, 2016).

[10]It should however be noted that while pragmatic processing must, on some level, incur cost, it may be sufficiently small and poorly localized that one would have difficulty detecting it using traditional online measures (eye-tracking, self-paced reading). Further, it is possible that the ease of semantic integration, in the case of conventionally habitual activities, would eclipse any difficulty due to the unpredictability of the utterance itself – although the two exactly cancelling each other out, either way, is relatively unlikely.

than needed), for example, is clearly inconsistent with this goal. Saying *more* than is needed, however, does not clearly impair one's ability to accurately transmit a message – hence, the general uncertainty over whether overinformativeness violates Gricean norms. In the information-theoretic tradition, however, the speaker's goal is to expend no more energy than needed to accurately transmit a message (Jaeger, 2010). Expending more effort than required to accurately transmit a message is inefficient from the speaker's perspective, and therefore not particularly rational, even while it is worse from a communication standpoint to not expend *enough* effort. The two traditions therefore make roughly similar predictions – fairly weakly in the Gricean case, and more strongly in the information-theoretic: that redundancy should be avoided.

What the Gricean tradition, however, adds to this mix is an idea of how comprehenders might interpret deviations from the communicative norm. Traditional information-theoretic accounts in general make no predictions about how perceived utterance meaning might be altered if there's a mismatch between the expected and received utterance form, beyond the possibility that intended utterance form or structure may be assumed to be something different from what was received, or that perception itself may be altered by comprehender expectations. Jaeger & Buz (2017) do note that if the speaker's aim is to accurately communicate a message, then they must take into account the signal, or *surface form*, that comprehenders expect to hear for that particular message. If they produce something that deviates from the expected signal, then even if the utterance is perceived accurately, and is believed to have been perceived accurately, comprehenders may be led to assume a different intended message, which is more compatible with the signal that was in fact produced. However, such an account must reach into the realm of pragmatic theory. In Chapter 7, I argue that the Rational Speech Act model framework, which takes into account cost-based pressures on production, while concerning itself with pragmatic utterance interpretation, may reconcile the various traditions – while clearly making those empirical predictions which, in isolation, they may not make.

While the above discussion covers rationality from the point of view of the speaker, information-theoretic models of language processing propose that comprehenders also have strong expectations for how things will be said, and encounter processing difficulty (reflected by a variety of online measures) when these expectations are violated. In this tradition, what comprehenders specifically have expectations about is the *form* of utterances. After hearing something like "*John went to the store*," they do not expect for "*He paid the cashier!*" to follow, as it is redundant: they are therefore surprised by the *form* the discourse has taken. The Gricean tradition similarly suggests that comprehenders have a base expectation that speakers will behave rationally, and interpret utterances literally or non-literally in a manner that will, generally, help to match this expectation (Grice, 1975; Levinson, 2000). I explicitly propose also that comprehenders have expectations as to the *state of the world* conveyed by the speaker. In the above example, the state of the world is precisely the one that is expected (i.e., one in which *John* has paid the cashier). The two forms of predictability

- *form-based* and *meaning-based* - are often treated as essentially identical, but as I discuss in the following section, need to be disentangled to make accurate predictions about language processing.

Surprisal

An area where this work might have particular implications is in formal modeling of language processing. The mathematical concept of *surprisal* (Hale, 2001; Levy, 2008), traditionally, represents how (un)predictable a word or a string of words is in context. Specifically, it is the negative log of the probability of encountering a specific word or utterance. As hinted in the name, words or utterances that one might expect to see in a given context have *low* surprisal values, and those that one would *not* expect to see in a given context have *high* surprisal values. Smith & Levy (2013) show that difficulty in processing a word (or string of words), as reflected in online measures like reading times, is proportional to the word's unpredictability in context, or *surprisal*. In other words, comprehenders read or process words or utterances that are predictable (low *surprisal*) quickly, and those that are unpredictable (high *surprisal*) slowly. An utterance you don't expect to see in ordinary contexts (*John paid the cashier!*) should incur some processing difficulty for comprehenders. However, a problem with this account is that it treats all forms of *predictability* similarly. Consider, for example, two utterances that one might be (hypothetically) equally likely to hear: *John paid the cashier!* and *John punched the cashier!*. Processing theories which take into account only the predictability of an utterance would predict similar processing difficulty or processing times for both.

However, this is not only conceptually problematic, but would likely make the wrong predictions. Considering only *surface-level* or *form-based* predictability (the predictability of the string of words, in context) doesn't take into account the fact that the utterances are unpredictable for entirely different reasons: dispreference for redundancy, vs. event unpredictability. Further, the first (*cashier-paying*) utterance may contribute additional processing cost due to encountering pragmatic abnormality, or due to the need to make a pragmatic inference to resolve the apparent redundancy. In this case, despite identical surface-form predictability, one would expect that conceptually redundant utterances would be associated with greater processing difficulty. Of course, it may also be the case that conceptually redundant utterances are relatively easy to process, due to the relative ease of semantic integration (there are no unusual or unexpected facts to integrate into one's world model[11]). Either case, however, poses problems for the link between *surprisal* and processing difficulty, as utterances matched on predictability (and, consequently, *surprisal*) would still not end up with identical processing difficulty or reading times.

[11]Of course, the experiments here make clear that many comprehenders do end up integrating an unusual common ground belief (*John is a habitual non-payer*) when trying to resolve the apparent pragmatic violation. For those comprehenders, one would predict that the processing cost would, in fact, be greater than the cost of simply integrating an unusual event into one's world model.

Several other interesting implications remain for surprisal theory, or the claimed link between *surprisal*, and reading times, or processing cost. First, it is commonly assumed that processing difficulty, in the context of this theory, is caused by encountering a particularly unexpected *form*. However, in the redundant *cashier-paying* example, the *form* of the utterance is unexpected *precisely because* the predictability of the utterance meaning is so high. In other words, in order for comprehenders to have expectations about the *form* of the utterance, they must also have expectations about the global *meaning* conveyed by the utterance, as it is precisely the meaning that renders the form surprising to comprehenders. I therefore consider it a significant shortcoming of these theories that they frequently either do not consider the predictability of meaning (what can also be termed *conceptual* predictability), beyond the truth-conditional meaning of an utterance, or treat the two probabilities, that of *form* and *meaning*, as essentially identical – whereas I have argued that they not only can influence each other, but in fact can diverge systematically at their extreme values. In the following subsection, I talk about this relationship in more detail, as well as the implications it has for what *types* of language models could in principle address the issues I've outlined.

Formal Models of Language Processing

The predictability of informationally redundant utterances, as mentioned, should be fairly low at the *surface* level, and reading times have been argued to reflect the predictability of *surface-level* linguistic events, rather than the conceptual predictability of the scenarios they describe, i.e., their broader *meaning* (Smith & Levy, 2013). There is evidence, however, that comprehenders predict at multiple levels: for example, the *event* (in the current case, *meaning*) level, as well as at the level of *surface form* (Kuperberg & Jaeger, 2016), although it remains unclear how these levels interface (e.g., if comprehenders expect something predictable at the *event* level to go unmentioned at the *surface* level). In the case of surprisal theory (Hale, 2001; Levy, 2008), this may have interesting implications, given that the surprisal values that have been linked to processing times have largely been obtained using formal (computational) language models. If informationally redundant utterances result in longer reading times, it's unclear how formal models could accurately generate the high surprisal values one would expect for those utterances[12].

For example, in the case of the *conventionally habitual* event utterance, in the *ordinary* vs. *wonky* common ground, the event description (*John paid the cashier*) consists of exactly the same string of words, with the preceding context identical stretching over multiple preceding sentences. The utterance is informationally redundant in the *ordinary* context, and non-redundant in the *wonky* context. Simple

[12]For that matter, if they result in shorter reading times, as speculated in the previous section, there would similarly be a problem given that the processing difficulty should not rely simply on *surface-level* probability, but also on *event* or *meaning* probability, which, as explained, current models cannot adequately integrate.

or even complex n-gram models, which can't represent long-distance dependencies, would not show any difference in predictability, and therefore would predict no differences in processing difficulty. Relatively sophisticated models which incorporate syntax or semantics, similarly, would not predict a difference, as there are no meaningful differences in syntactic structure, and semantic models would not have access to the relevant event-based information which distinguishes the utterances.

Models of event sequences, which estimate *event* (vs. string) probability, may be able to estimate differences in predictability, and, consequently, processing difficulty, between utterances describing script-congruent and script-incongruent events (e.g., events likely and unlikely to be a part of *grocery shopping*). However, the general prediction such models would make is that the more congruent an event is with an invoked script (i.e., the more predictable the event is given the script), the more predictable (and easy to process) the utterances which describe that event should be. There is no principled way, within this framework, to divide activities up into different 'grades' of predictability, such that utterances describing *moderately habitual* activities are easier to process than those describing *not-so-habitual* activities, yet those describing *very habitual* activities incur difficulty. In light of this, I suggest that to predict any processing difference between informationally redundant and non-redundant utterances, formal models of language comprehension would need to incorporate some form of pragmatic reasoning, and in Chapter 8, I discuss what such a model may look like (although the Rational Speech Act model does is not yet equipped to make any predictions about processing times).

Although attempts to build formal or computational language models may appear to have limited relevance to how humans process language – which is typically thought of as a seamless integration of information from the surrounding context – it should be recognized that humans do not make predictions about upcoming material based simply on the preceding string of words, as formally assumed by simple models of language processing and prediction. The vast majority of word/utterance sequences have never been previously encountered by a comprehender, and predictions concerning upcoming material cannot be based on them alone. Regardless of the modeling approach one takes, it must be concluded that humans also make predictions by keeping track of certain cues - semantic, syntactic, lexical, and pragmatic (e.g., whether a speaker is generally adhering to conversational norms). Thus, determining the specific cues that are necessary to accurately model language processing is also relevant to understanding how humans accomplish the same task, and what information they must keep track of in order to do so. There are two ways of elucidating which linguistic and contextual cues influence language comprehension: one may manipulate relevant cues in tightly controlled stimuli, and observe their influence on interpretation, or online measures such as reading times; or one may build formal language models which make specific, testable predictions regarding the influences of these cues on processing and comprehension. I believe that a combination of the two is likely to be the most fruitful approach.

To summarize, I argue that it would be informative to investigate the processing of informationally redundant utterances, using online measures such as eye-tracking of self-paced reading. On the one hand, there are many claims, but still relatively little data on the online processing of pragmatic inferences, and little is known about the cost (or efficiency) of pragmatic reasoning (Degen & Tanenhaus, 2016). The data that does exist is for the most part limited to scalar implicatures, which are often argued to be computed relatively automatically (but see Huang & Snedeker, 2009). On the other hand, determining whether informational redundancy contributes to the processing cost of utterances, above and beyond the surface predictability of those utterances, is critical to determining whether formal language models need to integrate pragmatic reasoning to correctly predict processing cost. The main challenge to using online measures is that to compare (for example) the reading times of utterances, they must be matched on all factors which may affect reading times, but are irrelevant to the experimental manipulation (in a case such as ours: length, word frequencies, etc.). This makes it very difficult to compare reading times for utterances that are not identical in their surface form. In the current case, one possibility is to compare reading times for otherwise identical phrases that are informationally redundant in one context, but not the other, as with the *cashier-paying* examples in the *ordinary* and *wonky* common grounds. I leave this to future work.

4.6.2 Perspectives for future work and conclusion

There are several avenues for further research. First, the range of inferences that comprehenders might draw from informationally redundant utterances may extend well beyond what was tested in this series of experiments. For instance, particularly in the absence of a possible pragmatically felicitous interpretation, like the one suggested by the response measure, comprehenders may simply assume that a speaker is being uncooperative, is having some production difficulty, or has unconventional speaking patterns (cf. Grodner & Sedivy, 2011; Pogue et al., 2016). There is also the possibility that informationally redundant event descriptions, especially as seen in Experiment 3, are initially interpreted as likely, and possibly aborted, temporal or causal anchors for more 'interesting' information. For example, in the context of a *grocery trip*, an 'informationally redundant description such as *John paid the cashier*, when followed by *with euros instead of dollars*, would likely not be considered anomalous. In this case, the description would not be redundant in its broader context, as it's part of a more extended description which overall contributes previously unknown, or not easily inferable information. These hypotheses might be investigated using rating studies, sentence or passage completion studies, or more naturalistic tasks where participants' behavior provides a clue as to their interpretation of these utterances.

Overall, the results described here strongly suggest that comprehenders perceive informational redundancy as anomalous. Comprehenders are able to accommodate the provision of this 'unnecessary' information by altering their pre-utterance beliefs

about individuals' behavior, or, more broadly, the common ground between speaker and comprehender. This demonstrates that there is a limit on how redundant a speaker may be without altering the interpreted meaning of the transmitted message, which is not accounted for by the base UID hypothesis. The results also complement work in the dialogue literature (Walker, 1993), which illustrates that informationally redundant utterances are frequently used to convey 'informationally useful' non-literal content. They raise presently important questions regarding which cues are systematically tracked by comprehenders, as well as how those cues are integrated during pragmatic interpretation – a question that is addressed in more detail in Chapter 8. Finally, they address the pragmatic interpretation of complex utterances, not bound to specific classes of lexical items, which to date have been largely treated as either too complex, or too idiosyncratic, to study systematically.

Chapter 5

Referring Expression Choice: Background

A speaker's choice of how to refer to an entity in discourse is one of the more well-studied phenomena of utterance choice. Speakers commonly have a variety of expressions available for this task – take, for example, a simple situation where the speaker wishes to express that an individual is located up the stairs. There are multiple ways in which they may refer to this individual in English, including pronouns, proper names, and definite descriptions:

(1) **He** is upstairs.

(2) **John** is upstairs.

(3) **The man** is upstairs.

All are grammatically licensed in most contexts, and although not all may be equally felicitous, typically the speaker may freely choose between a pronoun, and either a name or a definite description, and expect to be understood with minimal contextual disambiguation. In this chapter, I present the background and motivation for a series of referring expression production experiments, which aim to address the broader question of whether and how message predictability affects utterance choice at the level of discourse, as it typically does at other levels of production (for a review, see Section 2.1.2).

In the case of the utterances above, the question of interest then is what exactly determines a speaker's choice of expression. One might imagine many different factors which influence how felicitous any given expression is, including: whether the interlocutors both know *John*; how relevant *John* as an individual is to the rest of the discourse; and whether *John* has been previously mentioned in this discourse (particularly by name); all of which make mention by pronoun or name more likely. What ties many of these disparate influences together is that all are correlated with

the predictability of *John's* mention in the discourse. Perhaps, then, what matters primarily for choice of referring expression is the degree to which the speaker and comprehender *expect* for *John* to be mentioned, at that point – in other words, the intended referent's *predictability* in context.

That predictability should predict referring expression choice is intuitive – after all, if *John's* mention is perfectly predictable, one may be able to afford using a relatively simple, ambiguous, and easily misperceived expression, such as *he*, and still expect that the comprehender will walk away having understood the intended message. This hypothesis, in various forms, has been proposed by a variety of authors, either under the guise of factors such as *salience* which are minimally highly correlated with (if not identical to) predictability, or explicitly under the guise of *predictability* or *expectancy*. At face value, this hypothesis is highly appealing: if predictable references are more easily *recoverable* in context, then rational and efficient speakers should choose the shorter and less effortful expression whenever they can get away with doing so.

However, unlike with most of the other phenomena studied to date, where speakers choose between variably effortful linguistic alternatives, the degree to which referent predictability influences referring expression choice is under significant debate (Arnold, 2001; Rosa & Arnold, 2017; Bott et al., 2018, p.c. July 25, 2018; Fukumura & van Gompel, 2010; Rohde & Kehler, 2014; Tily & Piantadosi, 2009; Kravtchenko, 2014; Resnik, 1996; Modi et al., 2017). As discussed in Section 2.1.2, there is theoretically a critical distinction between making a choice among meaning-equivalent alternatives (such as *math*/*mathematics*), and making a choice among utterances which communicate substantially different amounts of information at the semantic or truth-conditional level.

The case of referring expressions, therefore, presents another challenging test case for the UID hypothesis. Linguistic alternatives at the discourse level, as a rule, are not truth-conditionally equivalent in the way that alternatives at other levels of production are. The potential failure of UID to play out straightforwardly in cases such as that of referring expressions would indicate that UID, contrary to what has been posited (Jaeger, 2010), does *not* have online effects at every level of production. However, it is also an empirical question as to what extent of semantic meaning-equivalence may be necessary for these effects to emerge. Looking back at informationally redundant utterances, while the two utterances *John went shopping. He paid the cashier* and *John went shopping* are not strictly meaning-equivalent, one may conclude that the latter will almost always be chosen over the former, in a typical scenario, given the predictability of *cashier-paying* in this context. In fact, the choice of the former in place of the latter is sufficiently unexpected by listeners that it triggers pragmatic inferences which substantially revises their ideas of the common ground.

Testing which factors affect the choice of whether and how to express intended discourse-level meaning, whether using corpora or natural language production experiments, is typically difficult and costly, partially due to the relatively uncontrolled variation in how a given meaning may be expressed (or not), to difficulty in de-

termining whether a particular meaning is *intended*, and to the logistics of setting up scenarios which will reliably provoke a speaker to *intend* to convey a particular meaning. However, the use and variety of referring expressions are highly regularized, referring expressions themselves are ubiquitous, and the meaning intended to be *conveyed* by them is nearly always recoverable (Brown-Schmidt & Tanenhaus, 2008). The *predictability* of certain referents, as will be discussed in Section 5.2.1, is also very easy to manipulate in a controlled fashion. Referring expressions therefore are highly suited to answering the question of whether the UID hypothesis holds at the discourse level.

Currently, there are a number of studies supporting the hypothesis that predictability affects referring expression choice (Tily & Piantadosi, 2009; Kravtchenko, 2014; Resnik, 1996; Arnold, 2001; Rosa & Arnold, 2017; Bott et al., 2018, p.c. July 25, 2018). On the other hand, there are at least an equal number of studies which show no effect of predictability on referring expression production. The factors which affect referring expression choice correlate with predictability, which suggests that minimally, referent predictability has a role in conventionalized or overall patterns of referring expression use (Arnold, 2008). However, every experimental paradigm used so far to test this hypothesis online produces severely conflicting results over whether there is a *local* effect of predictability on referring expression choice.

The use of different types of corpora, different stimulus design, different experimental paradigms, differing degrees of paradigm naturalness and interactivity, and different cues to manipulate predictability, has confounded attempts to trace the source of those effects that are detected. What does appear to be clear, however, is that to the degree that there *is* an online effect of predictability on referring expression production, this effect is relatively weak, possibly specific to certain contexts and types of discourse, and may emerge only in scenarios that call for more audience design on the part of the speaker. The experiments I describe in Chapter 6 are intended to bridge several gaps in the previous literature and previous experimental designs. I take one of the most commonly used and easily controlled paradigms, and use several manipulations that, on the basis of theory and prior literature, should maximize the likelihood of detecting an effect within this paradigm.

Throughout the rest of this chapter, I will briefly summarize the phenomenon of referring expression use as it pertains to this question. I will then present a more thorough literature review of the accounts of referring expression use that have been posited to date.

5.1 Literature Review

As discussed in Chapter 2, the main claim that the UID hypothesis makes is that humans aim to communicate efficiently by trading off between conciseness and message recoverability. In the context of referring expressions, this means that given

a more predictable meaning, or predictable *referent*, shorter expressions should be preferred. Foremost, this means that pronouns should be preferred to names and definite descriptions when the referent is more predictable. Although there has been a similar claim (Tily & Piantadosi, 2009) that names should be preferred to definite descriptions (which are typically longer) when the referent is more predictable, the two are relatively rarely licensed in the case contexts. It would be highly irregular, for example, to refer to someone already known by name to both interlocutors as *the woman* or *the man*, no matter how unpredictable the referent is in context – so such a trend could typically only be observed on average, rather than within a controlled experiment.

In this section, I first discuss the various theories of referring expression use that have been proposed to date, pointing out where they intersect in various ways with the claim that more predictable referents should be referred to with shorter expressions. I will then cover specific theoretical frameworks which link referent predictability with referring expression choice, as well as any empirical attempts at testing those theories. In the final section following the literature review, I propose a series of experiments, which attempt to bridge some of the gaps in previous work producing contradictory results.

5.1.1 Referring Expression Production

Multiple accounts have been proposed of which factors govern the use of third-person personal pronouns, proper names, definite descriptions, and other types of referring expressions (such as null anaphora). Commonly it's claimed that referents must be particularly prominent, salient, accessible, or 'activated,' in order to license the felicitous use of a personal pronoun over a proper name or definite description. A selection of the literature will be covered here.

The Role of Topicality

Some of the first attempts at capturing what licenses the use of pronouns appeal to the concept of *topicality*, of the referent or antecedent. Topicality, in general, has been used to account for a wide range of linguistic phenomena, including the licensing of specific intonation (Halliday, 1967), word orders (Sgall et al., 1986), or referring forms (Ariel, 1990; Gernsbacher, 1990). The main argument is that topical referents or antecedents license the use of pronouns, or other short or simple referring expressions.

However, attempting to define what makes something *topical* leads to one of the main problems with this account, which is that the term is typically vague and can be variably defined. Commonly, it's defined simply as that which a sentence or discourse is *about* – but at the level of discourse particularly, *aboutness* is rather difficult to quantify or determine. In general, topicality can be viewed as either a binary property

(either something is a topic, or it isn't), or as a gradient property (e.g., Givón, 1983). In the binary view, relevant to the discussion of experimental work in Section 5.2.1, typically the sentence subject, or sentence-initial element, is seen as the *topic* (in topic-marked languages, the topic is trivially determined). However, this runs into the immediate problem of grammatical subjects not necessarily appearing sentence-initially, and otherwise non-subject sentence-initial elements often arguably being that which the sentence is *about*.

Givón (1983) argued for topicality as a *continuum* – where all elements in a sentence are topical to some degree, but some are more topical than others. Those elements which are more topical are more likely to be pronominalized, and to license subsequent pronominalization. Topicality can be determined by how recently the entity was last mentioned, how many other referents appeared in the interim, and how long the referent remains in the discourse. Topics, on the one hand, reflect the status of the referent in the discourse so far. On the other hand, they also reflect the status that the referent is intended to have in the subsequent discourse.

According to Givón, topicality correlates with both current and future choice of referring expression. As conceptualized by him, the notion of topic has much to do with the notion of *salience* or *accessibility*, which is covered in the following sections.

The Accessibility Hierarchy

Unlike topichood, the notion of *accessibility* explicitly concerns the cognitive status of the referent. Ariel (1990) suggests that referents can be described in terms of their *accessibility* to the addressee, and the referring form chosen by the speaker depends on their understanding of the referent's cognitive status to the *comprehender*. There are several factors which affect how *accessible* any given referent is: the distance to the antecedent; the number of other referents in discourse that could in principle be referred to; the saliency (roughly, topichood, and for Ariel, subjecthood) of the antecedent; and whether the antecedent shares roughly the same world, time frame, paragraph, or other feature of the current discourse.

The accessibility of referents then correlates with the use of certain referring expressions, as proposed by (Ariel, 1990), with a wide range of referring expression types ranging from 'low-accessibility' full names with modifiers ('Joan Smith, the president') to 'high-accessibility' referring expressions like null pronouns. Similar scales, with entities higher on the scale being more likely to license the use of pronouns, have been proposed by Chafe (1994) (given > accessible > new) and Gundel et al. (1993) (in focus > activated > familiar > uniquely identifiable > referential > type identifiable). Prince (1992) proposes that the givenness scale be split into the following categories: new, inferable, and old. Although I will not discuss the specific criteria of each scale category here, I will reference this idea again in Section 5.1.2. These scales are similar in essence and use to the scale proposed by Ariel.

Psycholinguistic Approaches

Outside the formal linguistics and computational literature, there are a variety of psycholinguistic approaches to reference production and interpretation, the results of some of which will be described here. For the most part, these approaches either implicitly or explicitly assume that similar influences play out in the production and interpretation of referring expressions, and that any patterns observed in interpretation must also generalize to production, and vice versa. A challenge to this assumption will be discussed in Section 5.2.1.

A primary pattern that has been observed in referring expression interpretation is that ambiguous pronouns are typically interpreted as coreferential with preceding subjects (Stevenson et al., 1994), although this may be tempered by biases towards re-mention of entities in particular thematic roles. Another pattern that has been observed is a tendency towards parallelism, or for entities in various grammatical positions to be associated with previous entities occupying the same grammatical positions (Sheldon, 1974, Corbett & Chang (1983); Grober et al., 1978; Springston, 1975) – although this work mostly only demonstrates the tendency for subjects to refer back to subjects. Increased ambiguity may decrease the likelihood of pronoun use, and affect the ease of reference resolution, when there are multiple candidate entities that a pronoun could refer to (Ariel, 1990, Givón (1983), McDonald & Macwhinney (1995)).

Overall, this paints a complicated and multifaceted picture of which (interacting) factors affect referring expression use. I will next look at more overarching theories which attempt to make specific predictions about referring expression use or interpretation in particular contexts, and given an intersection of some of the above factors.

Centering Theory

Centering theory is one attempt to make specific, quantifiable predictions about the use of pronouns or other entities in the context of a particular discourse. Here, I will provide a brief summary. The primary claim of Grosz et al. (1995) is that some entities in an utterance are more *central* than others, and that this quality determines which referring expressions a speaker is likely to use for them.

Any given sentence offers the opportunity of referring to multiple entities, of *forward-looking centers* (C_f). Entities are ordered in terms of prominence, based on their grammatical roles: subject > direct/indirect object > adjunct NP. If any entities are referred to within the sentence, one of the expressions may refer back to the topic of a preceding utterance: C_b, or the *backward-looking center*. The rules of referring expression choice, roughly, operate as follows:

1. If any of the entities in the current utterance (C_f) are pronominalized, the *backward-looking center* (C_b), or that referent which is the topic of the preceding

utterance, must be one of them.

2. Topic continuations are preferred to referring back to the topic of an utterance preceding the immediately preceding one, which are preferred to topic shifts.

While this is a much more formalized system than the topicality- or accessibility-based accounts, the core insight of the first rule remains the same: pronominalization should be reserved for the most prominent entities of a discourse (if it is used at all). The second rule suggests, minimally, that more prominent entities are more likely to be mentioned again, and are more likely to remain prominent entities. These rules both aim to predict pronoun use, and to account for pronoun interpretation.

5.1.2 Predictability-Based Accounts

One of the problems with the accounts discussed in the preceding section, aside from the often vague definition of terms such as *topicality* or *salience*, is that the heuristics which account for pronoun production, while relatively intuitive and cognitively plausible, essentially remain just-so stories. The idea that one might prefer to use a shorter but more ambiguous or confusable expression for a more topical, salient or accessible referent is highly appealing, and empirically supported. However, the core idea of why expressions may be more recoverable or felicitous in certain contexts is often lacking, or simply refers back to observable facts. The role of predictability in assisting message recovery in a noisy channel, and in prompting the efficient use of maximally concise expressions, however, provides a fairly straightforward account.

Kuno (1972) and Prince (1981) first equated predictability with givenness, and Givón (1988) similarly equated it with accessibility. Later researchers observed the established effects of predictability on acoustic reduction, and proposed that the same mechanism might influence reduction of referring expressions (Arnold, 2008; Givón, 1988, 1989). Fowler et al. (1997) noted that highly predictable referents tend to be associated with both acoustically and lexically reduced referring forms. Further, factors which have been observed to correlate with predictability and acoustic reduction, such as recency or frequency of mention (Bard & Aylett, 1999; Fowler & Housum, 1987; Fowler et al., 1997) have also been observed to influence referring expression choice.

Arnold (2008) introduces the *Expectancy Hypothesis*, which proposes that referent *predictability* is the primary influence on referring expression choice, as well as interpretation. She argues that the various factors which so far have been argued to influence referring expression choice – distance to last mention, parallelism, subjecthood of last mention, and so forth – all correlate with the *predictability* of a referent in context. *Predictability*, unlike topichood, givenness, and accessibility, has the advantage of being measurable and formally defined. Further, the mechanism which it plays in production and interpretation is well-described and motivated by basic communicative principles.

The *Expectancy Hypothesis* in essence makes the same predictions as the UID hypothesis, but limits itself to the use of referring expressions. Arnold (2001), Arnold (2008) lays out more clearly the factors which influence referent predictability, and, consequently, referring expression choice. At face value, this hypothesis is extremely attractive. It neatly accounts for the existence of those patterns of referring expression choice that have been observed, with no appeal to otherwise unmotivated cognitive principles. It takes its inspiration from empirical data across multiple linguistic domains, showing that reduced forms are more likely to be used for predictable meanings. One question this hypothesis leaves is whether there is evidence that conventionalized patterns of referring expression choice stem from principles of (on average) efficient communication. The other question, central to the following chapter, is whether the effects of predictability on referring expression choice can be detected in online production.

5.2 The Empirical Evidence

While there is evidence that referent predictability, or next-mention bias, affects comprehension or processing (Rohde & Kehler, 2014; Stevenson et al., 1994), this does not necessarily extend to referring expression *production* (for more on the potential asymmetry, see Section 9.1). In this section, I will focus specifically on empirical evidence for the effect of predictability on referring expression production – specifically, either pronominalization or subject omission. First, I briefly discuss the experimental paradigms that have been used to date.

Two experimental paradigms, with some variation in setup, have predominantly been used to investigate whether predictability affects referring expression choice. The first is written passage completion, which allows for tight control of experimental prompts, typically at the expense of naturalness. The other is based on corpus analysis, where coreference-annotated corpora are further annotated for referent predictability through use of Shannon game-like guessing games, as described below. In both cases, the focus is on whether, after controlling for possible influences on referring expression production, referent predictability continues to exert a significant influence on referring expression choice.

5.2.1 Evidence from the Passage Completion Paradigm

The most commonly used paradigm is written passage completion. In this paradigm, participants are given a set of written prompts – typically, single sentences with multiple referents. A single variable, typically the verb, is manipulated so that either one referent or the other (typically, the subject of the object) is more likely to be mentioned next. The empirical likelihood of mention, given the manipulation, is typically assessed in the same study. All other factors are typically held constant.

For example, *causality* verbs bias participants towards subsequently talking about the *cause* for the described action (Stevenson et al., 1994). Since causality verbs, as discussed in Section 6.2 can be biased either towards continuations about the subject, or about the object, it's possible to test whether the more likely continuation, or *more predictable referent* is more likely to be pronominalized in written passage completion, relative to the less predictable referent. In Section 6.5, I offer up a critique of this paradigm.

Evidence from this paradigm is highly mixed. Arnold (2001) used this paradigm with transfer-of-possession verbs, which are biased towards goal completions. Since the goal can be in either subject (X caught Y from Z) or object (X threw Y to Z) position, it's possible to independently look at the effect of antecedent grammatical position on referring expression choice, and the effect of semantic continuation bias (referent predictability) on referring expression choice:

(4) I hate getting sick. It always seems like everyone gets sick as soon as it's vacation. Marguerite caught a cold from Eduardo two days before Christmas. _____

(5) There was so much food for Thanksgiving, we didn't even eat half of it. Everyone got to take some food home. Lisa gave the leftover pie to Brendan. _____

Participants were free to complete the passages as they wished. In the case of 4, since _ Marguerite_ is the goal and more likely continuation, references to her should be more likely to be pronominalized than those to *Eduardo*. In the case of 5, since *Brendan* is the more predictable referent, references to him should be more likely to be pronominalized than those to *Lisa*. At minimum, more predictable subjects should be pronominalized more often than less predictable subjects, with the same for objects. This is, in fact, what Arnold (2001) found.

However, Fukumura & van Gompel (2010) point out that these results may not be conclusive. First, they note that the effect of predictability on referring expression choice was stronger in the case of the goal object than the goal subject, although the goal subject was more predictable than the goal object. Second, and more critically, they note that there were many factors that differed between the two conditions, aside from the manipulation of interest. The context preceding the prompts was different, the object transferred was different, and the final segment was different (an adjunct in 4, and the goal in 5). In other words, factors other than predictability may have influenced the choice of referring expressions in both conditions.

Fukumura & van Gompel (2010) employ a similar paradigm to Arnold (2001), with some alterations. First, instead of Transfer-of-Possession verbs, they use implicit causality verbs, which may be either subject- or object-biased, but do not require one to control, for instance, for a transferred object. They omit the preceding context, add some material at the end of the prompt which is identical between stimulus

pairs, and ensure that the only factor manipulated between subject- and object-biased constructions is the verb:

(6) Gary scared Anna after the long discussion ended in a row. This was because _____

(7) Gary feared Anna after the long discussion ended in a row. This was because _____

What may be critical, as will be discussed in Section 5.2.4, is that rather than letting participants complete the passage as they wish, they indicated to participants which referent they should refer to next. This ensured that there were sufficient instances of the less predictable referent mention to make robust conclusions about pronominalization likelihood. What they found is that predictability had no influence on pronominalization

Rohde & Kehler (2014) point out that the results in Fukumura & van Gompel (2010) could be confounded by the fact that they used opposite-gender entities in their prompts. If entities are of different genders, then the pronouns used to refer to them are unambiguous as to their reference. If part of the reason for reducing mentions of predictable referents is their recoverability, then unambiguous pronouns give speakers no incentive for nominalizing less predictable referents. As I further show in Section 9.2, the use of opposite-gender entities in prompts would indeed to predicted to reduce or remove any effect of predictability on pronominalization.

Rohde & Kehler (2014), like Fukumura & van Gompel (2010), use implicit causality verbs. Unlike Fukumura & van Gompel, they use same-gender entities in the prompts, omit the additional phrase at the end of the prompt, and use a free-completion task rather than a constrained-choice task. The primary stimuli (from Experiment 1) look like the following:

(8) [Subject-biased IC verb, no-pronoun] John infuriated Bill. _____

(9) [Object-biased IC verb, no-pronoun] John scolded Bill. _____

(10) [Non-IC verb, no-pronoun] John chatted with Bill. _____

(11) [Subject-biased IC verb, pronoun] John infuriated Bill. He _____

(12) [Object-biased IC verb, pronoun] John scolded Bill. He _____

(13) [Non-IC verb, pronoun] John chatted with Bill. He _____

Like Fukumura & van Gompel (2010), they find no effect of referent predictability on pronominalization. Rohde & Kehler observe a mismatch between production, on which predictability appears to have no effect, and interpretation, where predictability clearly biases comprehenders towards certain interpretations. They account for this by positing that while pronominalization, from the speaker's point of view, is solely dependent on grammatical factors, interpretation is also guided by the listener's expectation of who will be mentioned. They reconcile this asymmetry by proposing a

Bayesian approach towards pronoun interpretation, which is discussed separately in Section 5.2.3.

Most of the evidence so far suggests that predictability plays no role in referring expression choice. However, Rosa & Arnold (2017) propose that the critical difference between those studies that find an effect of predictability, and those that don't, is Arnold's use of transfer-of-possession verbs, and the other studies' use of implicit causality verbs. Further, they posit that the use of a written passage completion paradigm, with tightly controlled stimuli devoid of any discourse context, divorces the task from natural discourse production. If reduction of more predictable elements is a matter of audience design, then more unnatural and less interactive tasks would be expected to decrease this effect[1]. To account for this, they run 3 experiments, using transfer-of-possession stimuli, which vary the degree of interactivity, and the degree to which the stimuli are part of a coherent narrative.

Experiment 1 of Rosa & Arnold (2017) is a highly interactive version of the passage completion paradigm. Participants were presented two pictures, first of a possession-transferring event between two entities, and then a second picture of one of the entities engaging in another activity. They spoke with a confederate who played the role of a 'detective' in a murder mystery. To simulate natural discourse, the series of pictures were designed to illustrate a coherent narrative, with repeated use of the same entities. The confederate read out either a subject-biased (goal-source) or an object-biased (source-goal) description of the activity in the first picture, and the participant was then prompted to describe the second picture, which featured either the goal or the source entity. Rosa & Arnold found that participants were far more likely to pronominalize subjects or non-subjects when they were the more predictable continuation.

In Experiment 2, they used the same stimuli, which still formed a coherent narrative with repeated use of the same entities, but embedded them within the standard written passage completion paradigm, using a constrained-completion rather than a free-completion paradigm. Participants again reliably used more pronouns for more predictable referents, when controlling for the grammatical function of the antecedent. In Experiment 3, the authors again chose a written completion task, but in this case the stimuli did not form a coherent narrative, and did not repeatedly mention the same participants. In this case, an effect of predictability on production was found only for same-gender prompts, but not for different-gender prompts. This asymmetry between results from same-gender and opposite-gender prompts is not adequately explained by the authors, but as I demonstrate in Section 9.2, it would in fact be expected if speakers are choosing referring expressions based on expected communicative utility.

In short, minimally it appears that effects of predictability on referring expression choice are detectable within a passage completion paradigm, so long as transfer-of-

[1]In fact, there is evidence for this, as I discuss in Section 6.5.

possession verbs are used, rather than implicit causality verbs. The weakness of this argument is that there is no clear *a priori* reason for why predictability should affect referring expression choice in one case, but not the other. Further, it also appears that the effect is far stronger within a fully interactive setting, and decreases in magnitude as the setting becomes less interactive, and the discourse context less rich and coherent. Overall, the results from Rosa & Arnold (2017) are compelling, but difficult to reconcile with previous work.

Aside from a lack of independently motivated explanation for the disparity with previous results, Rosa & Arnold (2017) has several shortcomings, which I attempt to address in the experiments described in Chapter 6. First, they use a severely limited set of Transfer-of-Possession verbs to test their hypothesis, leaving it unclear whether the effect generalizes to all similar verbs. Second, the experiments may individually be underpowered due to highly stringent participant exclusion criteria. Nevertheless, given the number of experiments run, and the strength of the effect in the first two experiments, the evidence that predictability *at least sometimes* has an effect on referring expression production is difficult to argue with. In Section 5.2.4, I introduce an account that attempts to reconcile these results.

5.2.2 Evidence from the Corpus Guessing Game Paradigm

Another paradigm which has been used is that of using referent guessing games to annotate coreference-annotated corpora for predictability, and seeing whether predictability contributes to referring expression choice after accounting for the influence of other known factors. In this case, researchers present portions of the corpus to participants, in the form of a guessing game. Participants are initially shown the text up to (but not including) the mention of the first referent, and are then asked to guess the identity of the upcoming referent (either 'something new,' or one of the referents already mentioned). After the participant makes their guess, they are shown the referent mention, as well as the following discourse, up until the next referent mention, which they must again guess the identity of. Average participant accuracy in guessing the identity of concealed referents, based on prior discourse context, is used as a proxy for predictability.

In this manner, a corpus can be annotated for referent predictability. When this predictability is negative log-transformed, it can be used as a measure of the *information* carried by the referring expression. At this point, a binomial or multinomial regression analysis is run to see if, after factoring in other common predictors of referring expression choice (distance to last mention, grammatical function of last mention, etc.), *predictability* or *information* continues to have a significant influence on referring expression choice. In this case, one expects that minimally, the shortest referring expressions are preferentially used for more predictable (lower information) referents, and longer referring expressions are used for less predictable (higher information) referents.

(I)$^{(1)}_{\text{P\_bather}}$ [decided]$_{\text{E\_wash}}$ to take a (bath)$^{(2)}_{\text{P\_bath}}$ yesterday afternoon after working out . Once (I)$^{(1)}_{\text{P\_bather}}$ got back home , (I)$^{(1)}_{\text{P\_bather}}$ [walked]$_{\text{E\_enter\_bathroom}}$ to (my)$^{(1)}_{\text{P\_bather}}$ (bathroom)$^{(3)}_{\text{P\_bathroom}}$ and first quickly scrubbed the (bathroom tub)$^{(4)}_{\text{P\_bathtub}}$ by [turning on]$_{\text{E\_turn\_water\_on}}$ the (water)$^{(5)}_{\text{P\_water}}$ and rinsing (it)$^{(4)}_{\text{P\_bathtub}}$ clean with a rag . After (I)$^{(1)}_{\text{P\_bather}}$ finished , (I)$^{(1)}_{\text{P\_bather}}$ [plugged]$_{\text{E\_close\_drain}}$ the (tub)$^{(4)}_{\text{P\_bathtub}}$ and began [filling]$_{\text{E\_fill\_water}}$ (it)$^{(4)}_{\text{P\_bathtub}}$ with warm (water)$^{(5)}_{\text{P\_water}}$ set at about 98 (degrees)$^{(6)}_{\text{P\_temperature}}$.

Figure 5.1: Example story from Modi et al. (2017).

Tily & Piantadosi (2009) use the corpus guessing game paradigm to test the prediction that pronouns would be used for more predictable referents than nouns, and nouns for more predictable referents than definite description. A coreference-annotated corpus of Wall Street Journal articles (Weischedel et al., 2008) was annotated for referent predictability using average guess accuracy from crowdsourced study participants who read passages from the corpus. Tily & Piantadosi find that pronouns are more likely to be used for predictable referents than definite descriptions, but not more likely than proper nouns, after controlling for other predictors of referring expression choice.

Kravtchenko (2014) used the same paradigm to investigate whether optionally null subjects in Russian were used for more predictable referents than overt subjects. Unlike Tily & Piantadosi (2009), I created and coreference-annotated a corpus that is more representative of natural, everyday discourse. Participants each read 2 of 24 passages, 8 each from interviews, personal blog posts, or dialogue in plays. After controlling for other predictors of referring expression choice, a significant effect of predictability on subject omission remained.

Modi et al. (2017) used a corpus of crowdsourced descriptions of everyday activities, and annotated it for predictability using crowdsourcing, as in the above studies. This corpus was significantly larger, with 3346 referring expressions and 182 stories (vs. 82 texts with 2211 referring expressions for Tily & Piantadosi (2009). After controlling for the same predictors of referring expression choice, they found no effect of predictability. However, there are several shortcomings of this study. The text in the corpus passages was fairly unnatural, as MTurk workers were asked to describe in detail, as if to a child, the steps in certain stereotyped event sequences. In practice, such detailed descriptions tend to be maximally redundant, avoiding the reduction of referring expression even when licensed to do so. Further, many of the entities in the passages were either first-person (with only one candidate referring expression), or non-animate, which are less likely to be reduced (Fukumura & van Gompel, 2011; Dahl & Fraurud, 1996; Yamamoto, 1999). An example of a passage from this corpus can be seen in Figure 5.1.

In short, it does not appear that the question of whether predictability affects referring expression choice has been settled in this paradigm. A separate criticism of these studies is that those control predictors that are factored out – such as the grammatical position of the last mention, distance to last mention, the number of times something has been mention – are in large part precisely those cues that comprehen-

ders use to predict which referent will be mentioned next. Arguably, these studies are partially factoring out the effects of predictability in this manner, although it makes a stronger test case if predictability separately continues to have a significant effect on production. In this study data, for example, the effect of predictability is stronger by orders of magnitude, prior to controlling for cues that are themselves correlated with predictability.

One possible way forward, which I do not however otherwise address in this thesis, is to run an experiment roughly the size of the Modi et al. (2017) experiment, but using relatively natural and informal written material, where reference to third-person animate entities is plentiful. I next address an attempt to reconcile the strong empirical evidence for an effect of predictability on reference processing, with the at best equivocal evidence for an effect of predictability on referring expression production.

5.2.3 A Bayesian Approach to Referring Expression Production and Interpretation

An issue which has not yet been discussed in detail remains: if predictability has a clear effect on reference resolution and processing, then why does it appear that it may not have an effect on production? Kehler et al. (2008) and Rohde & Kehler (2014) provide an elegant and highly plausible solution to this issue, by way of a Bayesian approach to pronoun interpretation. As Stevenson et al. (1994) first suggested, they argue that pronoun interpretation is influenced both by the grammatical function of the antecedent, and the semantic (continuation) bias of the verb – or, the likelihood that a given referent will be mentioned.

Under Rohde & Kehler account, pronoun production is governed by only one factor: the grammatical function of the referent's antecedent. That is, if the referent's last mention is the preceding subject, then the speaker will tend to use a pronoun, and if it is the preceding non-subject, the speaker will tend to use a proper name (or noun). However, on the comprehender's end, comprehension proceeds in a Bayesian manner, with the comprehender taking into account both the likelihood of referent mention (given verb semantic biases), and the likelihood of pronominalization (given the grammatical role of the potential referent's last mention):

$$P(\text{referent} \mid \text{pronoun}) = \frac{P(\text{pronoun} \mid \text{referent})P(\text{referent})}{\sum_{\text{referent} \in \text{referents}} P(\text{pronoun} \mid \text{referent})P(\text{referent})} \tag{5.1}$$

In this case, $P(\text{referent} \mid \text{pronoun})$ is the interpretation bias: the likelihood that a given referent is being referred to, given the pronoun that was used: determining this is the comprehender's task. I then use Bayes' rule to calculate this, given $P(\text{pronoun} \mid \text{referent})$ – which is the likelihood of using a pronoun for a given referent, and something Rohde & Kehler argue is simply a grammatical bias – and $P(\text{referent})$,

which is the likelihood of a given referent being mentioned in the first place. Critically, Rohde & Kehler's assertion that the production bias is determined solely by grammar is directly at odds with the Expectancy Hypothesis or the UID hypothesis, which minimally predicts that the likelihood of pronominalization is proportional to the likelihood of referent mention:

$$P(\text{pronoun} \mid \text{referent}) \propto P(\text{referent}) \tag{5.2}$$

This account necessitates that production and interpretation are necessarily asymmetrical, and influenced by referent predictability and grammatical biases to different degrees. However, it must be noted that there is nothing about this approach that likewise *necessitates* that only grammatical biases influence referring expression production – this is assumed solely on the basis of the empirical data. Indeed, *a priori* it is at least equally plausible that production is similarly probabilistically determined, and that speakers choose referring expressions based on how likely they are to communicate the intended meaning to the listener.

A further issue is that subjecthood or topichood as the main influence on referring expression production is not independently theoretically motivated. In contrast, accounts such as the Expectancy Hypothesis can independently explain the influence of grammatical role on referring expression choice, by referencing the fact that overall, subjects are more likely to be re-mentioned (Arnold, 1998). Since subjects are as a consequence on average more predictable, it would be expected that they are, on average, associated with shorter referring expressions, and that this tendency may even become conventionalized. A probabilistic account that derives this grammatical bias, a variably strong effect of predictability on production, and the production-comprehension asymmetry described by Rohde & Kehler (2014), is presented in Chapter 9.

5.2.4 A Possible Reconciliation

Focusing for the time being on passage completion tasks only, it remains quite unclear which factors are responsible for the varying results of these experiments. One possible generalization that appears to emerge, first brought forward by Bott et al. (2018), p.c. July 25, 2018, is that moderate effects of predictability appear when prompts contain same-gender (rather than opposite-gender) stimuli, and in constrained-completion (but not free completion) tasks. Bott et al. (2018) has replicated this pattern several times with German referring expressions, and has been able to consistently obtain a small effect of predictability on object pronominalization, using implicit causality verbs. I propose some possible reasons for why such a pattern might emerge.

First, although this is not directly addressed by Bott et al. (2018), the effect of predictability appears to be much stronger in more interactive and natural tasks, such as the first two experiments in Rosa & Arnold (2017), where they emerge even

in the case of opposite-gender prompts. In non-interactive tasks, on the other hand, where participants complete written passages, opposite-gender prompts fail to elicit an effect of predictability. I argue that this is due to there being little communicative utility in using a name instead of a pronoun, even when a referent is unpredictable, given that the pronoun's reference is unambiguous (I present a formal argument for this in Section 9.2). I similarly propose an argument for why an effect might still be detectable in more naturalistic, interactive tasks in Section 6.5.

Second, a potential reason for why constrained-reference tasks might elicit an effect of predictability, whereas free-completion tasks do not, is twofold. As Fukumura & van Gompel (2010) observed, not constraining reference may result in so few references to the non-preferred entity, that the conclusions are statistically unreliable. Further, one might imagine that in free-completion tasks, even in those cases where a non-preferred entity is mentioned, that entity must meet some threshold of contextual predictability *from the point of view of the producer*, in order to be mentioned at all, given the instructions to write the 'most likely' continuation.

Those referents that are more contextually predictable *from the producer's point of view*, even if not on average, are likely relatively more likely to be pronominalized. If the effect of predictability on production is already relatively weak, the fact that in a free-completion task, the producer will alternate only between referents that are *moderately to highly* contextually predictable, from their point of view, may wash any effect of predictability on production out. In contrast, in a constrained-completion task, participants are forced to refer to referents which, from their point of view, may be highly contextually unpredictable.

This suggests the following: contrary to what Rosa & Arnold (2017) put forward, their results are not due to their use of transfer-of-possession verbs, given that Bott et al. (2018) also obtains an effect of predictability using implicit causality verbs. It therefore appears that effects arise so long as at least one of the following holds:

1. The task is highly interactive, with a rich discourse context.

2. The task uses constrained reference, and same-gender prompts.

In the following section, I describe my attempt to design an experiment that tests this hypothesis, and addresses the other issues I've raised in the preceding sections.

5.3 Experiment Background

In the following chapter, I first of all test Rosa & Arnold (2017)'s assertion that an effect of predictability on referring expression choice arises when transfer-of-possession verbs are used to manipulate continuation biases, but not when implicit causality verbs are used. To date, with the exception of Bott et al. (2018)'s recent work, effects of predictability have indeed been consistently found only when transfer-of-possession

verbs were used. As I argue in the preceding section, however, Rosa & Arnold (2017)'s assertion is likely to be false. I therefore use both implicit causality and transfer-of-possession verbs to manipulate referent continuation likelihood, with otherwise no variation in the task or stimulus design, to determine whether verb type is in fact critical to eliciting an effect of predictability.

Second, I test a hypothesis I separately developed, that a more lengthy or costly referring expression should result in a greater tendency to pronominalize subsequent references to that expression. In short, given an efficiency-based account, the primary motivation for using a pronoun over a more complex expression is conciseness and conservation of articulatory effort. However, pronouns and proper names, particularly as used in most of the experimental designs above, are hardly different in terms of length and effort: *Mary*, for instance, is only one letter longer than *she*. This suggests that in order to see an effect of efficiency on pronominalization, it's important to ensure that the alternative way of referring back to an entity is significantly different in length. To this end, in addition to using name-name pairs in my stimuli, I also use definite description-definite description pairs as antecedents. A referent such as *the older woman in the park*, even if shorted to *the woman*, is still substantially longer than a pronoun.

In short, if pronominalization is indeed motivated by a desire for greater efficiency, then having the alternative expression be substantially longer should both increase pronominalization overall, and increase pronominalization specifically in those contexts where the intended referent would still be recoverable (i.e., contexts where the referent is predictable). Tily & Piantadosi (2009) found that pronouns were more likely to be used for predictable referents than definite descriptions, but not more likely than nouns, providing some empirical support to this assertion.

Finally, I specifically test Bott et al. (2018)'s claim that the critical difference between the different passage completion experiments is that of task design, and that using a constrained-reference paradigm and same-gender entities in stimulus prompts elicits at least a modest effect of predictability. To this end, I run both a free-completion and a constrained-completion version of the experiment, with the free-completion experiment using only opposite-gender stimuli, and the constrained-reference experiment using only same-gender stimuli. If differences are found between these two task designs, as would be predicted by this account, this also provides motivation to further investigate which manipulations precisely contribute to the emergence of a predictability effect, and under what conditions.

To quickly look ahead, neither of my experiments show any effect of predictability on referring expression production. I am unable to substantiate Bott et al. (2018)'s hypothesis, and I do not find any effects of greater referring expression length and verb type on whether predictability affects referring expression choice. I am thus unable to fully reconcile the disparity in results detailed above, or to account for the positive results reported in Rosa & Arnold (2017). Even so, in Chapter 9, I demonstrate how a Rational Speech Act model can straightforwardly account for this pattern of

observations. The task of seeing if this model can account for future experimental results, if Rosa & Arnold (2017) can be fully replicated (and whether effects would emerge if implicit causality verbs were used in their paradigm), and if Bott et al. (2018)'s observations can be replicated outside of German, is left to future work.

Chapter 6

Referring Expression Choice: Results

In this chapter, I present a series of passage completion experiments I conducted to build upon the existing set of results in the domain of referring expression choice, and the role of referent predictability in determining said choice. As mentioned in the preceding chapter, one of the current challenges is the lack of clarity over whether the effect of predictability modulates production only in the context of some verb-modulated continuation biases, but not others.

Rosa & Arnold (2017) argue that predictability modulates referring expression choice in the context of transfer-of-possession verbs, but not in the context of implicit causality verbs (which are used in Rohde & Kehler, 2014; Fukumura & van Gompel, 2010). However, none of the published experiments to date have compared implicit causality (IC) and transfer-of-possession (ToP) verbs head-to-head using the same paradigm and task design. I aim to answer the question of whether there is a fundamental difference in the effects of predictability, in the context of these two verb types, by using experimental stimuli incorporating both verb types, and testing them using the same tasks, and on the same participant population. I further expand the number of unique ToP verbs tested from 15 to 24.

Second, all experiments to date focus solely on the choice between pronouns and proper names. As mentioned in the preceding chapter, this is problematic, given that first proper names and pronouns differ little in their length, and the articulatory effort required by the speaker. I therefore use experimental stimuli with both proper name antecedents, and lengthy definite description antecedents. Finally, I test Bott et al. (2018)'s hypothesis that the effects of predictability on referring expression choice can be detected so long as one uses a constrained-reference paradigm, and same-gender antecedents in prompts (to ensure that any pronouns used in the continuation are ambiguous in reference).

To look ahead – what I find is that regardless of the manipulation, the results show a null effect of predictability on referring expression choice. I do, however, find that speakers are in general more likely overall to use pronouns when the alternative

is using a lengthy noun phrase (rather than a short proper name). Similarly, I find, following Bott et al. (2018), that speakers are less likely to use pronouns in the context of same-gender antecedents – where the pronouns would be ambiguous. In Section 9.2, I present a potential account for these results, as well as possible steps forward. In Chapter 9, I further present a set of probabilistic models which predict the hypothesized results, and well as at least partially account for the set of results found in the literature to date. The data I provide in this chapter does not provide sufficient empirical data to test the validity of these model predictions, but they similarly provide a way forward to determining, ultimately, whether referent predictability affects referring expression choice in a limited set of contexts.

6.1 Materials and Methods

6.2 Prompt Design

In this section, I describe the motivation behind the design of experimental prompts. Relevant factors manipulated, within and between experiments, are verb type (transfer-of-possession vs. implicit causality), antecedent form (proper name vs. definite description), and same vs. different antecedent gender (which determines whether pronoun reference is ambiguous, or not).

Verb Type

To date, almost all passage completion experiments which have shown an effect of referent predictability on referring expression choice have used transfer-of-possession (ToP) verbs to manipulate referent predictability. ToP constructions, which describe the transfer of an object from a "source" to a "goal" entity, which may appear in either subject or object position, consistently bias towards goal continuations (Stevenson et al., 1994; Rohde, 2008):

(1) John$_{source}$ gave the book to Mary$_{goal}$. → *Mary* more likely continuation

(2) John$_{goal}$ took a book from Mary$_{source}$. → *John* more likely continuation

On the other hand, most of the passage completion experiments which show no effect of referent predictability on referring expression choice use implicit causality verbs. Implicit causality verbs introduce a semantic bias towards continuations referring back to the *causal* entity, or *stimulus* (vs. *experiencer*), which can appear in either subject position (subject-biased IC verbs) or object position (object-biased IC verbs):

(3) John$_{stimulus}$ amazed Mary$_{experiencer}$. → *John* more likely continuation

(4) John$_{experiencer}$ admired Mary$_{stimulus}$. → *Mary* more likely continuation

One exception to this pattern is Bott et al. (2018), who show an effect of referent predictability on referring expression choice in the case of implicit causality verbs. However, this still leaves open the possibility, as argued by Rosa & Arnold (2017), that there is some special property of the thematic roles (for example) in ToP constructions which enables a greater and more replicable effect of predictability on reference production. This is difficult to determine, as past experiments have consistently used different prompt and task designs, verb types aside. As a result, I include both implicit causality and transfer-of-possession prompts in both experiments I run, otherwise controlling for prompt, experiment population, and task design.

In the case of ToP verbs, I constructed 24 stimulus pairs using distinct transfer-of-possession verbs found in Levin (1993). In the case of each stimulus pair, as in the case of 1 and 2, I controlled for the identity of the entities (counterbalancing name/description grammatical position), the transferred object, and the activity type, while varying the direction of transfer. This addresses some of the criticisms made of Arnold (2001), where stimulus prompt pairs differed significantly in discourse context. The full stimulus set can be found in Appendix E.2.2.

In the case of IC verbs, I used the full set of 40 stimulus items, including 20 subject-biased IC verbs, and 20 object-biased IC verbs, found in Rohde (2008) and used in Rohde & Kehler (2014). I altered only entity names (and using definite descriptions in place of names, as described below). The full stimulus set can be found in Appendix E.2.1.

Entity Form

To date, all studies have used proper names as antecedents. This is problematic, assuming that speakers choose pronouns for more predictable names in order to be efficient, given that there is little difference, by way of word length or effort, between, say *she* and *Mary*. I therefore hypothesize that using longer definite description antecedents, such as *the old woman by the bench*, may prompt more pronominalization overall, but also more pronominalization of predictable (and therefore recoverable) elements.

All stimuli, which were distributed across different lists, either used two proper names as antecedents, or two lengthy definite descriptions as antecedents:

(5) *Mary* gave the book to *John*.

(6) *The old woman in the park* gave the book to *the man behind the counter*.

All adjective and adjunctive modifiers in the definite descriptions were chosen to be maximally semantically neutral, and to describe physical appearance (hair color, age, clothing item color), physical location, or identifiers as to where the individual is known from (e.g., from class, from the gym, from work). To control for any residual semantic effects of the modifiers or entity gender, the grammatical position/order of the entities in the construction was alternated across lists.

Antecedent Gender

Some studies have used opposite-gender antecedents in prompts (Arnold, 2001; Fukumura & van Gompel, 2010), as it eases the task of disambiguating intended reference in passage continuations. However, using opposite-gender antecedents in prompts also lowers the relative utility of names in uniquely identifying antecedents – reducing participant motivation to use names for less predictable antecedents (as there is no utility in doing so). In Section 9.2, I show that, as long as a clean channel is assumed, the use of opposite-gender antecedents would in fact be expected to nullify any effects of referent predictability. These effects should still be attenuated, relative to prompts with same-gender antecedents, even given an assumption of a noisy channel (Section 9.2).

However, Rohde & Kehler (2014) show no effect of referent predictability on pronominalization, despite using same-gender antecedents. The experiments in Rosa & Arnold (2017) use both same-gender and opposite-gender antecedents, but show no significant difference in effect magnitude between the two types of prompts, save in their last experiment (where an effect was seen for same-gender, but not opposite-gender antecedents). Bott et al. (2018) show an effect of referent predictability on pronominalization using same-gender antecedents in prompts. Overall, it remains unclear to what degree the gender configuration in prompts matters.

I therefore run two experiments – one less likely to elicit an effect (opposite-gender antecedents in prompts), and one more likely to elicit an effect (same-gender antecedents in prompts):

(7) John gave the book to Mary.

(8) John gave the book to Bill.

The two experiments also differ in task type, as discussed in the next section.

6.2.1 Procedure

Two experiments were run. One employed a free completion paradigm – the most commonly used (e.g. Stevenson et al., 1994; Arnold, 2001; Rohde & Kehler, 2014). The second employed a constrained completion paradigm, where it was indicated to participants which referent should be mentioned first in the continuation (cf. Fukumura & van Gompel, 2010; Rosa & Arnold, 2017; Bott et al., 2018, p.c. July 25, 2018). Bott et al. (2018), as discussed in Section 5.2.4, have hypothesized specifically that effects of predictability on pronominalization arise given constrained completion paradigms, and not (or to a lesser degree) in free completion paradigms.

In *free completion*, by far the most dominant paradigm used so far, participants are free to complete the passage as they wish, mentioning either one of the antecedents (or something else entirely) first. The benefit of this approach is that it's arguably more natural, as the participant is not artificially constrained in how they express

themselves. However, it is arguably *less* natural in another respect – normally speakers do not have to construct a scenario from scratch, given a context-free utterance, with no indication even as to which entity the discourse is "about." In free completion, the participant is presented with only the prompt, and then asked to write what they think to be the most likely continuation:

(9) John gave the book to Mary. _____

In the case of *constrained completion*, participants are 'told,' either via pictures which show one of the entities engaging in a subsequent activity (Rosa & Arnold, 2017), or by framing or underlying the relevant entity, which entity to refer to next, or which entity is to be central in the following discourse. The benefit of this approach is that it makes the passage completion paradigm somewhat more natural, with respect to producers "knowing" (independently of their own intuition) who is more 'central' to the following discourse. Another benefit, noted by Fukumura & van Gompel (2010), is that more data is gathered for the 'less predictable' referent, making conclusions more statistically reliable. The downside, of course, is that this paradigm unnaturally constrains how participants may express themselves. In constrained completion, participants are presented with the prompt, and then asked to write (in this case) the most likely continuation *about the entity contained in the box*:

(10) John gave the book to [Mary]. _____

As I show in Section 9.3, and discuss in Section 5.2.4, if one assumes that participants avoid lower-predictability referents in free-completion tasks, but are forced to refer to them in constrained-completion tasks, then effects of predictability on pronominalization should be greater in constrained-completion tasks.

As mentioned above in Section 5.2.4, effects of predictability on pronominalization are also expected to be greater when using same-gender antecedents in prompts, than when using opposite-gender antecedents. I therefore run two experiments: one which uses the *free completion* paradigm and opposite-gender antecedents in prompts, and one which uses the *forced completion* paradigm, and same-gender antecedents in prompts. The first experiment therefore uses a prompt design and passage completion paradigm *less* likely to show an effect of predictability on pronominalization, and the second experiment used a prompt design and completion paradigm *more* likely to show an effect, according to Bott et al. (2018)'s hypothesis, and the distribution of experimental results to date.

6.2.2 Presentation and Exclusion Criteria

Stimuli and stimuli conditions were distributed among multiple lists (16 in the case of the first experiment, and 32 in the case of the second). All participants saw 10 of

40 implicit causality stimuli, 12 of 48 transfer-of-possession stimuli (only one of each pair), and 10 non-IC, non-ToP filler stimuli (see Appendix E).

In both cases, participants were recruited on Amazon Mechanical Turk. Data from participants who reported their native language as other than American English, or did not follow instructions, was excluded prior to analysis (see individual experiments for details), and additional participants were recruited to replace them. Participants were asked to write the most likely continuation ('about' the framed entity, in the constrained completion task), while avoiding humor, and without turning the different items into one story.

Continuations in which the first-mentioned entity was in an unusual syntactic position, such as a cleft sentence or a subordinate clause, were not included in the analysis, in order to control for the effect of syntactic position or topicality. Cases where there was lexical material more than a few words long preceding the first entity mention, beyond temporal markers or connectives, were similarly excluded, also to control for the effect of length to last mention.

In all cases, I separately analyzed the full remaining data set, as well as a subset of the data in which participants used a minimum of 2 pronouns, and 2 lengthier expressions, in their continuation, to ensure that participants were fully engaged in the task and not simply repeating the same constructions (cf. Rosa & Arnold, 2017). I report results from the dataset with some variability in expressions used, except where results from the full data set differ.

6.3 Experiment 1: Free Completion with Opposite-Gender Prompts

In this experiment, participants were instructed to write the most likely continuation to the prompt. Prompts contained two opposite-gender antecedents, removing any ambiguity of pronominal reference used by participants in continuations. The predicted effects are the following:

1. If Rosa & Arnold (2017)'s hypothesis is correct, then an effect of predictability on pronominalization should be observed in the case of transfer-of-possession verbs, but not implicit causality verbs. If Rohde & Kehler (2014)'s hypothesis is correct, then no effect of predictability on pronominalization should be observed in either case. If Bott et al. (2018)'s hypothesis is correct, then effect magnitude should not differ substantively by verb type, but may be minimal to nonexistent, given the use of opposite-gender antecedents in prompts, and free completion as a paradigm.

2. If an effect of predictability on pronominalization is detected, then it should be stronger in the case of stimuli with definite description antecedents, as par-

ticipants may have comparatively little cause to use pronouns instead of short proper names.

3. References back to definite description antecedents should overall be pronominalized more than references back to proper name antecedents, due to the increased effort needed to produce definite descriptions.

6.3.1 Participants

244 participants were recruited on Amazon Mechanical Turk, with IP addresses constrained to the US. 11 participants were excluded from analysis due to reporting their native language as other than American English, and more participants were recruited to replace them. 14 participants were excluded due to either not following instructions (e.g., writing one-word continuations) or making numerous evidently non-native grammatical and word choice errors in their continuations, with more participants recruited to replace them, leaving 244 eligible participants of an originally planned 240[1].

This left two sets of data. The 'full' set contained all participants who were native speakers of English, and followed instructions (n=244). A more constrained set included only those participants who used at least two pronouns, and at least two names or definite descriptions, in their continuations, ensuring that they were paying adequate attention and not simply repeating the same construction (n=142). In all cases below, I report results from the more constrained dataset, except where the results differ from the 'full' dataset, in which case I report both.

6.3.2 Results

In this experiment, as well as the next, I used logistic mixed effects models with the maximal converging by-subject and by-item random effects structure (Barr et al., 2013), as implemented in the `lme4` package (Bates et al., 2020). Models with a maximal random slope structure often failed to converge, in which case the slopes or slope interactions accounting for the least amount of variance were removed, until the model converged. *P*-values were obtained using the Satterthwaite approximation for degrees of freedom, as implemented in the `lmerTest` package (Kuznetsova et al., 2017).

Continuations

First, I see if the IC and ToP continuation biases are replicated. In fact, I find that I replicate the implicit causality bias, although it is much weaker than in much of

[1] 4 extra participants beyond what was originally planned were able to submit their data through the experimental portal.

the published literature (despite using the same stimuli as Rohde & Kehler, 2014). Table 6.1 summarizes this effect for names ($p < .001$), while Table 6.2 summarizes this effect for descriptions ($p < .01$); see Figure 6.1 for an illustration of both results.

Table 6.1: Experiment 1: Replication of thematic role continuation biases (IC verbs; names). The predicted element is significantly more likely to be referred to.

	β	SE(β)	**z**	**p**
Intercept	-1.06	0.13	-8.30	**<.001**
Predicted Continuation Bias	0.88	0.20	4.46	**<.001**

Table 6.2: Experiment 1: Replication of thematic role continuation biases (IC verbs; descriptions). The predicted element is significantly more likely to be referred to.

	β	SE(β)	**z**	**p**
Intercept	-0.24	0.15	-1.60	0.1
Predicted Continuation Bias	0.80	0.28	2.82	**<.01**

The transfer-of-possession bias, in contrast, appears quite robust. Table 6.3 shows this effect for names ($p < .001$), while Table 6.4 shows this effect for descriptions ($p < .001$); see Figure 6.2 for both. To note, the data for both verb types was collected from the same population, with all participants presented with both prompt types, so this difference is not credibly attributable to the population or task.

Table 6.3: Experiment 1: Replication of thematic role continuation biases (ToP verbs; names). The predicted element is significantly more likely to be referred to.

	β	SE(β)	**z**	**p**
Intercept	-0.78	0.12	-6.35	**<.001**
Predicted Continuation Bias	1.74	0.20	8.81	**<.001**

Names: Effect of Predictability on Pronominalization

In the case of names, it is clear that the likelihood of referent mention does not affect the likelihood of pronominalization, of either the subject or the object, in the case of both IC ($p = 0.2$) and ToP verbs ($p = 0.3$). Table 6.5 summarizes this effect for names, while Table 6.6 summarizes this effect for descriptions; see Figure 6.3 for an illustration of both.

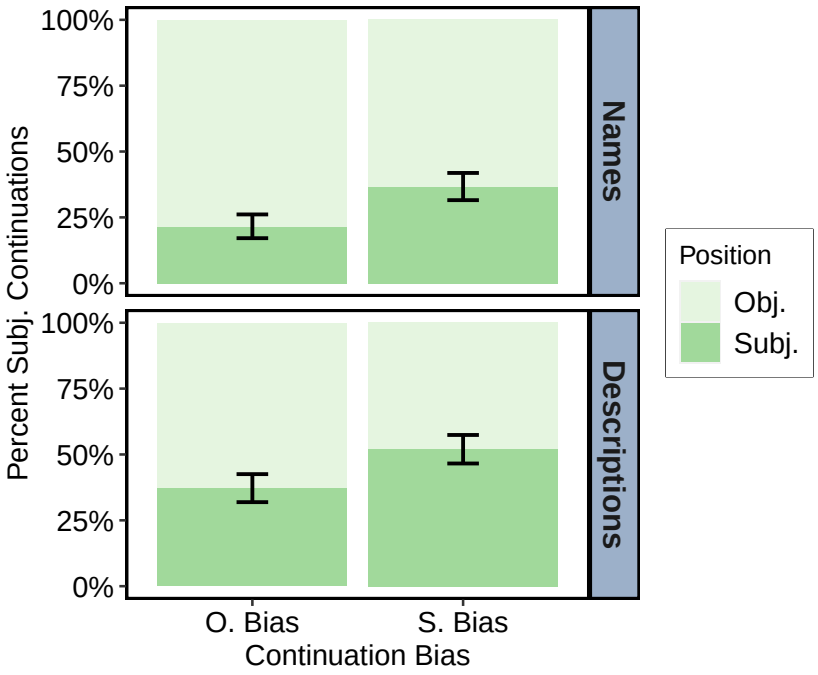

Figure 6.1: Replication of thematic role continuation biases (IC verbs).

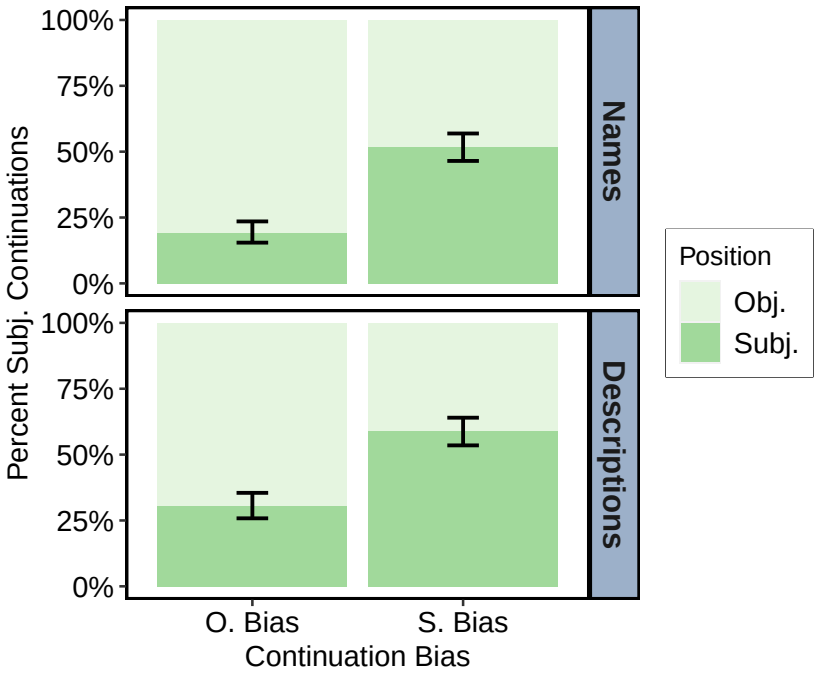

Figure 6.2: Replication of thematic role continuation biases (ToP verbs).

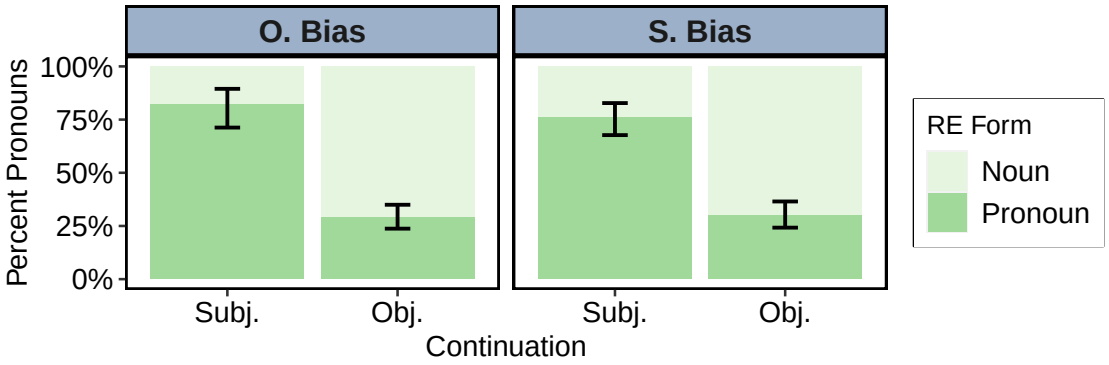

Figure 6.3: Experiment 1: Pronominalization of proper name antecedents (IC verbs).

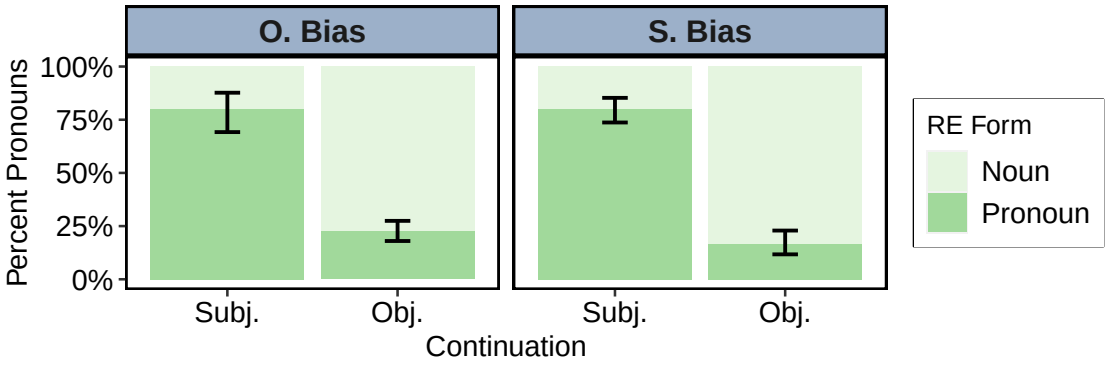

Figure 6.4: Experiment 1: Pronominalization of proper name antecedents (ToP verbs).

Table 6.4: Experiment 1: Replication of thematic role continuation biases (ToP verbs; descriptions). The predicted element is significantly more likely to be referred to.

	β	SE(β)	**z**	**p**
Intercept	-0.20	0.13	-1.47	0.1
Predicted Continuation Bias	1.43	0.28	5.10	**<.001**

Table 6.5: Experiment 1: Pronominalization of proper name antecedents (IC verbs). There is a significant effect of subject/object reference on pronominalization rates, but no effect of how likely the referent is to be mentioned.

	β	SE(β)	**z**	**p**
Intercept	-0.27	0.20	-1.36	0.2
Reference	-2.88	0.35	-8.35	**<.001**
Bias	0.19	0.27	0.71	0.5
Reference * Bias	0.94	0.68	1.38	0.2

Table 6.6: Experiment 1: Pronominalization of proper name antecedents (ToP verbs). There is a significant effect of subject/object reference on pronominalization rates, but no effect of how likely the referent is to be mentioned.

	β	SE(β)	**z**	**p**
Intercept	0.41	0.26	1.56	0.1
Reference	-4.04	0.47	-8.61	**<.001**
Bias	0.18	0.26	0.68	0.5
Reference * Bias	-0.59	0.55	-1.07	0.3

Descriptions: Effect of Predictability on Pronominalization

In the case of descriptions, one again sees that the likelihood of mention does not affect the likelihood of pronominalization, despite the increased length of the expressions speakers must use to refer back to the entities, in the case of both IC ($p = 0.5$) and ToP verbs ($p = 0.1$); Table 6.7 and Figure 6.5 show this effect for names, while Table 6.8 and Figure 6.6 show this effect for descriptions.

Effect of Increased Antecedent Length on Pronominalization

What is clear, however, is that the prediction that speakers are *overall* more likely to pronominalize lengthier and more effortful references pans out. This is the case for both IC ($p < .001$) and ToP verbs ($p < .001$). Table 6.9 and Figure 6.7 show this

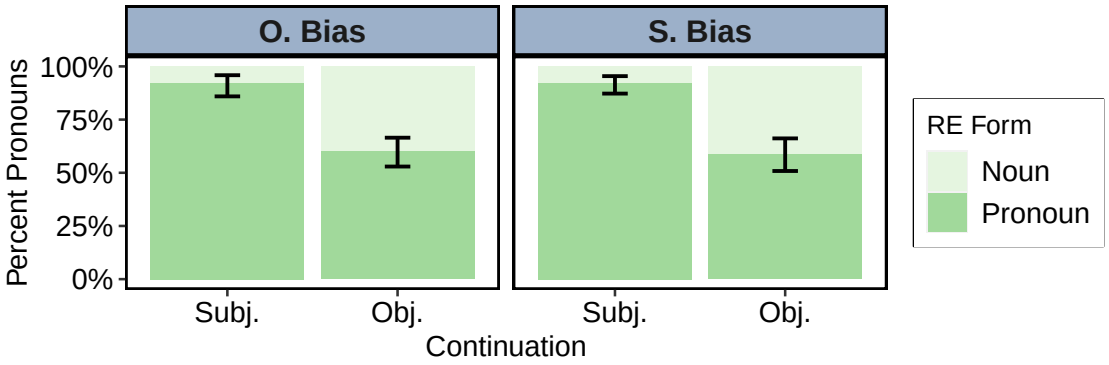

Figure 6.5: Experiment 1: Pronominalization of definite description antecedents (IC verbs).

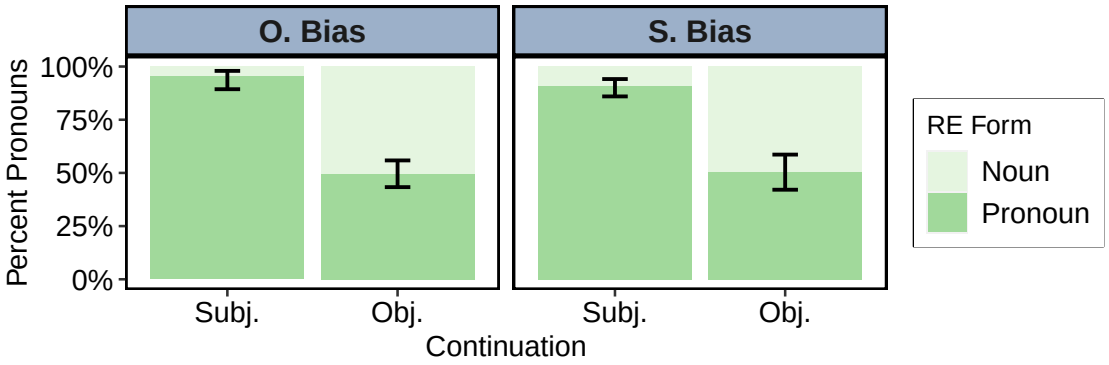

Figure 6.6: Experiment 1: Pronominalization of definite description antecedents (ToP verbs).

Table 6.7: Experiment 1: Pronominalization of definite description antecedents (IC verbs). There is a significant effect of subject/object reference on pronominalization rates, but no effect of how likely the referent is to be mentioned.

	β	SE(β)	**z**	**p**
Intercept	1.99	0.25	7.96	**<.001**
Reference	2.63	0.34	7.68	**<.001**
Bias	-0.09	0.29	-0.31	0.8
Reference * Bias	-0.41	0.61	-0.66	0.5

Table 6.8: Experiment 1: Pronominalization of definite description antecedents (ToP verbs). There is a significant effect of subject/object reference on pronominalization rates, but no effect of how likely the referent is to be mentioned.

	β	SE(β)	**z**	**p**
Intercept	2.25	0.43	5.27	**<.001**
Reference	3.86	0.76	5.07	**<.001**
Bias	-0.39	0.37	-1.05	0.3
Reference * Bias	-1.36	0.76	-1.79	0.1

effect for names, while Table 6.10 and Figure 6.8 show this effect for descriptions.

Table 6.9: Experiment 1: Effect of antecedent length on pronominalization (IC verbs). Increased antecedent length significantly increases the likelihood of pronominalization.

	β	SE(β)	**z**	**p**
Intercept	-0.34	0.16	-2.06	**<.05**
Length	1.82	0.17	10.45	**<.001**

Table 6.10: Experiment 1: Effect of antecedent length on pronominalization (ToP verbs). Increased antecedent length significantly increases the likelihood of pronominalization.

	β	SE(β)	**z**	**p**
Intercept	-0.50	0.17	-2.99	**<.01**
Length	1.81	0.18	10.20	**<.001**

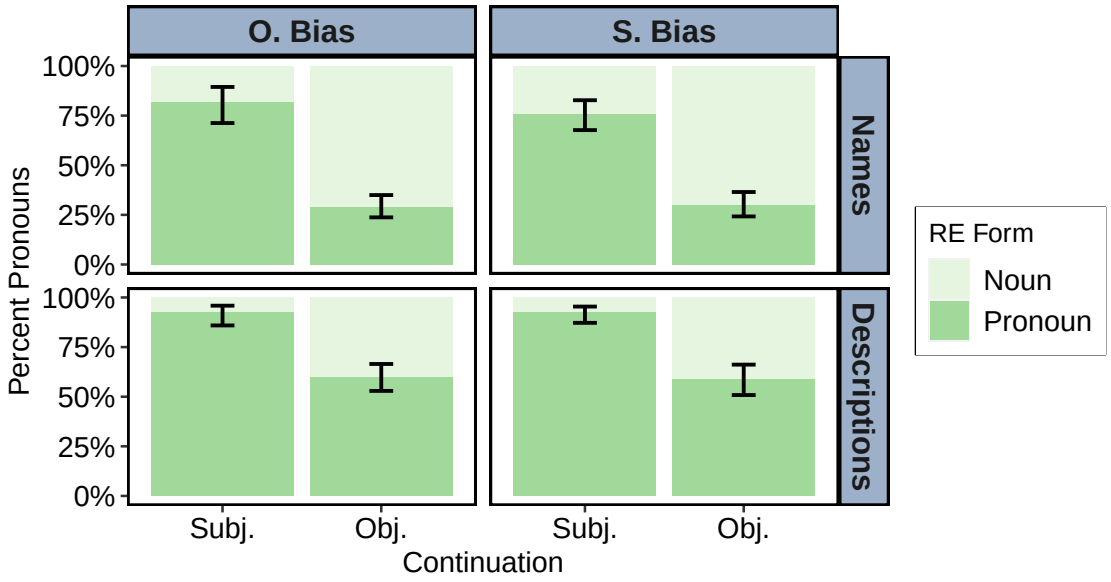

Figure 6.7: Experiment 1: Pronominalization of antecedents by length (IC verbs).

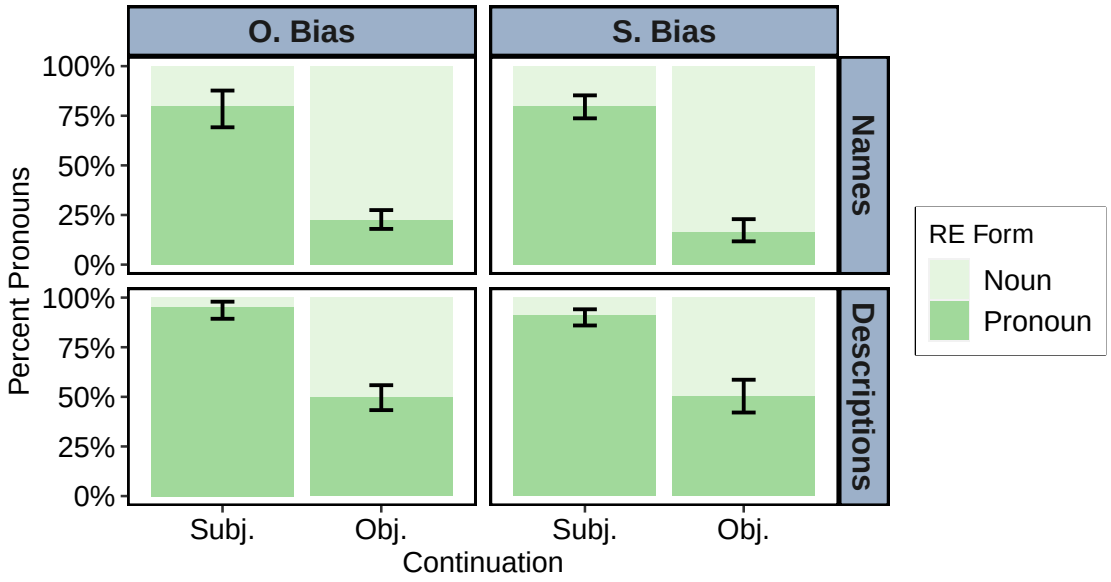

Figure 6.8: Experiment 1: Pronominalization of antecedents by length (ToP verbs).

6.3.3 Discussion

Overall, it appears that there is no effect at all of predictability on choice of referring expression, for either verb type. This is at odds with the argument put forward by Rosa & Arnold (2017) that one should be consistently able to see an effect of referent predictability on referring expression choice, in the case of transfer-of-possession verbs. In this case, the results for both verb types, when population and task are matched, are remarkably similar.

These results are similarly consistent with the predictions the Rohde & Kehler (2014) account puts forward, as well as the predictions the Bott et al. (2018) account puts forward. The Rohde & Kehler (2014) account would argue that this is evidence that referent predictability does not affect referring expression choice – though it should again be noted that it is currently unable to account for the results in Rosa & Arnold (2017) or Bott et al. (2018)[2]. The Bott et al. (2018) account would argue that the lack of effect is due to: a) the use of opposite-gender vs. same-gender antecedents in prompts, and b) the use of a free completion, vs. constrained completion paradigm.

The following experiment, which uses a constrained completion task and same-gender antecedents, is meant to disambiguate between the two accounts. The Rohde & Kehler (2014) account would predict no effect of referent predictability on referring expression choice. The Bott et al. (2018) account, however, would predict that the results of this experimental paradigm and prompt design should show an effect of referent predictability on pronominalization rates.

6.4 Experiment 2: Constrained Completion with Same-Gender Prompts

In this experiment, participants were instructed to write the most likely continuation *about the entity in the frame*, in effect forcing participants to refer both to less likely entities, as well as more likely entities. Prompts contained two same-gender antecedents, so that any pronominal reference in the continuation would be ambiguous. The predicted effects are the following:

1. If Rohde & Kehler (2014)'s hypothesis is correct, then there should again be no effect of predictability on referring expression choice. If Bott et al. (2018)'s account, on the other hand, is correct, then one should see a significant effect of predictability on referring expression choice, at least in the case of objects, which appear to be less vulnerable to a ceiling effect, with respect to pronominalization.

[2]To note, it is possible that Rohde & Kehler (2014)'s account only holds for English, and the system of referring expressions in German is a separate case.

2. If the account proposed by Rosa & Arnold (2017) is (partly) correct, then the effect should be larger for transfer-of-possession verbs than for implicit causality verbs. As can be seen in the first experiment, the continuation bias for transfer-of-possession verbs is quite robust, compared to that for implicit causality verbs.

3. If an effect of predictability on pronominalization is detected, then it should be stronger in the case of stimuli with definite description antecedents.

4. References back to definite description antecedents should overall be pronominalized more than references back to proper name antecedents.

5. As Bott et al. (2018) observed, and as has been previously observed by Arnold & Griffin (2007), overall one should see less pronominalization when reference is ambiguous (second experiment), rather than unambiguous (first experiment).

6.4.1 Participants

417 participants were recruited on Amazon Mechanical Turk, with IP addresses constrained to the US. 10 participants were excluded from analysis due to reporting their native language as other than American English, and more participants were recruited to replace them. 18 participants were excluded due to not following instructions (e.g., writing one-word continuations), with more participants recruited to replace them, or making numerous evidently non-native grammatical and word choice errors in their continuations, leaving 389 eligible participants of an originally planned 384[3].

Again, this leaves two sets of data. The 'full' set contains all participants who were native speakers of English, and followed instructions (n=389). A more constrained set includes only those participants who used at least two pronouns, and at least two names or definite descriptions, in their continuations (n=168). In all cases below, I report results from the more constrained dataset, except where the results differ from the 'full' dataset, in which case I report both.

6.4.2 Results

Names: Effect of Predictability on Pronominalization

In the case of names, it is again clear that likelihood of referent mention does not affect the likelihood of pronominalization, of either the subject or the object, in the case of both IC ($p = 0.9$) and ToP verbs ($p = 0.6$). Table 6.11 shows this effect for names, while Table 6.12 shows this effect for descriptions; see Figure 6.9 for both.

[3]As in the previous experiment, a handful of additional participants were able to submit their data through the experiment portal.

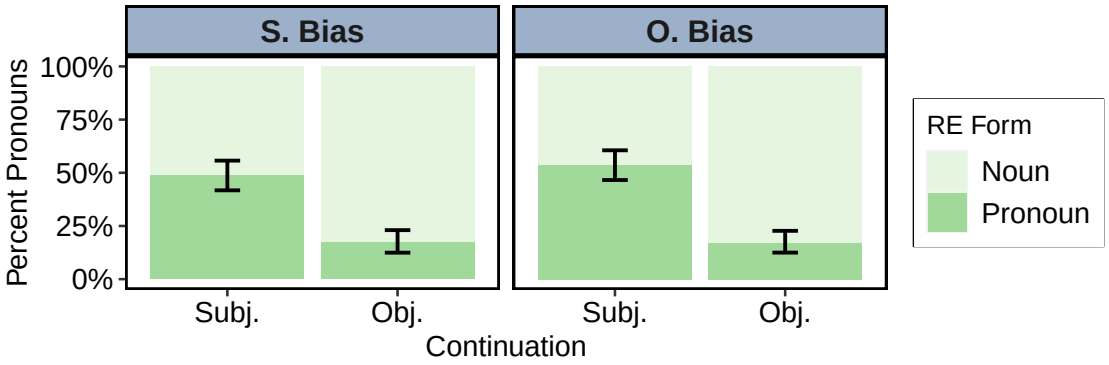

Figure 6.9: Experiment 2: Pronominalization of proper name antecedents (IC verbs).

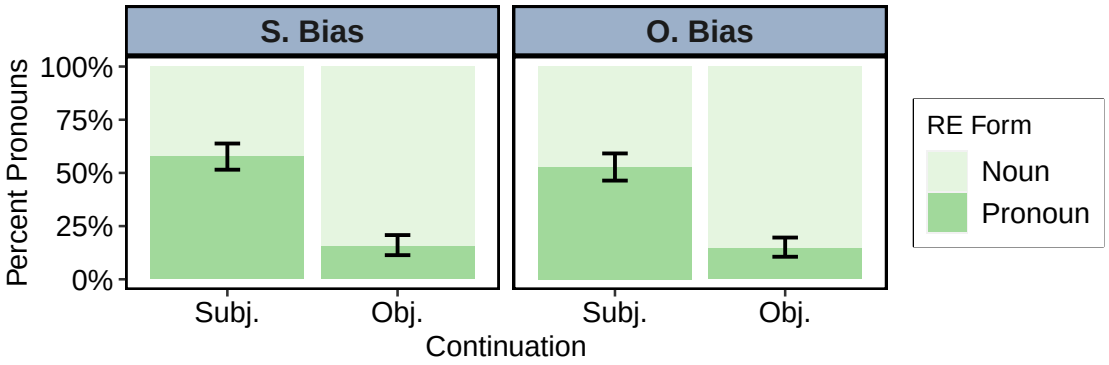

Figure 6.10: Experiment 2: Pronominalization of proper name antecedents (ToP verbs).

Table 6.11: Experiment 2: Pronominalization of proper name antecedents (IC verbs). There is a significant effect of subject/object reference on pronominalization rates, but no effect of how likely the referent is to be mentioned.

	β	SE(β)	**z**	**p**
Intercept	-1.47	0.27	-5.47	**<.001**
Reference	2.93	0.31	9.48	**<.001**
Bias	-0.16	0.23	-0.68	0.5
Reference * Bias	-0.07	0.46	-0.15	0.9

Table 6.12: Experiment 2: Pronominalization of proper name antecedents (ToP verbs). There is a significant effect of subject/object reference on pronominalization rates, but no effect of how likely the referent is to be mentioned.

	β	SE(β)	**z**	**p**
Intercept	-1.30	0.23	-5.59	**<.001**
Reference	3.26	0.29	11.15	**<.001**
Bias	0.30	0.21	1.44	0.2
Reference * Bias	0.20	0.42	0.48	0.6

Descriptions: Effect of Predictability on Pronominalization

In the case of descriptions, it is again clear that the likelihood of mention does not affect the likelihood of pronominalization, in the case of both IC ($p = 0.1$) and ToP verbs ($p = 0.9$). Table 6.13 and Figure 6.11 show this effect for names, while Table 6.14 and Figure 6.12 show this effect for descriptions.

Table 6.13: Experiment 2: Pronominalization of definite description antecedents (IC verbs). There is a significant effect of subject/object reference on pronominalization rates, but no effect of how likely the referent is to be mentioned.

	β	SE(β)	**z**	**p**
Intercept	0.84	0.27	3.11	**<.01**
Reference	2.23	0.37	5.98	**<.001**
Bias	0.09	0.22	0.40	0.7
Reference * Bias	-0.72	0.45	-1.60	0.1

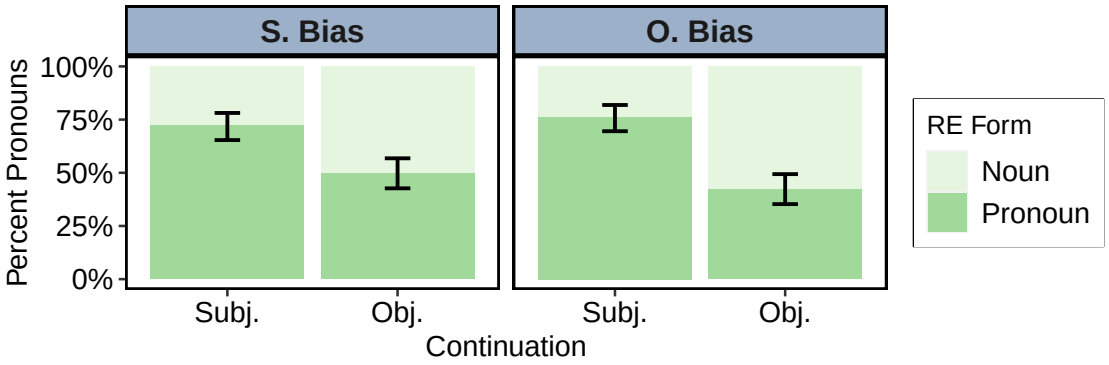

Figure 6.11: Experiment 2: Pronominalization of definite description antecedents (IC verbs)

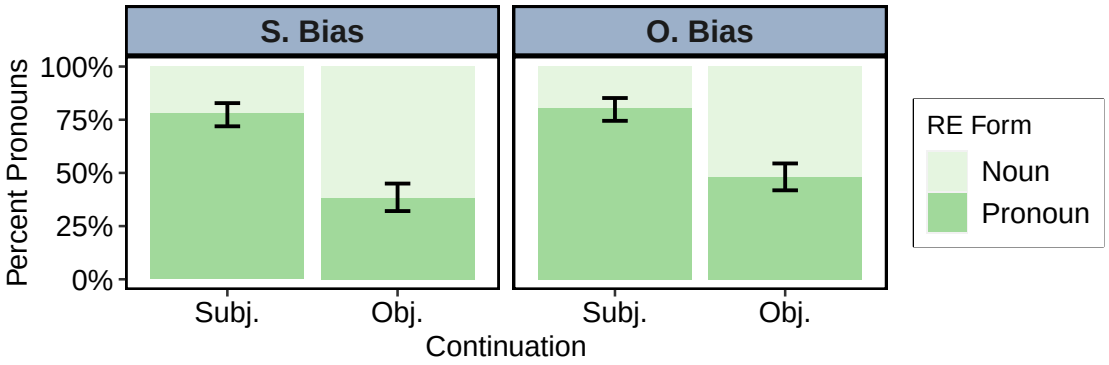

Figure 6.12: Experiment 2: Pronominalization of definite description antecedents (ToP verbs).

Table 6.14: Experiment 2: Pronominalization of definite description antecedents (ToP verbs). There is a significant effect of subject/object reference on pronominalization rates, but no effect of how likely the referent is to be mentioned.

	β	SE(β)	**z**	**p**
Intercept	1.08	0.23	4.64	**<.001**
Reference	2.99	0.29	10.18	**<.001**
Bias	-0.57	0.29	-2.01	**<.05**
Reference * Bias	0.05	0.62	0.09	0.9

Effect of Increased Antecedent Length on Pronominalization

Again, it is clear that speakers are *overall* more likely to pronominalize lengthier and more effortful references. This is the case for both IC ($p < 0.001$) and ToP verbs ($p < 0.001$). Table 6.15 and Figure 6.13 show this effect for names, while Table 6.16 and Figure 6.14 show this effect for descriptions.

Table 6.15: Experiment 2: Effect of antecedent length on pronominalization (IC verbs). Increased antecedent length significantly increases the likelihood of pronominalization.

	β	SE(β)	**z**	**p**
Intercept	-1.08	0.18	-6.02	**<.001**
Length	1.68	0.16	10.54	**<.001**

Table 6.16: Experiment 2: Effect of antecedent length on pronominalization (ToP verbs). Increased antecedent length significantly increases the likelihood of pronominalization.

	β	SE(β)	**z**	**p**
Intercept	-0.86	0.15	-5.69	**<.001**
Length	1.56	0.14	10.87	**<.001**

Effect of Ambiguity on Pronominalization

Here, I look at whether increased ambiguity overall decreases pronominalization, by comparing pronominalization rates in the second experiment (ambiguous reference) to those in the first (unambiguous reference). The results can be seen in Table 6.17 for IC verbs, and Table 6.18 for ToP verbs. Figure 6.15 further illustrates the results.

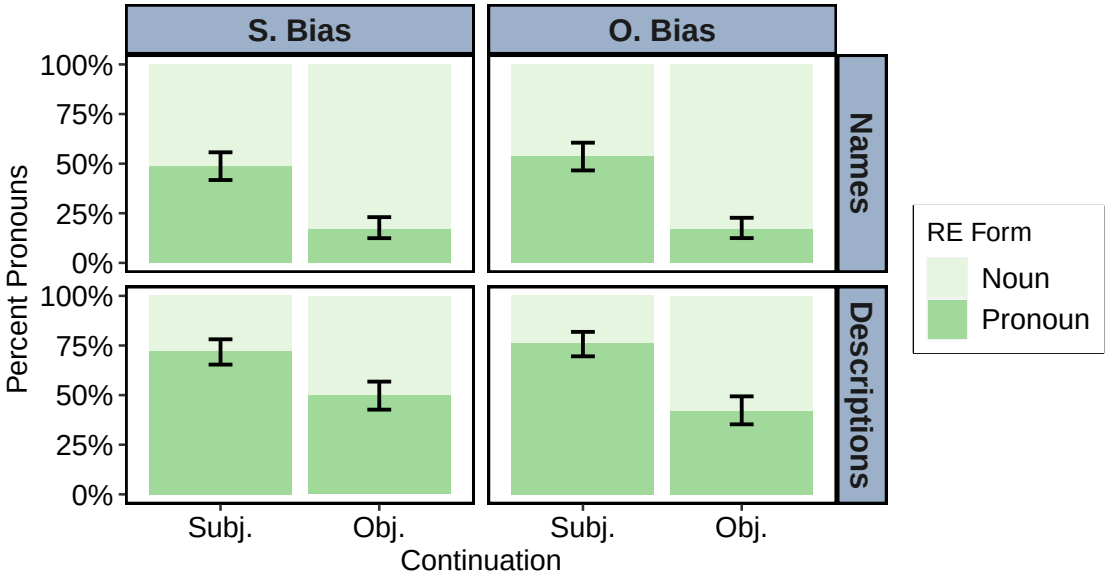

Figure 6.13: Experiment 2: Pronominalization of antecedents by length (IC verbs).

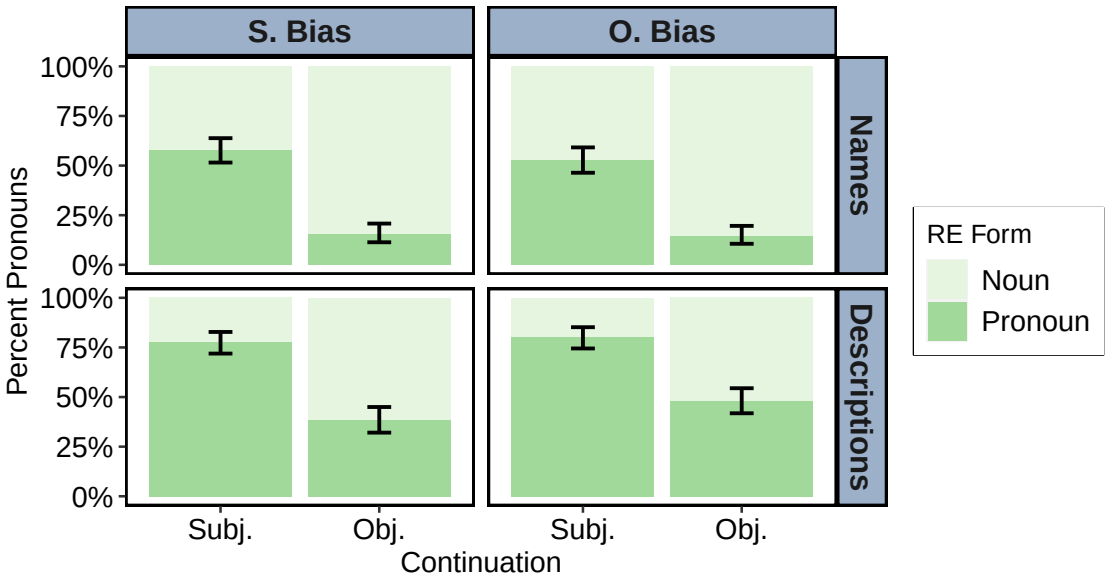

Figure 6.14: Experiment 2: Pronominalization of antecedents by length (ToP verbs).

Table 6.17: Experiment 1 vs. 2: Effect of antecedent ambiguity on pronominalization (IC verbs). Pronominalization rates decrease when reference is ambiguous.

	β	SE(β)	**z**	**p**
Intercept	-0.14	0.06	-2.41	$<$**.05**
Ambiguity	0.49	0.08	6.45	$<$**.001**

Table 6.18: Experiment 1 vs. 2: Effect of Antecedent Ambiguity on Pronominalization (ToP verbs). Pronominalization rates decrease when reference is ambiguous.

	β	SE(β)	**z**	**p**
Intercept	-0.09	0.06	-1.62	0.1
Ambiguity	0.28	0.07	3.87	$<$**.001**

From these results, it appears clear that participants are less likely to use pronouns when those pronouns are ambiguous, in the case of both IC ($p < 0.001$) and ToP verbs ($p < 0.001$), confirming that ambiguity (and perhaps, some degree of audience design) plays a role in speaker choice of referring expression. This replicates the results of Bott et al. (2018), as well as Arnold & Griffin (2007).

6.4.3 Discussion

Overall, the second experiment shows no effect of referent predictability on speaker choice of referring expression. These results are inconsistent with the arguments put forward for Rosa & Arnold (2017): there is no evidence that the effect is reliable or easily replicable (or greater) for transfer-of-possession verbs. Further, although these results by themselves cannot account for the results of the three experiments presented in Rosa & Arnold (2017), or the results of Bott et al. (2018), they are generally inconsistent with the proposal that referent predictability affects referring expression choice.

The validity of these results, however, is supported by the fact that the effect of ambiguity on pronominalization rates is very clearly replicated, as well as the consistency of the results between the two paradigms used in these experiments.

6.5 General discussion

The results of these two experiments, in isolation, are clearly most consistent with the argument put forward by Rohde & Kehler (2014): that speakers choose which referring expression to use based only on the grammatical position or topicality of

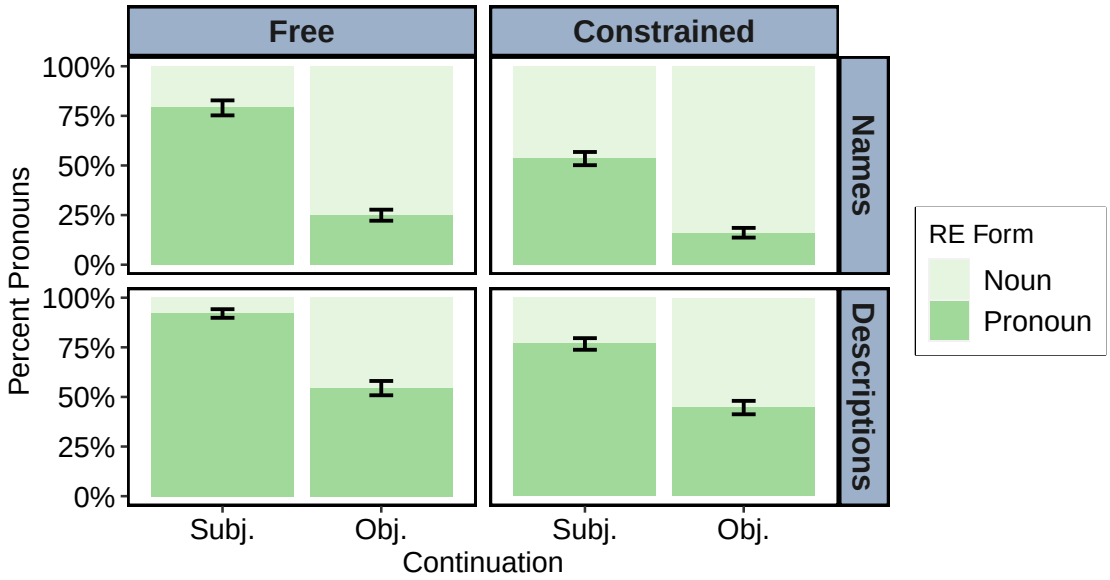

Figure 6.15: Pronominalization by antecedent length and ambiguity. Pronouns are less likely to be used when there is ambiguity as to who the intended referent is.

the antecedent. However, neither their results, nor mine, are able to account for the English data in Rosa & Arnold (2017), nor the German data in Bott et al. (2018).

At this time, the following appears clear: the effect of referent predictability on referring expression choice is at best fragile, and even in Bott et al. (2018)'s studies, is rather weak. The results of Rosa & Arnold (2017) could be an artifact of their limited inventory of transfer-of-possession verbs, indicating that the effect does not generalize beyond that set of words. However, I do not consider this a likely explanation, given the following.

If the data from the above two experiments is limited to only those verbs tested by Rosa & Arnold, the effect remains null – inconsistent with the idea that there is anything 'special' about their verb inventory. Second, the effect is robust, and found in all three of their experiments, although in the last one, it was only seen for items with ambiguous antecedents. Third, although the inventory of verbs is rather limited, most of the transfer-of-possession verbs that exist, or are used in these experiments, are synonyms of, or express a specific variation of, the basic verbs used by Rosa & Arnold (2017). For example, many of the verbs used in my experiments, as categorized in Levin (1993)'s book, are 'give' verbs, and are unlikely to display substantially different behavior from the more basic verb *give*, as used by Rosa & Arnold (e.g., *hand*, *bequeath*, *gift*). There is similarly little appreciable difference between verbs such as *take*, *wrest*, or *grab*. Finally, Rosa & Arnold, with the exception of their last (and weakest) experiment, use a unique task and stimulus design, which to date has not been exactly replicated by others.

Rosa & Arnold (2017) use a highly interactive task design, with a rich discourse context, as described in Section 5.2.1. The weakest results were obtained in their last

experiment, which used a non-interactive passage completion design, where the items were rewritten so that they no longer formed a coherent story. It is possible that, assuming that increased pronominalization of more predictable referents is an element of audience design, that speakers are more likely to choose referring expressions with listener expectations in mind when: 1) there is, in fact, a listener; and/or 2) the discourse is naturalistic, and resembles more typical communication.

To note, there is well-known precedent for online and local effects of predictability (or audience design) appearing only in highly interactive, naturalistic paradigms. Brown & Dell (1987) and Dell & Brown (1991) conducted a series of studies in which participants were exposed to stories in which various acts were carried out using a variety of predictable and unpredictable instruments (e.g., *killing* someone with a *knife* vs. an *ice pick*). Participants were then instructed to describe the stories to a confederate. The hypotheses tested were that 1) typical, or predictable instruments (e.g., *knife*, in the above example) would be more likely to be omitted in story retellings; and 2) participants would be more likely to omit instruments if they had reason to believe that the confederates were already familiar with the story, i.e. could predict or recover the instruments without them being mentioned explicitly. While the first hypothesis was supported, the second was not – participants appeared not to take their listener's prior knowledge and beliefs (and expectations) into account.

However, Lockridge & Brennan (2002) replicated these studies with the original items, but used highly trained confederates, who were specifically instructed to naturalistically interact with the participants in determining what occurred in the stories being retold. In this case, the second hypothesis above was supported: participants were sensitive to the prior knowledge and beliefs of the confederates. The conclusion the authors reached was that local, online effects of audience design may not surface unless the speaker is faced with a believable, interactive audience. Although it would require further replication of the original studies, it minimally appears plausible that the strong effects found in Rosa & Arnold (2017) are due to the far more naturalistic paradigms used, rather than an otherwise unexplainable artifact of stimulus design. It may also account for the effects found in Tily & Piantadosi (2009) and Kravtchenko (2014), which used texts that, while less naturalistic, are nevertheless aimed at a specific audience.

Ultimately, it appears clear that the basic UID model offers a fairly poor explanation of this set of results. If the data from this experiment is taken in isolation, then it makes plainly false predictions with respect to the pronominalization of more predictable referents. There is no clear account for why this might be the case, further. On the other hand, if the set of empirical results to date is taken as a whole, UID continues to offer a fairly poor explanation for why effects are present in some paradigms and with some stimulus designs, but not others, and why they exhibit wildly different strengths – although it would indeed predict more pronominalization in the case of the less predictable entities referred to in constrained completion tasks. As I show in Chapter 9, which remains speculative but hopefully can help guide

future experiments, the Rational Speech Act model, in contrast, accounts variably neatly for almost all of the empirical patterns noted to date, including the following (when matching for task and stimulus design), which are not accounted for by a comparatively simple UID model:

1. The RSA model predicts very straightforwardly that effects should be stronger in the case of same-gender antecedents in prompts, and pronoun ambiguity. The clean channel RSA model, in fact, predicts clearly that *no* effects should be seen in designs that use opposite-gender antecedents. While this does not account for the results of the first two experiments in Rosa & Arnold (2017), I show that incorporating the notion of the noisy channel, discussed in the preceding chapters, predicts a weaker, but stable effect, in the case of opposite-gender antecedents, and lack of pronoun ambiguity.

2. Similarly to Rohde & Kehler (2014) – although I do not explore pronoun comprehension in my experiments – the RSA model, likewise being a probabilistic Bayesian account, straightforwardly accounts for the asymmetry between production and comprehension reported by them. Neither Arnold (2008)'s *Expectancy Hypothesis*, nor the UID model, can straightforwardly account for this insight. This insight is further important enough that it arguably needs to be represented by any model which attempts to account for pronoun production and comprehension.

3. More speculatively, I argue that the RSA model *may* account for the lack of results (or presence of only relatively weak results) in less interactive, less discourse-rich paradigms. What may be occurring is that speakers are either on average only very minimally optimizing their utility, such that effects are too small to be detected, or that what we have on hand, in the case of less interactive and naturalistic paradigms, are by and large *very* unsophisticated speakers, who are concerned only with stating what is true, and not with modeling their listeners' beliefs or expectations (cf. Franke & Degen, 2016) – which is essentially what is assumed by Rohde & Kehler (2014). In contrast, the more interactive, naturalistic designs in Rosa & Arnold (2017) may be prompting speakers to, on average, become far more sophisticated agents, who take into account the utility of their utterances to the literal (or pragmatic) listener. This, however, would have to be empirically determined by future studies.

Chapter 7

Rational Speech Act Model: Background

As discussed in Chapter 2, dominant pragmatic theories are able to account for the rich variety of non-literal meanings that comprehenders attach to utterances which at face value appear to violate communicative norms. However, these theories face the deficit of not making clear quantitative predictions, as well as largely not accounting for the influence of prior beliefs about the world or speaker on utterance interpretation. They are further, unlike UID, not primarily intended as comprehensive theories of speaker utterance choice, and do not clearly address the influence of utterance cost-based constraints on speaker production, although these may have downstream effects on utterance interpretation. Further, the formalism of these theories primarily provides a taxonomy of inference types or pragmatic principles, rather than clearly specifying a process by which a listener might reason from an utterance to an implied meaning, at least at the computational level (Marr, 1982) – and which information they make use of while doing so.

Theories of utterance choice, such as UID, in contrast, concern themselves solely with the speaker's success in transmitting the truth-conditional meaning of an intended utterance to a comprehender, without much regard for how pragmatic reasoning may influence the message that is in fact received, or for a speaker's potentially variable tendency towards communicating optimally rationally in particular contexts. As shown in Chapter 4, this results in UID not accounting for the potential of redundancy (which is barely, if at all, penalized) to substantially distort the speaker's intended message; and as shown in Chapter 6, it can also result in incorrect predictions for utterance choice preferences at the discourse level (as well as an inability to account for them). As I argue in 4.6.1, a more comprehensive theory of utterance choice must integrate pragmatic reasoning into a formal model of language production. A formal model of pragmatic reasoning and utterance choice has the potential to interface with other computational models of language processing, to yield rich in-

sights about how, when, and to what degree pragmatic reasoning influences language processing.

In the following chapters, I focus primarily on the Rational Speech Act (RSA) model (Frank & Goodman, 2012), a commonly used probabilistic model of utterance production and interpretation, which describes the process by which listeners make inferences under uncertainty. As will be discussed in the rest of this chapter, this framework is extremely flexible[1], and able to represent reasoning on the basis of prior beliefs about the world and speaker, the influence of cost-based considerations on speaker utterance choice, fuzzy semantics, as well as the consequences of the speaker and listener communicating through a noisy channel. A growing body of empirical data supports the use of this model, which has been used to account for phenomena such as common ground inferences, irony and metaphor, inferences due to contrastive prosody, fragment interpretation, hyperbole, and cases of vagueness and ambiguity.

In this and the following two chapters, I discuss the limitations of and extensions to the base RSA model, which, as first demonstrated by Bergen & Goodman (2015), is unable to derive distinct inferences, or inferences of different strengths, given truth-conditionally equivalent utterances. As a result, the base RSA model cannot account for significantly different inference strength in the case of more, or less effortful informationally redundant utterances (see Chapter 8). It is similarly unable to account for the full set of empirical data on referring expression choice, as it predicts that one should never see an effect of referent predictability in the case of unambiguous pronominal reference (see Chapter 9) – which appears to not be the case.

As I demonstrate in Chapter 8, if one makes the assumption that accurate utterance storage and recall is a noisy process similar to that of utterance perception, it is possible to fully account for the intuitive observation that more effortful utterances generate stronger inferences (cf. Wilson & Sperber, 2004). I ultimately argue that the assumption of a clean channel (whether with respect to perception, or to memory) in the base RSA model is problematic, and that the effects of communicating through a noisy channel on pragmatic reasoning are both far-reaching, and easy to miss. The assumption of a noisy channel should therefore be integrated into the basic Rational Speech Act model toolkit. In Chapter 9, I show that the RSA model, similarly incorporating the assumption of a noisy channel, shows unique promise in accounting for the current set of empirical data on predictability and referring expression choice.

7.1 Literature Review

Since roughly the mid-90s, there have been multiple attempts at creating formal models of pragmatic reasoning, using a mathematical or explicitly probabilistic framework.

[1]The flexibility of this model can also be a cause for concern, as a model which can be modified to suit any empirical result has limited predictive and theoretical value. However, as I argue in Section 7.2, the modifications I make use of in this thesis have independent theoretical motivation, and apply to a wide range of phenomena, rather than a limited set of empirical results.

These models have been used to yield empirically testable predictions about utterance interpretation, and in determining which constraints speakers and comprehenders are subject to in utterance production and interpretation. They have also yielded insight into which information speakers and comprehenders must make use of in, respectively, selecting utterances to communicate specific messages, and interpreting which message a given utterance is most likely to be communicating. In addition to the Rational Speech Act model, introduced by Frank & Goodman (2012), many models have made use of a game-theoretic framework (e.g., Parikh, 1991; Benz & Rooij, 2007; Franke, 2009; Jäger, 2012), which I will not discuss further here – although it is possible that a game-theoretic framework could account similarly for the data I present. Most frameworks assume rational agents, which reason about each other recursively, with speaker utterance choices constrained by concerns of efficient information transfer.

Approaches such as the Rational Speech Act model consider the semantic contributions of an utterance separately from the pragmatic inferences it may trigger. The model takes the semantic content as input, and uses other known facts or beliefs about the world, the conversational setting, or the agents themselves (as well as constraints on agents' actions), and computes the most likely intended meaning of the utterance. In the next section, I describe the most basic version of such a model, and illustrate how it operates step-by-step.

7.1.1 Rational Speech Act Model

Here, following up on the discussion in Section 1.3, I will describe the baseline Rational Speech Act model in greater detail. In this model framework, speakers and listeners reason iteratively about each other, given certain starting assumptions, with the overarching goal of the listener arriving at the meaning that is intended by the speaker. Most baseline versions of the model assume, consistent with the core insights of the UID hypothesis, that speakers attempt to balance two periodically conflicting goals. On the one hand, speakers aim to produce those utterances which are maximally likely to get across the intended meaning to the comprehender. On the other hand, they aim to conserve articulatory effort by producing less effortful utterances (typically, so long as the comprehender is still sufficiently likely to infer the intended meaning). Comprehenders are assumed to interpret utterances by imagining, via Bayesian inference, which message the speaker must have intended to transmit (given their goals of accurate yet efficient communication). The process may continue recursively, with comprehenders reasoning about yet more sophisticated speakers, and vice versa, although there is relatively limited empirical evidence for more than one level of recursion (but see Franke & Degen, 2016).

The base RSA model is typically structured as described below. One assumes a particular world state (w), as well as a set of utterance alternatives (u) which may be used to describe that state. For practical purposes, in each case the set of alternative utterances is constrained to those deemed most likely. The goal of the speaker model

is to determine which of the alternative utterances is most likely to be used given a particular world state (or speaker intent, background world knowledge, and so forth). The goal of the listener model, correspondingly, is to determine which world state (or other condition) is most likely to hold, given a particular utterance, in the context of the specified alternative set.

Both the speaker and listener start out with shared prior beliefs about the base likelihood of various world states (denoted by $P(w)$), and a shared knowledge of the cost of each alternative utterance ($C(u)$). The cost determines the base probability of each utterance, independently of its utility, with shorter or more efficient utterances being *a priori* more likely to be produced, all else being equal. Another element of shared knowledge is the α parameter, which denotes the presumed rationality of the speaker. Typically, it is presumed to apply both to the degree to which a speaker attempts to optimize the utility of the utterance, and the degree to which they attempt to minimize cost. I assume, in contrast, that speakers may place different weights on optimization of utility vs. cost, which are fundamentally different concerns, and use a separate */lambda* parameter to denote a speaker's tendency to optimize cost (although using a single parameter can be a useful simplifying assumption).

The Literal Listener

In the basic Rational Speech Act model, the literal listener (P_{L0}, infers the state of the world (w), given the utterance the speaker selected (u), its semantic denotation, and their prior beliefs about the likelihood of various possible world states, via Bayesian inference[2]:

$$P_{L_0}(w \mid u) \propto [u](w) \cdot P(w) \tag{7.1}$$

$[u](w)$, or the semantic denotation of u in the context of w, has a binary True (1) or False (0) value, and denotes whether the utterance is semantically compatible with a given world state, or not.

To illustrate, consider the signaling game first described in Section 2.3.1, where a pragmatic listener must infer which of several objects, of various shapes and colors, a speaker is referring to, given an ambiguous one-word shape or color descriptor. In this game, the speaker may refer to one of three objects, which vary along the shape dimension (squares or circles), and the color dimension (blue or green), but only with a single word describing either the shape or the color. Take, for instance, the scenario in Figure 7.1.

In this context, while "*green*" and "*circle*" identify objects unambiguously, the "*blue*" and "*square*" descriptors are ambiguous. Further, there is no one-word description that will uniquely identify the green square. For simplicity's sake, I assume

[2]However, it should be noted that the literal listener is not typically assumed to be an actual interlocutor, but rather represents the semantic meaning of an utterance and any common ground beliefs, which is used by the pragmatic speaker as a basis for utterance choice.

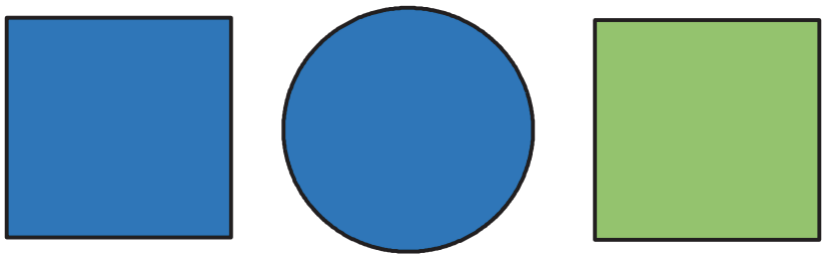

Figure 7.1: Example of an experimental stimulus, from Frank & Goodman (2012).

that all objects are equally likely to be referred to (which gives a uniform prior for $P(w)$). "*blue*" may refer both to the blue circle, and the blue square (i.e., the semantics of the word are equally consistent with both objects), but not the green square. Given a uniform prior likelihood of referring to either, the literal listener assumes that it is equally likely that either the blue circle, or the blue square is being referred to

The Pragmatic Speaker

The pragmatic speaker is a utility-maximizing agent who chooses among a set of alternative utterances, preferentially selecting those that optimally maximize utility (the ability of the literal listener to infer the intended message) while conserving articulatory effort, as denoted by the cost function C – which means that it is not always the utterance with the greatest utility that is most likely to get selected by the speaker. The α parameter denotes how *rational* the speaker is; i.e., to what degree they attempt to maximize utility. As mentioned above, it is frequently assumed that a speaker is identically *rational* in maximizing the utility and cost-efficiency of their utterances – alternately, the α parameter may be applied to utility only. Here, as well as in the models I present in the next two chapters, I assume, in contrast, that speakers do not necessarily give equal weight to *rationality* in the domains of utility and cost-efficiency (I do assume, however, that speakers may variably optimize cost-efficiency). I therefore introduce a λ parameter, which denotes the degree to which a speaker optimizes cost-efficiency.

The cost of a given utterance is presumed to govern the likelihood of said utterance being selected, *a priori*. The cost function is therefore arguably more accurately represented as the probability of the utterance, given the optimizing parameter λ and the cost function C: $P(u; \lambda, C)$. The following two formulas are identical; in later chapters, however, I use the latter:

$$P_{S_1}(u \mid w) \propto \exp(\alpha \cdot \log P_{L_0}(w \mid u) - \lambda \cdot C(u)) = \tag{7.2}$$

$$P_{S_1}(u \mid w) \propto P(u; \lambda, C) \cdot \exp(\alpha \cdot \log P_{L_0}(w \mid u)) \tag{7.3}$$

In our signaling scenario, assuming the speaker wishes to refer to the *blue square*, they may choose either the signifier "*blue*", or the signifier "*square*" to do so; they are

approximately cost-equivalent, and each has equal utility in identifying the object. However, given a scenario where the speaker wishes to refer to the *blue circle*, while they may choose between the signifier "*blue*" and the signifier "*circle*" to do so, the latter has much greater utility in uniquely identifying the object, as it does so unambiguously. Similarly, while "*square*" has only a 50% likelihood of correctly identifying the *green square*, "*green*" identifies it with 100% reliability. As a result, speakers show a significant preference towards using the signifier "*green*" rather than the signifier "*square*".

The Pragmatic Listener

The pragmatic listener, P_{L_1}, infers the state of the world (w) via Bayesian inference, given the utterance (u) chosen by the speaker. The listener considers the relative likelihood that the speaker would have produced this particular utterance to communicate any given message, given the utility and costs of possible utterance alternatives in communicating said message:

$$P_{L_1}(w \mid u) \propto P_{S_1}(u \mid w) \cdot P(w) \tag{7.4}$$

In the case of the signaling scenario, the listener will consider the relative utility of using a particular signifier for any given object, and for any given signifier, will conclude that the object which that signifier has the highest utility in correctly identifying is most likely to be the one the speaker is referring to. For example, the signifiers "*blue*" and "*square*" have a 50% likelihood of correctly identifying the *blue square*, from the speaker's point of view – but the other two objects have other signifiers which identify them *uniquely*. The listener therefore concludes that had the speaker wished to refer to the *blue circle*, or *green square*, they would most likely have used the unique identifiers for those objects. If the speaker is using the terms "*blue*" or "*square*" instead, then the listener concludes that they are most likely referring to the *blue square*.

More Sophisticated Agents

Theoretically, ever-more pragmatically sophisticated speakers (S_n) and listeners (L_n) may continue to recursively reason about each other (e.g., a speaker who reasons about how a pragmatic, rather than a literal listener is likely to interpret an utterance, and a listener who reasons about a speaker who reasons about a pragmatic listener, and so forth). The standard model does not specify that the reasoning must and at the L_1/S_1 level. However, empirical evidence for deeper levels of recursion is limited. A series of studies (Stiller et al., 2011; Vogel et al., 2013; Degen et al., 2012) show little evidence for recursion beyond the first level; comprehenders generally did not arrive at interpretations requiring deeper levels of recursion. On the other hand, Franke &

Degen (2016) provide some evidence from complex reference signaling games that a small number of participants (around 15%) may engage in more complex reasoning.

Bergen & Goodman (2015), which looks at the interpretation of contrastive prosody, assumes deep (up to 10) levels of recursion; however, in this case it is arguably unclear whether this describes online reasoning by speakers and listeners, or perhaps a longer-term process by which contrastive prosody came to more conventionally signal exhaustive interpretations. Similarly, Levy (2018) accounts for the distribution of optional function word use using a clean-channel RSA model with arbitrarily deep levels of recursion and a utility function penalizing peaks and troughs in information density, avoiding some circularity problems within the standard UID model. However, here it is likewise unclear whether the model is an argument for the existence of hyper-sophisticated agents, or whether it could rather represent "intergenerational change in language transmission arising from differential communicative success or acquisition of different utterance variants." In the remaining chapters, I do not assume more than one level of recursion, and find that a single level of recursion is sufficient to derive the given phenomena.

Utterance Costs

Initially, the cost parameter was not included in the utility function. However, evidence has emerged since then that not only producers (as would be predicted by UID), but also listeners are sensitive to utterance cost. Bergen et al. (2012), for example, showed that utterance 'cost' (in virtual dollars) modulated both producer choice of utterance, and the inferences that comprehenders made in light of the costs of alternative utterances. Similarly, Degen et al. (2013) looked at whether increasing production costs (in this case, the amount of time required to type a message) influenced utterance interpretation, as well as decreasing likelihood of production, and found both to be the case. Many models since then have incorporated cost parameters to account for pragmatic phenomena (most notably models incorporating latent threshold variables; e.g. Lassiter & Goodman, 2013, 2017; Qing & Franke, 2014; Schöller & Franke, 2017). At this point, the cost parameter is conventionally included in most models, provided that alternative utterances are not cost-equivalent.

To illustrate, consider the set of scalar alternatives "*some*" and "*all*," where, as discussed in Section 2.2, the use of "*some*" leads to the pragmatic inference that *not all* students passed the test (because, had *all* students passed the test, the speaker would have used the more informative alternative). However, in this scenario, why would speakers frequently choose to say "*some*," rather than "*some, but not all*," which describes the situation unambiguously, and therefore has higher utility in this context (cf. Bergen et al., 2016)? And further, why would the listener so reliably infer *but not all* from "*some*," in this scenario, even when there is an unambiguous alternative? In this case, "*some, but not all*" has a significantly higher cost than either "*some*" or "*all*," although it has higher utility than the former in identifying a situation in

which *all* does not hold. It is therefore less likely to be used to indicate *not all* than "*all*" is to be used to communicate *all*, as well as relatively less likely to be used to indicate *not all* than "*some*." As a result, listeners reason that speakers are relatively less likely to use "*some, but not all*" to communicate *not all*, which leads them to infer *not all* from "*some*."

7.1.2 Applications of the Base RSA Model

The base RSA model was initially successful in accounting for the use and interpretation of ambiguous reference in signaling games, such as that illustrated in Section 7.1.1 above. Frank & Goodman (2012) tested the utility of the RSA model in this context by setting up an online one-shot signaling game, where participants initially indicated which object they thought was most salient (which corresponds to $P(w)$ above), on the assumption that salience influences the prior likelihood of referring to any particular object. Participants were then instructed to act as either speaker or listener in a signaling game. As a speaker, they indicated which one-word ambiguous reference they would use to refer to a particular object, and as a listener, they indicated which object they thought a given ambiguous reference referred to. The results of this experiment showed a tight fit to those predicted by the base RSA model, at both the speaker and listener level. Critically, the base RSA model can similarly account for the classic "some but not all" Quantity implicature (cf. Goodman & Stuhlmüller, 2013).

7.1.3 Joint Reasoning in the RSA Framework

The base RSA model represents the process of reasoning about intended messages, and utterance utility, on the basis of the semantic meanings, utilities, and costs of alternative utterances. However, one of the greatest strengths of the RSA framework is that it can easily also represent pragmatic reasoning on the basis of common ground beliefs about the world, uncertainty about the speaker's rationality and the question under discussion, and uncertainty about the speaker's knowledge state (for example). This is easily represented through the concept of *joint reasoning*.

In joint reasoning RSA models, listeners jointly reason about the likelihood of an utterance communicating a particular world state, as well as the likelihood of it corresponding to a particular speaker goal, background world knowledge, or belief state. Speakers select utterances in accordance with their goals and knowledge, taking into account similarly the likelihood of particular utterances communicating the intended world state, goal, or other non-semantic information to the listener. The joint reasoning model requires the addition of at least one parameter, which I here term g, here using the idea of a higher-level speaker *goal* as an example.

First, it should be noted that there is some variation between joint reasoning models in whether they consider the information represented by the additional parameter

to be something that the literal listener reasons about. In many cases, it can be argued that inferences about speaker goals, background knowledge, and so forth, are fundamentally *pragmatic* inferences, and should not be represented at the literal listener level as something the literal listener *infers* (cf. Degen & Tanenhaus, 2015). In these chapters, I assume that the literal listener does *not* reason about anything that is not a part of the semantic meaning of the utterance, but otherwise do not address this discrepancy.

In a basic joint reasoning model, then, the literal listener first infers the world state, given the utterance and potential speaker *goals* (for example):

$$P_{L_0}(w \mid u, g) \propto [\![u]\!](w) \cdot P(w \mid g) \tag{7.5}$$

The pragmatic speaker then chooses the utterance that best communicates the intended world state, given their goal in communicating this information[3]:

$$P_{S_1}(u \mid w, g) \propto \exp(\alpha(\log P_{L_0}(w \mid u, g) - C(u))) \tag{7.6}$$

Finally, the pragmatic listener jointly reasons about the likely world state and speaker goal, given the utterance. In other words, they consider which world state and speaker goal is most likely to hold, given the utterance that was chosen by the speaker:

$$P_{L_1}(w, g \mid u) \propto P_{S_1}(u \mid w, g) \cdot P(g) \cdot P(w) \tag{7.7}$$

As mentioned above, the speaker *goal* here may equally well represent any aspect of or factor influencing speaker knowledge and utterance choices – whether it be the speaker's beliefs about the background world state, the speaker's intended conversation topic, the possible meanings they attach to words, their affective state, and so forth. As it is inarguable that all of these factors influence intended utterance meaning, as well as utterance interpretation, the joint reasoning RSA model is in many cases a much more realistic representation of the process by which listeners attach pragmatic meaning to utterances (although in many cases, the base RSA model is sufficient to describe it). In Chapter 8, I further argue that an accurate and descriptively valid representation of this process must also incorporate the fact that signal transmission and encoding occurs in a noisy channel – extrapolating from the level of perception (as described in Bergen & Goodman, 2015) to the higher-level process of memory storage and recall.

[3]To note, other representations of pragmatic speaker reasoning use slightly different conventions, but are mathematically equivalent.

7.2 Applications of the Joint Reasoning RSA Model

The joint reasoning RSA model has shown wide descriptive capacity, capturing phenomena such as hyperbole, sarcasm, and metaphor – as well as 'fuzzy' semantic interpretation, and communication through a noisy channel. In this section, I provide an overview of the inference types that have to date been described by joint reasoning models.

For instance, Goodman & Stuhlmüller (2013) found that varying levels of speaker uncertainty affected listener judgments about the meaning of "some." In their model, speakers and listeners reason about utterances and likely world states on the basis of their shared knowledge of the speaker's *knowledge access* – meaning, whether the speaker's knowledge of the situation is partial, or complete. For instance, a speaker may know that two of three apples in a bag are red, but be unaware of what the color of the third apple is. In this context, "*some*" is relatively more likely to be interpreted as *at least some, and possibly all*, given that the speaker lacks the knowledge to say whether *all* of the apples are red, or not.

Hyperbolic utterances deliberately exaggerate the qualities of various states, and are intended more to express the speaker's *attitude* towards those states, than to describe the states accurately. Kao et al. (2014b), for instance, looked at utterances such as *"The electric kettle cost $1,000"*. A listener *may* interpret such an utterance as describing the amount the kettle, in fact, cost – but in a typical context, should be much more likely to interpret the stated price as a deliberate exaggeration, which has the purpose of pointing out that the speaker considers the kettle very expensive (with the kettle, in reality, most likely costing far less money). In this case, the listener reasons jointly about the most likely price of the kettle (given their prior knowledge of the distribution of kettle prices), as well as the speaker's intent in communicating the price (either a straightforward factual statement of the kettle's price, or a statement expressing the speaker's *attitude* towards the price), given the utterance at hand. The joint reasoning model represents the intuition that the more implausibly high the price given is, the more likely it is that the statement is intended to communicate the speaker's *attitude* towards the price, rather than the price itself.

Irony, similarly, involves sarcastic statements such as *"The weather is amazing"* (in the context of demonstrably bad weather). Kao & Goodman (2015) model the interpretation of such statements by assuming that listeners are attempting to jointly infer the actual state of the weather, as well as the speaker's *attitude* towards the weather and the resulting emotional *arousal*, given their prior beliefs about the likely current weather status. For instance, if the pragmatic listener has a high expectation that the weather is *good*, and they hear the speaker describe the weather as "terrible," they are more likely to believe that the speaker is communicating a high level of emotional *arousal* due to the (excellent) weather, rather than simply describing the weather state. This model improves upon the aforementioned model of hyperbole by considering the possibility that the speaker may not only be attempting to commu-

nicate their emotional state to the listener, but may additionally be attempting to communicate the *strength* of their emotional arousal.

Kao et al. (2014a) accounts for the interpretation of metaphor, such as *"John is a shark"*. In this case, the listener jointly infers *John's* status of being either a shark or a human, as well as the qualities that *John* is likely to possess (such as ruthlessness), given the more likely speaker goal. The speaker, in this case, may be attempting to respond to vague questions regarding what John is *like*, or specific questions regarding whether *John* possesses certain qualities. The model accurately predicts that when the goal is to answer vague questions, the listener is more likely to interpret the statement as literal (i.e., *John* is in fact a shark), and when the goal is to address *John*'s specific qualities, the listener is more likely to interpret the statement as metaphorically describing *John*'s shark-like qualities. In a typical context, the listener's expectations are that *John* is, in fact, a person. They are therefore more likely to determine that the speaker is attempting to communicate that *John* shares certain qualities with sharks.

Joint reasoning models have also been used to account for the interpretation of certain labels, such as *expensive* and *tall*, which can only be interpreted in reference to a comparison class (Lassiter & Goodman, 2013). In this case, the listener reasons jointly about the class-specific threshold (e.g. 180cm, or $50) that must hold for a class member (e.g., a human male, or an electric kettle) to be referred to as *tall*, or *expensive*, considering the typical distributions of height or prices (for example) for the given class, and the speaker's desire that the term still be *informative* (if fuzzy). Scontras & Goodman (2017) use a similar threshold-based joint reasoning model to account for which gradable adjectives permit collective interpretations (e.g., '*the boxes [jointly] are tall*'), vs. those that do not (e.g., '*the boxes [jointly] are big*').

Yet another application of the joint reasoning model considers the possibility that word meanings are not fully fixed, and that some degree of semantic vagueness exists (Bergen et al., 2016). In this case, the pragmatic listener must derive M-implicatures – where the listener interprets statements worded in an *unusual* manner (but semantically equivalent to statements worded typically) as indicating that a *non-prototypical* state is in fact being described. Bergen et al. (2016) prove, furthermore, that the base joint reasoning model as described above, which considers semantic meaning fixed, principally cannot derive different inferences, or inferences of different strength, for utterances with the same semantic meaning, as discussed further in Section 7.2.3. In this particular case, the conundrum is fixed by allowing utterance meaning to be *fuzzy*, or a distribution of varyingly likely world states.

Most relevant to the following chapters, however, are those joint reasoning models which represent the influence of background world states (unknown to the listener, who is only aware of their prior likelihood), and models which account for the observation that speakers and listeners communicate through a noisy channel, with potential for signal distortion (which must be accounted for by both speakers and listeners).

7.2.1 Reasoning about the Background World State

Degen et al. (2015) looks at whether, when faced with an implied utterance meaning that is inconsistent with a comprehender's beliefs about the world, the comprehender resolves the conflict by revising those beliefs. Take for instance the following example, where the literal meaning of 1 and 2 is inconsistent with a naive comprehender's presumed beliefs about the world - which is that marbles will generally sink when thrown into a pool:

(1) John threw the marbles into the pool.

(2) Some of the marbles sank.

Indeed they find that comprehenders revise their background knowledge, coming to the conclusion that in the particular context under discussion, marbles do *not* invariably sink into pools. As Degen et al. (2015) demonstrate, the base RSA model is unable to account for this variety of joint reasoning over world states and unstated assumptions about the world under question. They therefore propose a joint reasoning model which in fact cleanly accounts for the observations in question, with s denoting the world state, u denoting the utterance, and w denoting world "wonkiness":

$$P_{L_0}(s \mid u, w) \propto [\![u]\!](s) \cdot P(s \mid w) \tag{7.8}$$

$$P_{S_1}(u \mid s, w) \propto \exp(\lambda \ln P_{L_0}(s \mid u, w)) \tag{7.9}$$

$$P_{L_1}(s, w \mid u) \propto P_{S_1}(u \mid s, w) \cdot P(s \mid w) \cdot P(w) \tag{7.10}$$

7.2.2 Communicating through a Noisy Channel

Bergen & Goodman (2015) look at whether joint reasoning models can account for the use and understanding of sentence fragments on the one hand, and for the pragmatic interpretation of prosodic emphasis on the other, as in the following example, with capital letters indicating prosodic emphasis:

(3) Who went to the movies?

(4) BOB went to the movies.

I will focus only on how interpretation of prosodic emphasis was modeled here. What Bergen & Goodman (2015) attempt to account for is how the second utterance is interpreted exhaustively - meaning that *Bob*, and *only* Bob went to the movies. What they propose is that as speakers and listeners are communicating through a noisy channel, additional prosodic emphasis is an intentional attempt by the speaker to reduce the capacity of noise to distort their utterance. This intentional action can result in the listener making a pragmatic inference - in this case, that the speaker has exhaustive knowledge of who went to the movies.

In the end, Bergen & Goodman (2015) demonstrate that to account for the inferences that listeners draw, one requires joint reasoning over the utterance meaning, as well as the likelihood of a given perceived utterance being the intended utterance, taking into account noise in the communication channel. The noisy channel model they propose is discussed in more detail in the following chapter, where I adopt it to account for the pragmatic interpretation of informationally redundant utterances.

7.2.3 Failure of Standard RSA Models to Derive Distinct Inferences under Semantic Equivalency

As pointed out in Section 3.2.3, it is intuitively obvious that if a speaker draws more attention to an utterance which has the potential to trigger a pragmatic inference, or which apparently violates some communicative norm, then the comprehender will be more likely to draw the inference, or to draw a stronger inference than they would otherwise. This is likely due to several factors. First, the comprehender must notice and accurately perceive an utterance in the first place, in order to draw any inferences from it. Further, the comprehender must be *motivated* to process the utterance at a relatively deep, rather than surface level – which intuitively is more likely to take place if the speaker has implicitly signaled, through increased effort, that the utterance is worth paying attention to. Finally, and connected to the previous point, the more *emphasis* the speaker puts on a given utterance, the more likely the comprehender is to conclude that there must have been a *specific reason* for the utterance deserving special emphasis – and the more likely they will be to infer the likely reason.

It is apparent then that the more articulatory effort is expended on producing an utterance, or the more words are used to express a particular point, the more likely the utterance is to draw a comprehender's attention – both due to increased perceptual prominence, and to the comprehender becoming more motivated to account for *why* the speaker would expend more effort than strictly necessary. One would intuitively expect that standard RSA models, which incorporate both the influence of utterance cost of utterance choice, and comprehenders' reasoning about the reasons for any given utterance choice, would be able to derive this effect. However, Bergen & Goodman (2015) and Bergen et al. (2016) provide a mathematical proof that they are not, due partly to the fact that they assume a clean communication channel, where every utterance is accurately and faithfully perceived by the listener.

Take, for example, the contrast in utterance strength demonstrated in Chapter 4:

(5) John went shopping. He paid the cashier. -> *weak habituality inference*

(6) John went shopping. He paid the cashier! -> *stronger habituality inference than 5*

(7) John went shopping. Oh yeah, and he paid the cashier. -> *stronger habituality inference than 5*

What was found is that the more perceptually prominent an utterance is (e.g., examples 6/7 vs. 5), the stronger an inference it triggers, on average. As I demonstrate in Chapter 8, neither the base RSA model, nor the standard joint reasoning model, is able to capture this.

As pointed out in Bergen et al. (2016), within a standard RSA framework, a more costly or perceptually salient utterance will never be of any advantage to the literal listener, as long as it is truth-conditionally equivalent to the less costly or perceptually salient utterance – both, in this case, communicate exactly the same information to the literal listener. As a result the more costly utterance never presents any advantage to the speaker, in terms of communicating their desired message, given that it does not present any advantage to the listener in comprehending it. The speaker's utterance choice, therefore, is in this case based solely on relative cost. This is at odds with intuition, as one would expect a speaker to use a more effortful, or perceptually salient utterance, to communicate less predictable meanings (as predicted by UID; see Section 2.1).

The downstream effect of this is that pragmatic listeners are unable to infer that more effortful or perceptually prominent utterances are more likely to communicate an unpredictable, or atypical meaning. This is due to the fact that the *atypicality* of the meaning in fact plays no role in the speaker's choice of utterance (rather, only cost does). As a result, all utterances which are truth-conditionally equivalent will generate the same inferences, and further, only inferences of exactly the same strength. The model proposed by Bergen & Goodman (2015) accounts for pragmatic inferences triggered by contrastive prosody by considering that utterances with increased prosodic emphasis are more likely to be perceived accurately, as described in Section 7.2.2. However, the question of accurate *perception* is more likely to apply at a relatively low level – e.g., mistaking one short word for another. It is relatively unlikely, in contrast, that a comprehender would fail to (sufficiently) accurately perceive a multi-word utterance, such as that in Example 1.

In the following section, I present a solution to this. The noisy channel model introduced by Bergen & Goodman (2015) can capture two crucial observations. On the one hand, speakers are more likely to use more costly and perceptually prominent methods of expressing themselves when wishing to communicate an unusual, unexpected, or particularly important message. On the other hand, comprehenders are more likely to interpret more costly and perceptually prominent utterances as expressing unusual, unexpected, or particularly important. I argue that this model can be expanded beyond the level of perception, to account for the inherent noisiness of the encoding and recall processes (cf. Bower et al., 1979).

7.3 Model Background

Here I give a brief account, expanded upon in the following two chapters, of how the base RSA model, the joint reasoning model, and the noisy channel model, can account for the *habituality* inferences described in Chapters 3-4; as well as how they can account for the somewhat confusing and contradictory data on whether referent predictability influences referring expression use.

7.3.1 Habituality Inferences

To recall, *habituality* inferences are cases in which a comprehender interprets an otherwise redundant description of a typically predictable activity as signifying that the activity is not, in fact, predictable in context. The following utterances, in a *typical* context, all generate *habituality* inferences of varying strengths:

(8) John went shopping. He paid the cashier. -> (*John* doesn't typically pay the cashier [weak])

(9) John went shopping. He paid the cashier! -> (*John* doesn't typically pay the cashier [relatively strong])

(10) John went shopping. Oh yeah, and he paid the cashier. -> (*John* doesn't typically pay the cashier [relatively strong])

Overall, the explicit description of the highly predictable activity is incompatible with the idea of an efficient and rational (in the information-theoretic sense; see Section 4.6.1) speaker. To reconcile this, comprehenders make an inference that the activity must in fact not be as predictable as they would have otherwise assumed. This, in itself, is fairly straightforwardly accounted for by a basic joint reasoning model, very similar to that in Degen et al. (2015) (see Section 7.2.1). In this model, described in Section 8.4, listeners reason jointly about the world state (whether the cashier was paid, or not) and the activity's *habituality*, and conclude that a *low* habituality, even if at odds with their initial assumptions, is the more likely circumstance given the explicit activity mention.

However, as I show in Section 8.4, truth-conditionally equivalent utterances cannot generate inferences of different strengths given a standard joint reasoning model (as discussed above, in Section 7.2.3). Here, I argue that incorporating a notion of the *noisy channel*, as first introduced in Bergen & Goodman (2015), generates stronger inferences for more attentionally prominent utterances, and reflects the intuition that speakers should use particularly prominent utterances to communicate unusual meanings. However, here I extend the notion of the *noisy channel* beyond the level of *signal processing* (i.e., whether the signal was accurately perceived), to the level of *message encoding* (i.e., whether the message communicated by the signal was accurately encoded in memory, and/or made available for recall).

Bower et al. (1979) shows specifically that upon reading stories describing stereotyped activities, participants are often subsequently unable to recall whether particular steps in said sequences were explicitly mentioned, or not. A reasonable assumption to make is that more effortful and perceptually prominent utterances are more likely to grab the comprehender's attention, and consequently are more likely to be processed on a deeper level, as well as to be encoded and recalled accurately. A model attaching a higher likelihood to less (vs. more) attentionally prominent utterances being misremembered as 'nothing' is highly psychologically plausible, and as I show in Section 8.5, performs well in representing both hypothesized speaker utterance preferences, and in representing the increased strength of inferences due to increased attentional prominence [cf. Wilson2004].

7.3.2 Choice of Referring Expressions

In the case of referring expressions, I tackle the rather puzzling pattern of results I describe in Section 6.5, regarding the question of whether referent predictability affects referring expression choice. So far, effects of predictability on referring expression choice have been seen in some paradigms, but not others, and appears to be modulated partly by task and prompt design, as discussed in Section 5.2.4. Although it makes sense intuitively that more interactive and naturalistic experimental designs would be more successful, perhaps prompting a greater degree of audience design on the part of the speaker (cf. Brown & Dell (1987) vs. Lockridge & Brennan (2002)), some of the other patterns described are either more difficult to account for, or have some intuitive but not formal support. The picture is further complicated by my results not supporting Bott et al. (2018)'s hypothesis that results can be consistently detected as long as prompt antecedents are same gender (and any pronouns used therefore ambiguous), and the task is using a constrained continuation paradigm.

This given, I argue in Chapter 9 that the RSA model may be uniquely suited to account for the pattern of results so far detected and hypothesized. First, it straightforwardly predicts that pronoun referential ambiguity (i.e., same-gender prompts in passage completion paradigms) would increase the magnitude of any effect of referent predictability on referring expression production, in line with Bott et al. (2018)'s hypothesis and account of the existing set of results. This does not account, however, for the fact that some experimental paradigms *do* show an effect, even if smaller, even when prompts include opposite-gender antecedents, and pronoun use is entirely unambiguous (Rosa & Arnold, 2017). This, however, is accounted for if the notion of a noisy channel is incorporated into the model – which is quite reasonable, as third-person pronouns in English are minimal pairs, and easily confused for each other in a noisy environment.

Secondly, the RSA model does provide an explanation for why constrained-completion tasks may detect an effect of referent predictability on referring expression choice, while free-completion tasks do not. Here, I assume, as discussed in Section 9.3, that

while producers will only refer, when instructed to freely write *the most likely continuation*, to referents that meet some arbitrary criterion of predictability - e.g., 0.40 likelihood of mention. However, when participants are constrained in which referent they refer to next, they are in effect forced to refer even to those referents that they consider highly unpredictable, given the prompt and verb bias. The wider the variation in the predictability of referents referred to in the continuation, the more likely are the effects of (un)predictability on referring expression choice to pan out. In Section 9.3, I demonstrate the predicted increased effect of predictability on referring expression choice in forced-completion tasks.

Finally, and somewhat less straightforwardly, one must account for why many paradigms and task designs yield no effects of predictability on referring expression choice (including my results, in Section 6.4.2, which are null despite using a forced-completion paradigm and same-gender antecedents in prompts). One possibility is that as the rest of the sentence typically disambiguates the intended referent, pronouns are never *truly* ambiguous, and therefore per the model described in Section 9.2, should be chosen based on cost and speaker grammatical biases, but not utility to the listener (which does not change depending on whether a noun or pronoun is used to refer back to the referent).

This, again, however, does not account for the results obtained by Rosa & Arnold (2017) and Bott et al. (2018). In the case of Bott et al. (2018), it may simply be the case that the referring expression choice in German is more flexible (for example) – but accounting for differences between English and German is beyond the scope of this thesis. Another possibility is that increased interactivity and naturalness of the paradigm – by creating a highly interactive environment, as in Rosa & Arnold (2017) – prompts speakers to increase the utility of their utterances (computationally, perhaps corresponding to modulation of the α parameter). This, however, would require far more empirical work to confirm.

7.4 Summary

In summary, as I demonstrate in the following two chapters, the Rational Speech Act model presents great potential in accounting for utterance choice at the discourse level, particularly once one incorporates the assumption of a *noisy channel*. In the first case study I present – that of *informationally redundant utterances* – the RSA model clearly demonstrates the capacity of informational redundancy to alter the message received by the listener, even if it does not substantively alter the semantic meaning of the utterance.

The Uniform Information Density hypothesis, due to concerning itself solely with signal perception and decoding, rather than the message ultimately transmitted to the listener, wrongly assumes that there is no communicatively motivated constraint on redundancy in speech, beyond the speaker's desire to be maximally concise and

conserve articulatory energy. However, as I demonstrate, the RSA model is similarly, and rather unintuitively, unable to account for the generation of stronger pragmatic inferences in the case of more attentionally prominent utterances, or a speaker's preference to use more prominent utterances for more unusual meanings. Adding the assumption of a *noisy channel* to the RSA model allows it to account for all empirical phenomena predicted and observed, and gives it better explanatory value than either UID or the base RSA models, alone.

In the second case study I present, I show that in many to most cases, UID appears to wrongly predict that referent predictability influences referring expression choice. While there is some difficulty in accounting for this in both the UID and RSA frameworks, the RSA model, unlike UID, predicts some of the patterns in results that have been observed to date, and accounts neatly for those circumstances in which effects of referent predictability on referring expression choice appear to be more likely to arise, due to taking into account not only the effect of predictability on production, but also the effect of utterance *utility* in communicating the truth-conditional utterance meaning. Due to these and other properties of the RSA model discussed in the following chapters, it appears far more suited than UID to represent utterance choice and interpretation at the discourse level.

Chapter 8

Rational Speech Act Model: Informationally Redundant Utterances

In this chapter, I present a joint reasoning RSA model which accounts for the *habituality inferences* presented in Chapter 3. I first show that the base RSA model is trivially unable to account for these inferences. I then demonstrate that the standard joint reasoning model derives these inferences, but is principally unable to derive stronger inferences for more effortful or attentionally prominent utterances. I finally show that integrating the noisy channel machinery first introduced in Bergen & Goodman (2015) into this joint reasoning model, and extrapolating from noisy perception to the possibility of noisy and attention-dependent information processing and retrieval, generates stronger inferences for more attentionally prominent utterances. This last model then accomplishes the following:

1. It generates habituality inferences from a speaker's use of informationally redundant utterances, as well as stronger inferences for more ostensive stimuli (Wilson & Sperber, 2004).

2. It integrates the intuition behind UID – that of communication through a noisy channel – into a probabilistic pragmatic model, increasing its psychological plausibility, and enabling it to make accurate predictions across a range of phenomena predicted by UID.

3. It goes beyond UID in demonstrating the consequences of speakers being redundant – rather than this simply resulting in a loss of efficiency, it results in a potential distortion of the speaker's intended message. A model which does not consider utterance interpretation, beyond the decoding of an utterance's

truth-conditional meaning, is unable to straightforwardly account for this (or to make strong claims about the putative success of message transmission).

Further, of specific interest to the field of experimental pragmatics, and modeling of pragmatic phenomena, it demonstrates a further instance of listeners reasoning about the background world knowledge of the speaker, and altering their beliefs about the common ground, in order to accommodate otherwise pragmatically anomalous utterances (cf. Degen et al., 2015, discussed in Section 7.2.1).

In this chapter, I consider utterances such as the following, introduced in Chapter 3:

(1) "John went shopping. He paid the cashier!"

(2) "John went shopping. Oh yeah, and he paid the cashier."

(3) "John went shopping. He paid the cashier."

(4) "John went shopping."

In Examples 1-3, assuming a typical common ground, stating explicitly that *John* paid the cashier is informationally redundant – *cashier-paying* is a very predictable activity in context, and should automatically be inferred simply given the mention of shopping (Bower et al., 1979). The predicted, and, in the case of points 1 and 2, empirically validated (see Chapter 4) effects associated with the use and comprehension of such utterances are:

1. As utterances 1-3 are informationally redundant, at face value they are pragmatically odd. Comprehenders resolve this pragmatic anomaly in part by determining that *cashier-paying* is not, in fact, typical for this individual and in this context, contrary to their prior beliefs.

2. Expending more effort on communicating an informationally redundant utterance, for example by using exclamatory prosody, should strengthen the inference. Increased articulatory effort (and increased attempts at grabbing the listener's attention) reflect greater speaker intent to transmit precisely this message to the listener, and should increase the listener's likelihood of noting, processing, and accurately recalling the message.

3. Speakers should preferentially use more attentionally prominent utterances to transmit particularly unusual or unexpected meanings, even when doing so is relatively costly. This is fairly straightforwardly predicted by UID, at least at lower levels of production (Jaeger, 2010; Jaeger & Buz, 2017).

In the following sections, I first briefly discuss the empirical *habituality* priors fed into the models which follow. I then discuss the basic setup for the models which follow, focusing on the states represented by each parameter. I follow by showing,

separately, why the base RSA and standard joint reasoning models cannot account for these inferences, and their respective strengths – principally in the former case, and somewhat unintuitively in the latter. I finally demonstrate that a noisy channel joint reasoning RSA model is able to derive all three effects described above, and similarly demonstrate that this model performs well in accounting for the interpretation of the *non*-redundant utterances described in Chapter 4.

8.1 Empirical Habituality Priors

Here, I discuss how the prior beliefs regarding the habituality of various activities are represented, and incorporated into the following models. *Prior* beliefs regarding the base likelihood of various activities occurring were collected empirically in the course of Experiments 1-3, presented in Chapter 4. This was done by asking comprehenders to rate, on a scale of 0 (never) to 100 (always), how often they thought someone engaged in a particular activity, in the context of a specific event sequence (such as *grocery shopping*), which, by common knowledge, habitually includes said activity:

(5) "How often do you think John usually pays the cashier, when grocery shopping?"

This question was asked after presenting comprehenders with either a neutral context mentioning a certain event sequence, or a "wonky" context mentioning said sequence. The participant at this point did *not* see utterances 1-4. The "wonky" context either hinted strongly, or explicitly stated, that the individual in question did not habitually engage in the normally-habitual activity. An example can be seen below:

(6) **neutral**: "John often goes to the grocery store around the corner from his apartment."

(7) **wonky**: "John is typically broke, and doesn't usually pay when he goes to the grocery store."

Additionally, as a control, and to use as a comparison to the "wonky" condition, above, I also collected ratings for *non-habitual* activities, which are consistent with the event sequence, but not necessarily *expected*: for example, buying apples when grocery shopping:

(8) "How often do you think John usually gets apples, when grocery shopping?"

For the rest of this chapter, I will only look at the "neutral context - habitual activity" condition – i.e., the only one where the activity description is informationally redundant:

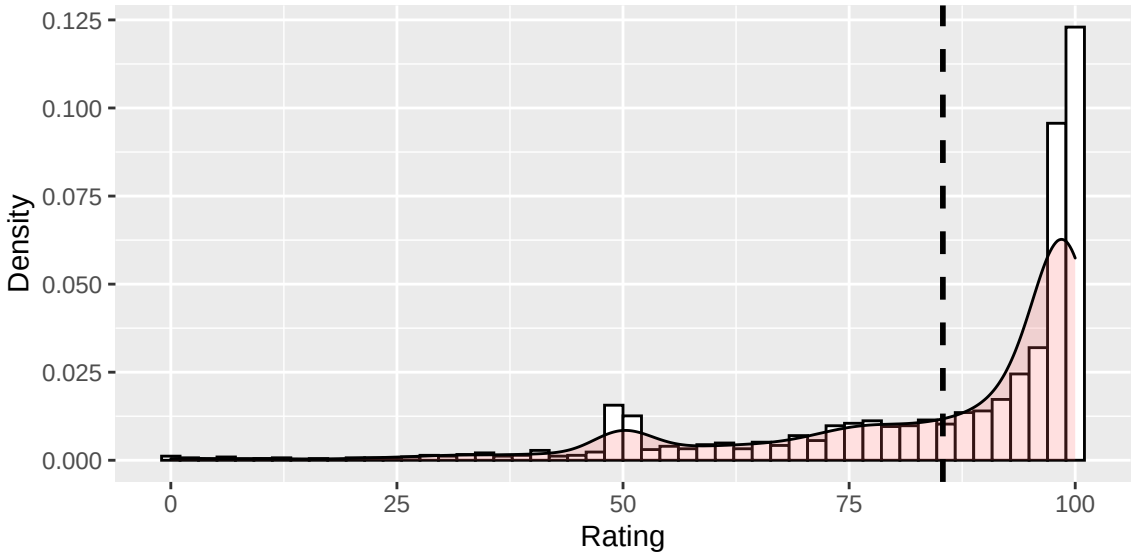

Figure 8.1: Distribution of prior ratings collected from participants.

(9) **Context**: "John often goes to the grocery store around the corner from his apartment."

(10) **Question**: "How often do you think John usually pays the cashier, when grocery shopping?"

In Figure 8.1, I plot the distribution of ratings collected from participants. Here, it is evident that given a neutral or typical context, the vast majority of comprehenders believe that (e.g.) *John* is a *habitual* cashier-payer.

In order to incorporate the empirical priors into the models, I fit beta distributions to each condition, using the `fitdistrplus` `R` package (Delignette-Muller et al., 2020). A beta distribution is appropriate here, as I am modeling a distribution of probabilities bounded on the interval (0, 1). Although the fitted distributions do not reflect the increase in ratings at the 50-point mark, I do not necessarily consider this to be a problem, given that a bias towards rating towards the middle of a scale is a well-known effect (Stevens, 1971)) that should not be assumed to reflect psychological reality.

8.2 Model Setup

Here, I briefly describe the possible states and distributions associated with the primary parameters used in the three models introduced in this chapter

Possible utterances (u), roughly ordered by increased effort, include the following:

- "(. . .)" = *no/zero/null utterance*
- "John paid the cashier." = *period*
- "John paid the cashier!" = *exclamation*

- "Oh yeah, and John paid the cashier." = *oh yeah, and...*

Possible world states (s), with respect to the activity in question include the following:

- Activity happened on the given instance of grocery shopping (for example) in question (i.e., *John* paid the cashier this time).
- Activity didn't happen on the given instance in question (i.e., *John* didn't pay the cashier this time).

Possible habitualities (h) are sampled from a beta distribution of world state probabilities. These may range from 0 to 1 (never to always), and denote the expected likelihood that the given activity will occur on any particular instance. For example, if the sampled probability is 0.9, then the likelihood of an activity such as *paying the cashier* taking place at any potential instance of *grocery shopping* is 0.9. The higher-skewed the beta distribution, the more *habitual* the activity.

Model parameters that cannot be directly estimated from the empirical data I have at hand, in contrast to the beta distributions that can be fit to empirical distributions of habituality estimates, include the α and λ parameters, as well as the confusion matrix for noisy perception that is proposed in Section 8.5. Here I briefly describe how I estimate α and λ. At their heart, both parameters, which represent speaker optimality/rationality with respect to utterance utility in the case of α, and the degree to which a speaker attempts to optimize utterance cost in the case of λ, are somewhat arbitrary, and can be set arbitrarily to demonstrate the effects of increasing or decreasing degrees of 'optimal' speaker behavior.

In the case of the base RSA model below, the model provides no mechanism for measuring a predicted habituality distribution on the part of listeners - for this reason, the model is principally inadequate. Lacking an empirical distribution that can be used to estimate optimal parameter values, I set these to match those of the joint reasoning hRSA model, which does output a predicted habituality distribution which can be compared to the empirical distribution. In the case of the joint reasoning hRSA model, I use the empirical distribution of (updated) habituality estimates to select the optimal values for the α and λ parameters, by performing a grid search for values producing minimum Kullback–Leibler divergence between the predicted and empirical distributions. In the case of the noisy channel hRSA model, as a parameter grid search would also need to incorporate the noisy channel parameters (which cannot be set empirically), I stick with the α and λ values selected for the hRSA model.

8.3 The Base RSA Model

The baseline RSA model is inherently unequipped to model changes in beliefs about activity *habituality* (h) that are independent of the current activity state (s):

$$P_{L_0}(s \mid u) \propto [\![u]\!](s) \cdot P(s) \tag{8.1}$$

$$P_{S_1}(u \mid s; \alpha, \lambda, C) \propto P(u; \lambda, C) \exp(\alpha \log P_{L_0}(s \mid u)) \tag{8.2}$$

$$P_{L_1}(s \mid u) \propto P_{S_1}(u \mid s; \alpha, \lambda, C) \cdot P(s) \tag{8.3}$$

Given that the literal meaning ($[\![u]\!]$) of *paid the cashier* does not directly communicate anything about activity *habituality*, the standard RSA model can predict only that the cashier was quite definitely paid in this case, given one of utterances 1-3, and that they *may or may not* have been paid in the case of 4. Reasoning about activity *habituality* in itself cannot be represented in the standard RSA model, since the meaning of all utterances is at face value equally consistent with all possible habitualities.

8.3.1 Model Predictions

Literal Listener

In Figure 8.2, it can clearly be seen that after encountering a *null* 'utterance' ("(. . .)"), *literal listeners* preferentially conclude that the activity occurred (although there is some uncertainty).

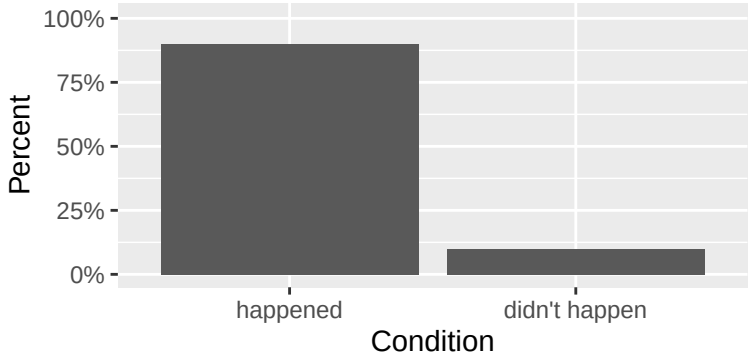

Figure 8.2: RSA literal listener input: "(. . .)" (probability that event happened: 0.9; probability that it didn't happen: 0.1).

Overt utterances are uniformly consistent only with the interpretation that the activity in question *happened*; see Figures 8.3-8.5.

Pragmatic Speaker

As expected, if the activity *happened*, then speakers preferentially say nothing, and only rarely use high-effort utterances; see Figures 8.6 and 8.7.

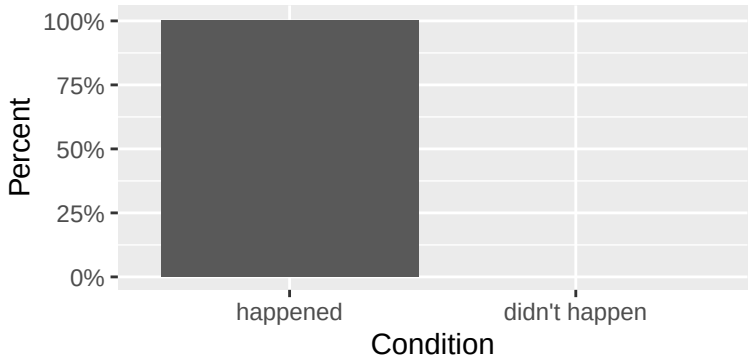

Figure 8.3: RSA literal listener input: "John paid the cashier." (probability that event happened: 1; probability that it didn't happen: 0).

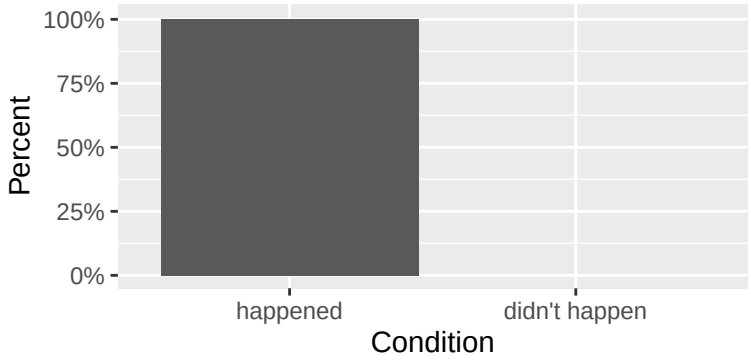

Figure 8.4: RSA literal listener input: "John paid the cashier!" (probability that event happened: 1; probability that it didn't happen: 0).

Pragmatic Listener

As expected, pragmatic listeners infer that if an activity went unmentioned, then it is slightly more likely, compared to baseline, to *not* have happened, given that the speaker has multiple viable alternatives for communicating unambiguously that it *did* happen. However, they still overwhelmingly conclude that it is far more likely that the activity occurred, than that it did not; see Figures 8.8-8.11.

8.3.2 Model Summary

As demonstrated, given that the literal meaning of *paid the cashier* communicates nothing about activity habituality directly, base RSA models accurately predict only that the *cashier* was definitely paid in the case of utterances 1-3, and that they may or may not have been paid given their prior beliefs about the habituality of *cashier-paying*, in the case of utterance 4. Activity *habituality*, in itself, cannot be modeled, since all utterances are at face value equally consistent with all possible habitualities.

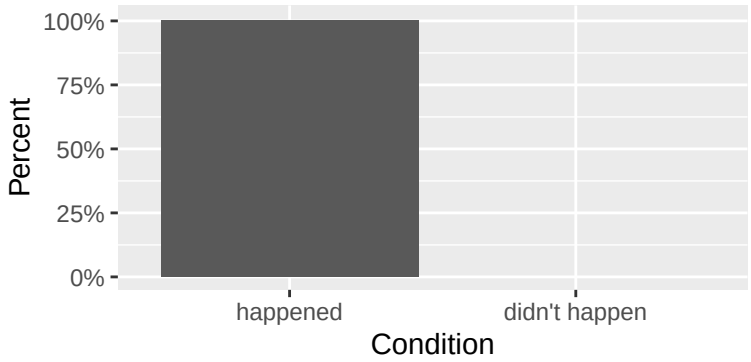

Figure 8.5: RSA literal listener input: "Oh yeah, and John paid the cashier." (probability that event happened: 1; probability that it didn't happen: 0).

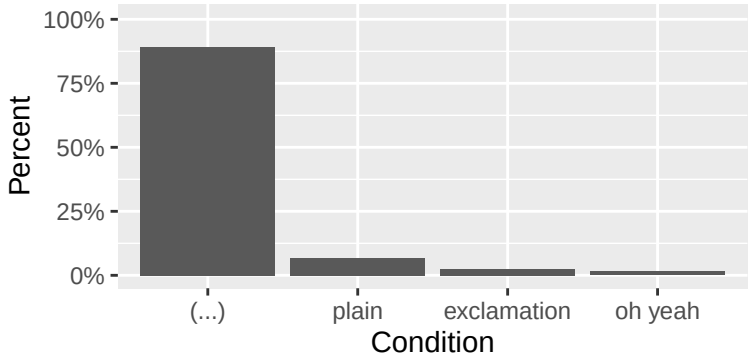

Figure 8.6: RSA speaker input: Action happened (probability of '(...)': 0.892; probability of 'plain': 0.068; probability of 'exclamation': 0.025; probability of 'oh yeah': 0.015)

8.4 The Joint Reasoning hRSA Model

A standard RSA model which incorporates joint reasoning (cf., Degen et al., 2015; Goodman & Frank, 2016) can represent both utterance-contingent changes in belief about the background world state, or activity *habituality*; and *habituality*- and utterance-contingent changes in beliefs about the current activity state. In this model, pragmatic listeners reason about the joint likelihood of a given *habituality* (h), and a given *activity state* (s), given a particular utterance (u):

$$P_{L_0}(s \mid u, h) \propto [\![u]\!](s) \cdot P(s \mid h) \tag{8.4}$$

$$P_{S_1}(u \mid s, h; \alpha, \lambda, C) \propto P(u; \lambda, C) \exp(\alpha \log P_{L_0}(s \mid u, h)) \tag{8.5}$$

$$P_{L_1}(s, h \mid u) \propto P_{S_1}(u \mid s, h; \alpha, \lambda, C) \cdot P(s \mid h) \cdot P(h) \tag{8.6}$$

The literal listener does not reason about activity habituality, as this is not a part of the literal interpretation of the utterance. In this case, it's possible to feed the empirical priors directly into the model, so that the likelihood of the activity occurring

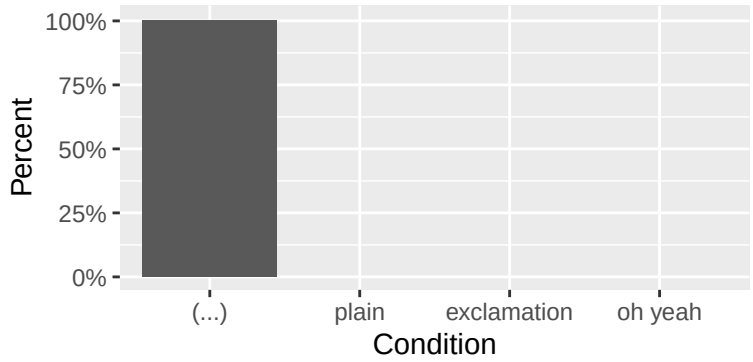

Figure 8.7: RSA speaker input: Action didn't happen (probability of '(...)': 1; probability of 'plain': 0; probability of 'exclamation': 0; probability of 'oh yeah': 0)

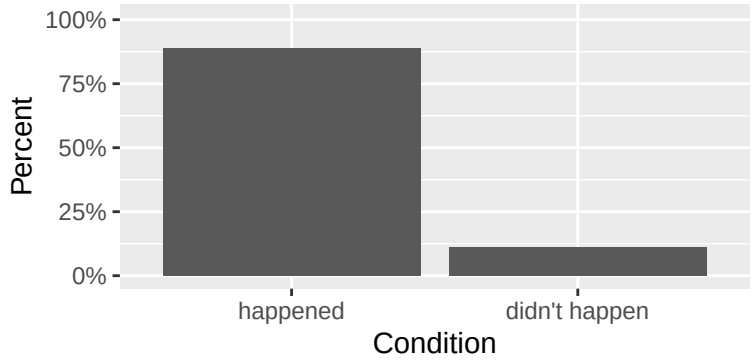

Figure 8.8: RSA pragmatic listener input: "(. . .)" (probability that event happened: 0.889; probability that it didn't happen: 0.111).

is conditional upon a sampled habituality. Whether the given activity occurred, or not (s), then, is simply a Bernoulli trial with $p = h$.

8.4.1 Model Predictions

Literal Listener

In Figures 8.12-8.14, it is clear that the literal listener determines that highly *habitual* activities almost certainly occurred; that moderately *habitual* activities may or may not have occurred; and that *non-habitual* activities almost certainly did not occur, provided there is some semantic ambiguity as to the fact.

Pragmatic Speaker

The pragmatic speaker is most likely to *not* describe a highly habitual activity explicitly, as expected, and particularly disprefers more effortful utterances; see Figure 8.15.

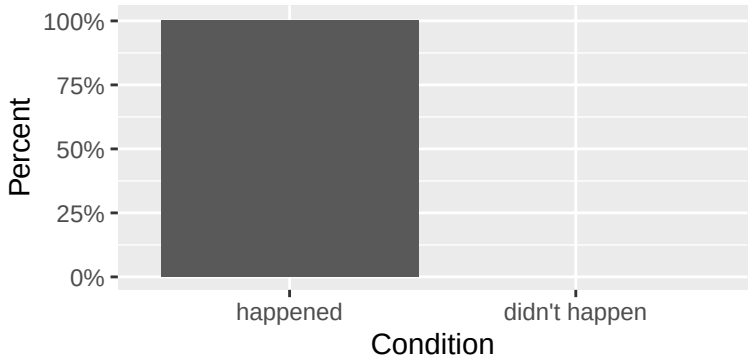

Figure 8.9: RSA pragmatic listener input: "John paid the cashier." (probability that event happened: 1; probability that it didn't happen: 0).

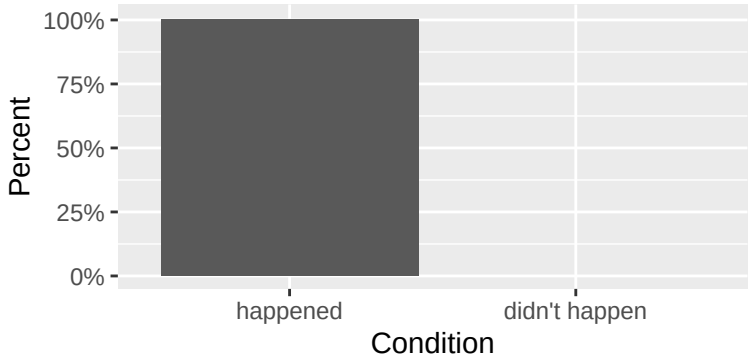

Figure 8.10: RSA pragmatic listener input: "John paid the cashier!" (probability that event happened: 1; probability that it didn't happen: 0).

In the case of moderately habitual activities, the speaker is far more likely to describe the activity explicitly, preferring the least effortful overt utterance; see Figure 8.16. To note, in the case of moderately predictable activities, it's unclear exactly how frequently one should expect for the activity to be mentioned overtly, but the qualitative pattern is consistent with predictions.

In the case of highly non-habitual activities, the speaker most often describes the activity explicitly, again preferring the least effortful utterance; see Figure 8.17. Critically, this model does *not* capture the intuition that speakers should choose more effortful utterances for particularly non-habitual activities - although, to be clear, this is an intuition about expected speaker behavior that is not yet empirically demonstrated.

Pragmatic Listener

The pragmatic listener considers unmentioned (typically habitual) activities to most likely be highly habitual, as one would expect; see Figure 8.18. However, explicitly mentioned activities are all equally likely to be interpreted as relatively non-habitual (see Figures 8.19-8.25), contrary to predictions that more effortful utterances should be perceived as relatively less habitual.

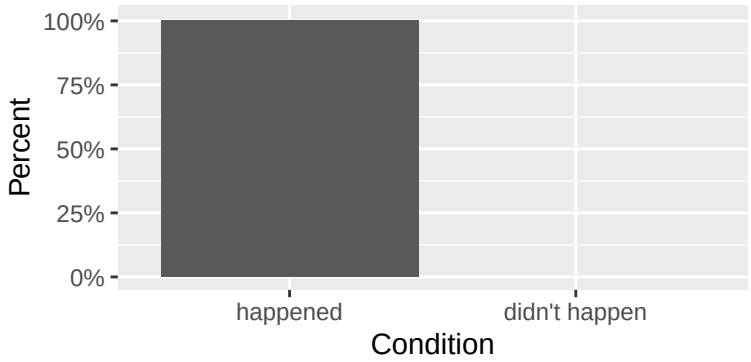

Figure 8.11: RSA pragmatic listener input: "Oh yeah, and John paid the cashier." (probability that event happened: 1; probability that it didn't happen: 0).

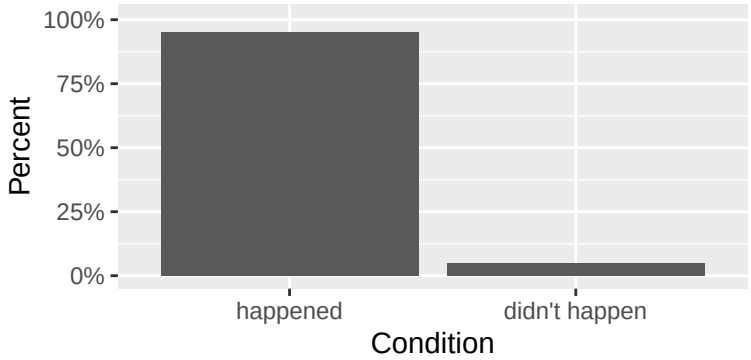

Figure 8.12: hRSA literal listener input: "(. . .)", 95% habituality (probability that event happened: 0.95; probability that it didn't happen: 0.05).

8.4.2 Model Summary

This model correctly captures the first empirical result: if a highly habitual activity is described explicitly, the listener is likely to interpret the habituality as *low*. Its shortcoming, however, is that there is no possibility of leveraging utterance costs to capture the second result (stronger inferences for more effortful utterances), and the last prediction (that speakers should use more effortful utterances to communicate unusual meanings.

There are 3 possible ways, in this case, of describing an activity explicitly: "plain," with a period at the end; using (implicit) exclamatory prosody; or with a discourse marker (*oh yeah, and. . .*) signifying the utterance's relevance to the discourse or listener – with the latter two more costly. As discussed in Section 8.4.1, the two more attentionally prominent utterances will never be of any advantage to the literal listener, in terms of effectively communicating the current world state. Likewise, they are of no advantage to the speaker, either in terms of likelihood of accurate message transmission to the listener or the speaker's presumed goal of conserving articulatory effort. As a consequence, the pragmatic listener will not infer that the more effortful

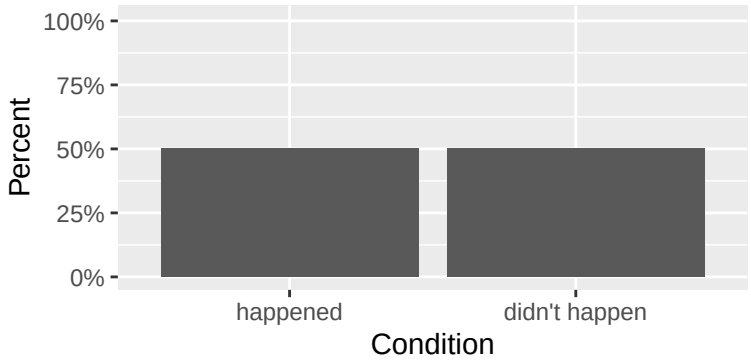

Figure 8.13: hRSA literal listener input: "(. . .)", 50% habituality (probability that event happened: 0.5; probability that it didn't happen: 0.5).

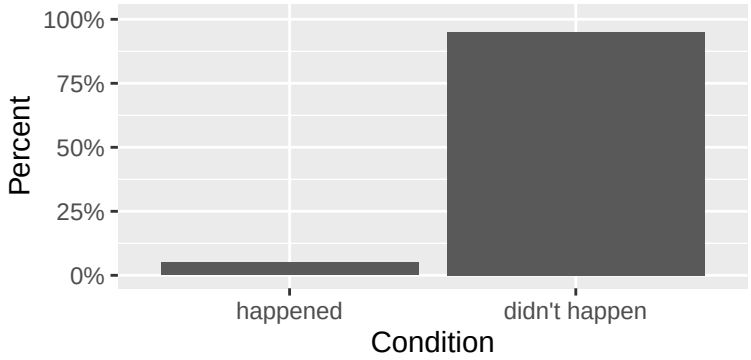

Figure 8.14: hRSA literal listener input: "(. . .)", 5% habituality (probability that event happened: 0.05; probability that it didn't happen: 0.95).

utterances are used in particularly unusual circumstances, compared to the "plain" utterance.

8.5 The Noisy Channel hRSA Model

As proven in Bergen et al. (2016), standard joint reasoning RSA models are unable to derive pragmatic inferences of different types or strengths, given semantically meaning-equivalent utterances. In order to capture the two remaining effects – a speaker preference for using more effortful utterance for unusual meanings, and stronger inferences for more effortful utterances – it is necessary to assign some communicative benefit to the more costly utterances, in terms of grabbing attention and facilitating recall, already active at the literal listener level. It is, in fact, quite plausible that comprehenders cannot accurately recall whether an activity has been explicitly mentioned, or not, as it has been shown that readers often cannot recall whether or not elements in a stereotyped activity sequence were explicitly mentioned (Bower et al., 1979). Further, informational redundancy, even at the multi-word level, in part has the purpose of ensuring that listeners attend to and accurately recall relevant information, implying that neither is guaranteed (Baker et al., 2008; Walker,

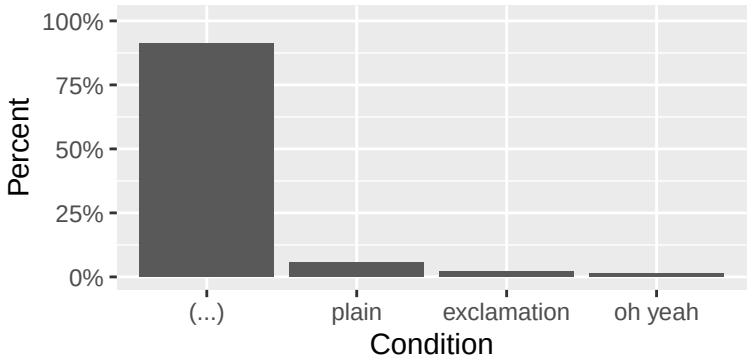

Figure 8.15: hRSA speaker input: activity happened, 95% habituality (probability of '(...)': 0.911; probability of 'plain': 0.056; probability of 'exclamation': 0.020; probability of 'oh yeah': 0.012)

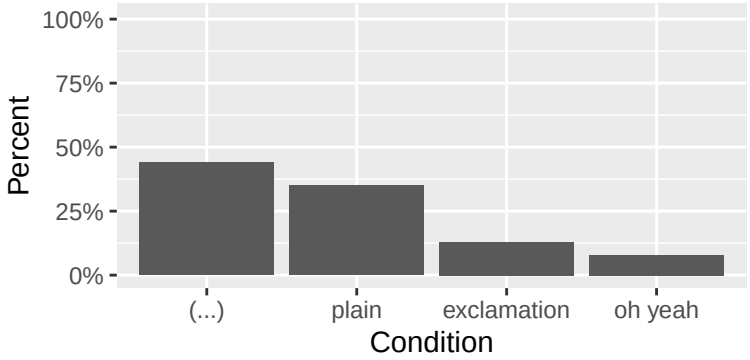

Figure 8.16: hRSA speaker input: activity happened, 50% habituality (probability of '(...)': 0.441; probability of 'plain': 0.351; probability of 'exclamation': 0.129; probability of 'oh yeah': 0.078)

1993).

The noisy channel RSA model proposed by Bergen & Goodman (2015), with fairly minimal modification, can successfully capture this intuition, as I show below. However, in this case I argue that this mechanism should be extrapolated to the notion of noisy encoding and recall, rather than simply perception, and that more effortful and attention-grabbing utterances should facilitate accurate encoding and recall, if only by virtue of the listener being more likely to attend to them.

$$P_{L_0}(s \mid u_r, h) \propto P(s \mid h) \cdot \sum_{u_i : [\![u_i]\!](s)=1} P(u_r \mid u_i) P(u_i) \tag{8.7}$$

$$P_{S_1}(u_i \mid s, h; \alpha, \lambda, C) \propto P(u_i; \lambda, C) \exp(\alpha \sum_{u_r} P(u_r \mid u_i) \log P_{L_0}(s \mid u_r, h)) \tag{8.8}$$

$$P_{L_1}(s, h \mid u_r) \propto P(s \mid h) \cdot P(h) \cdot \sum_{u_i} P_{S_1}(u_i \mid s, h; \alpha, \lambda, C) P(u_r \mid u_i) P(u_i) \tag{8.9}$$

In this model, it's assumed that every utterance has a non-trivial likelihood of not being actively attended to, and being mistaken for or mis-recalled as something akin

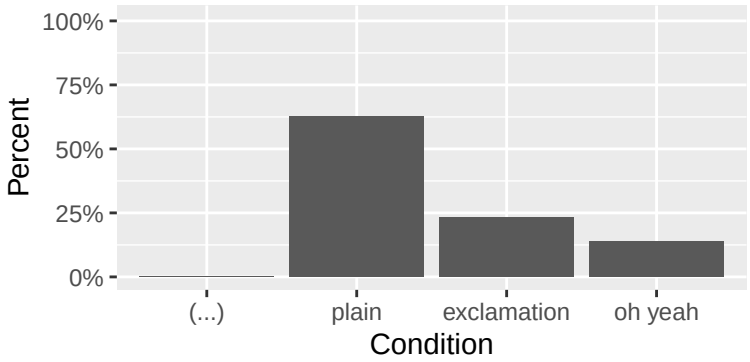

Figure 8.17: hRSA speaker input: activity happened, 5% habituality (probability of '(...)': ∼0; probability of 'plain': 0.628; probability of 'exclamation': 0.231; probability of 'oh yeah': 0.140)

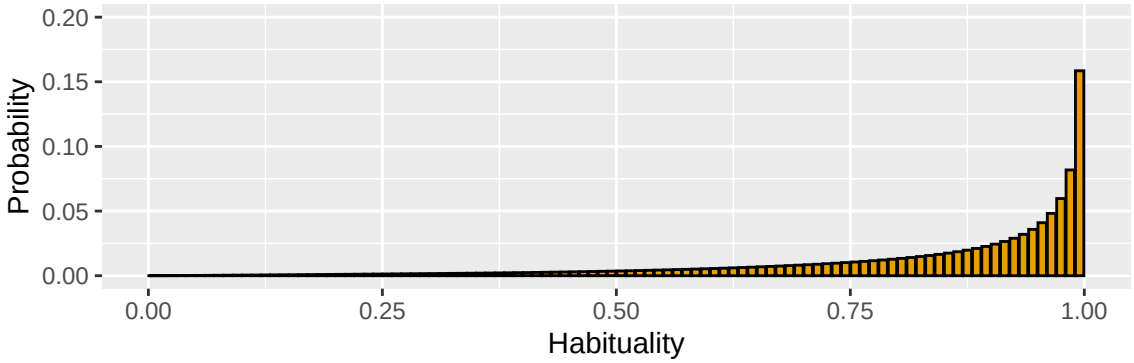

Figure 8.18: hRSA pragmatic listener input: "(. . .)"; habituality only

to its "perceptual neighbors" (as well as there being a very small chance of it being mis-recalled as a non-neighboring utterance). The "plain" utterance is considered to be perceptually "neighboring" to the two more effortful utterances, which are further moderately perceptually neighboring to each other. The "null" utterance is relatively perceptually neighboring to the "plain" utterance, although this relationship is possibly asymmetrical, as comprehenders may be more likely to misremember highly typical activities are not having been mentioned, than the other way around (this is, however, not critical for the functioning of the model). The values found in the confusion matrix (Table 8.1) are meant to reflect these qualitative judgements, but as empirical data on utterance confusability is lacking, are otherwise set somewhat arbitrarily.

At every level, the speaker or listener choose, or interpret, the utterance taking into account the possibility that it may not be (or have been) recalled correctly. As a result, speakers should prefer utterances less likely to be mis-recalled (i.e., more effortful utterances) when the intended meaning is otherwise unlikely to be inferred by the listener (i.e., the listener does not expect the activity to occur). Similarly, pragmatic listeners should interpret more effortful utterances as signifying that the meaning communicated is *unusual*; i.e., otherwise not likely to be inferred (which

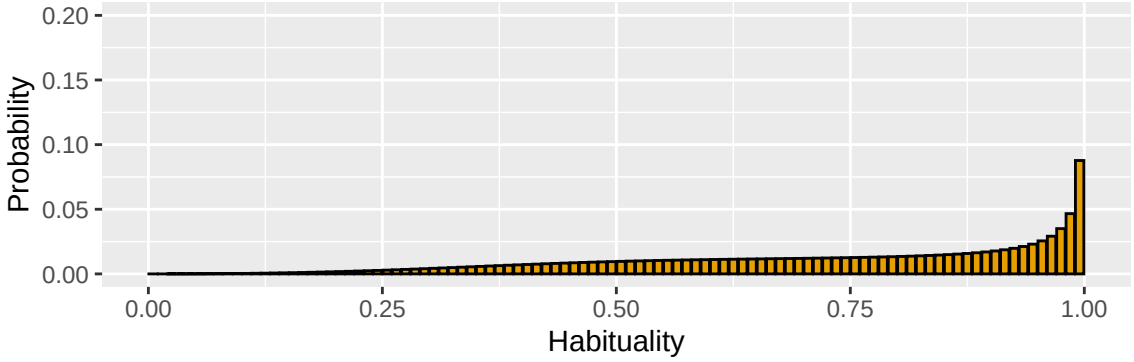

Figure 8.19: hRSA pragmatic listener input: "John paid the cashier."; habituality only

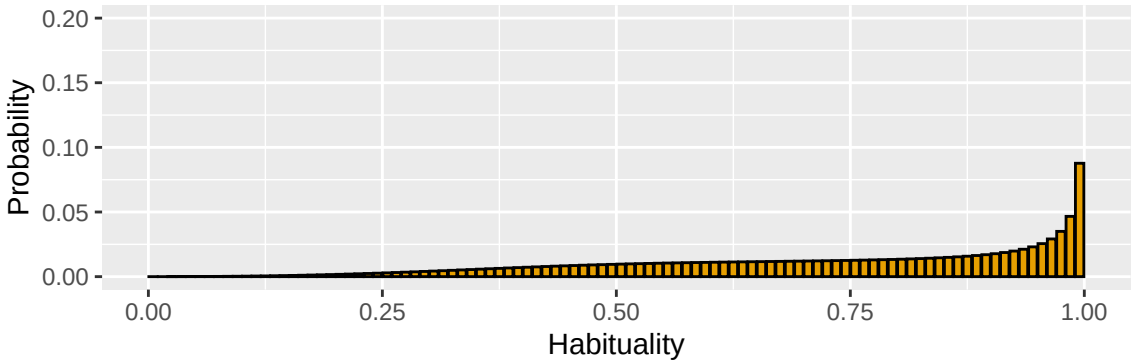

Figure 8.20: hRSA pragmatic listener input: "John paid the cashier!"; habituality only

Table 8.1: Confusion matrix which shows the estimated likelihood of mistaking one utterance for another.

Utterance	*(...)*	*He paid.*	*He paid!*	*Oh yeah...*
(...)	0.9900	0.01	0.0001	0.0001
He paid.	0.0100	0.95	0.0200	0.0200
He paid!	0.0001	0.02	0.9700	0.0100
Oh yeah...	0.0001	0.02	0.0100	0.9700

would not be the case if the activity described were, in fact, habitual).

8.5.1 Model Predictions

Literal Listener

The literal listener, as expected, perceives highly habitual activities as having most likely occurred, with a progressively lower expected likelihood of occurrence in the case of moderately habitual and non-habitual activities. As can be seen in Figures 8.26-8.28, relative to activity habituality, they are slightly more likely to assume that

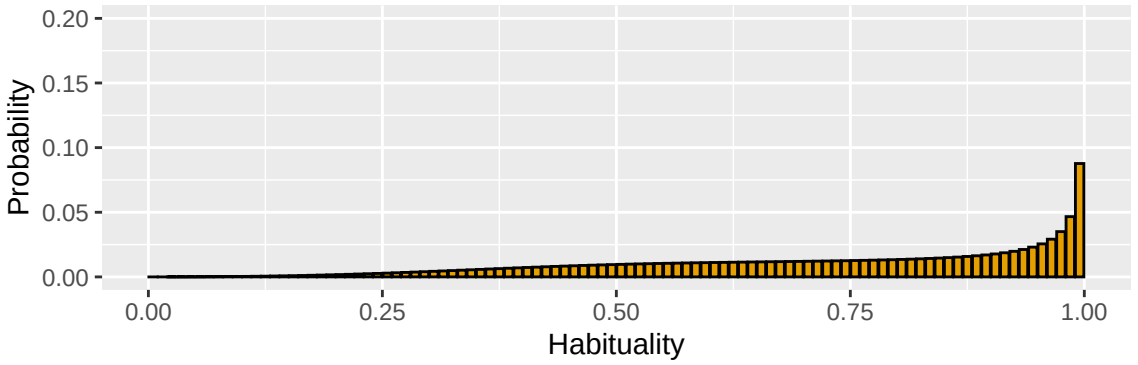

Figure 8.21: hRSA pragmatic listener input: "Oh yeah, and John paid the cashier."; habituality only

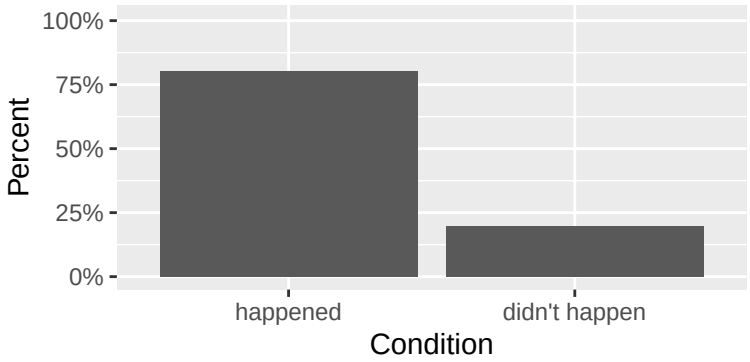

Figure 8.22: hRSA pragmatic listener input: "(. . .)"; state only (probability that event happened: 0.802; probability that it didn't happen: 0.198).

the activity occurred. This is due to a relatively elevated likelihood that an activity will be incorrectly recalled as having been described explicitly (in which case it is certain to have occurred).

Pragmatic Speaker

In the case of high-habituality activities, as before, speakers are very unlikely to describe the activity explicitly – and if they do, they tend towards less effortful utterances (see Figure 8.29).

Moderately habitual activities are only moderately likely to be mentioned, and again speakers gravitate towards less effortful utterances; see Figure 8.30. This is consistent with expectations, as moderately predictable activities are less likely to be assumed to not have occurred – it is therefore not quite as important to grab the listener's attention to ensure that they do, in fact, understand that the activity took place.

Non-habitual activities are almost always described explicitly, and as can be observed, speakers prefer to use a higher-effort utterance which is more likely to be attended to, and less likely to be mis-recalled as not having been uttered (see Figure

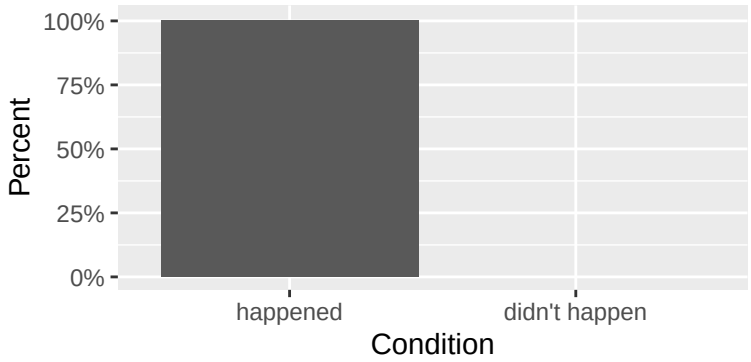

Figure 8.23: hRSA pragmatic listener input: "John paid the cashier."; state only (probability that event happened: 1; probability that it didn't happen: 0).

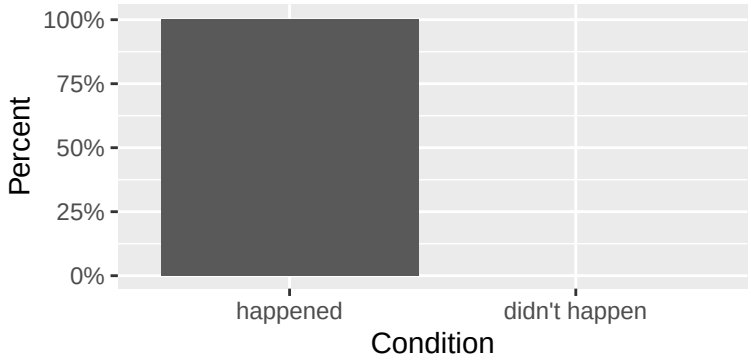

Figure 8.24: hRSA pragmatic listener input: "John paid the cashier!"; state only (probability that event happened: 1; probability that it didn't happen: 0).

8.31). This matches the last predicted effect: that speakers should use more effortful utterances for less predictable meanings.

Pragmatic Listener

As can be seen in Figures 8.32-8.35, pragmatic listeners perceive activities described overtly as less habitual – and furthermore, perceive activities described with more effortful utterances as less habitual than those described with a less effortful utterances. This confirms the two expected listener effects: that listeners interpret highly habitual activities which are explicitly mentioned as *less* habitual, and that these *habituality* inferences are stronger for more effortful utterances.

As can be seen in Figures 8.36-8.39, the comparatively low-effort "plain" utterance has a small likelihood of being remembered as not having been uttered – with a far smaller chance of the same in the case of high-effort utterances.

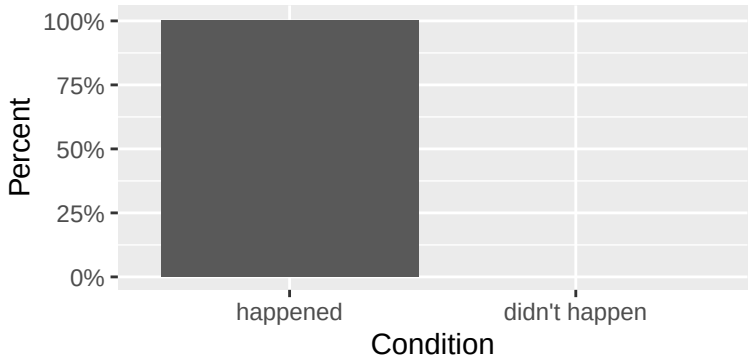

Figure 8.25: hRSA pragmatic listener input: "Oh yeah, and John paid the cashier."; state only (probability that event happened: 1; probability that it didn't happen: 0).

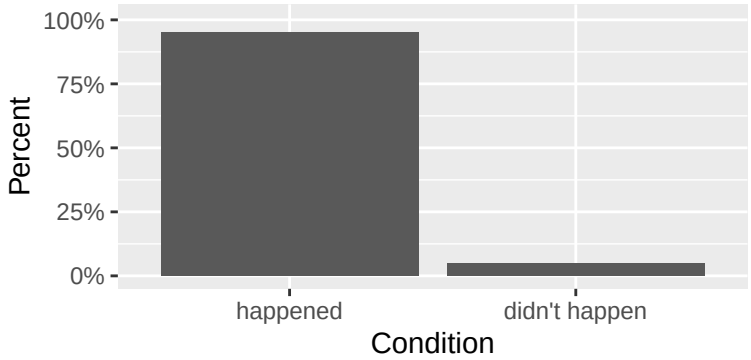

Figure 8.26: Noisy hRSA literal listener input: "(. . .)", 95% habituality (probability that event happened: 0.95; probability that it didn't happen: 0.05).

8.5.2 Comparison to Empirical Results

Overall, the results of the noisy channel hRSA model are a fairly close match, qualitatively, to those empirically measured in the experiments described in Chapter 4. To demonstrate this, the results and model predictions of interest are plotted side-by-side in Figure 8.40. In the case of the "null" utterance, I use participant-provided *habituality* ratings for cashier-paying in the "typical context - non-habitual activity" condition (*John went shopping. He got some apples!*), to control for the extra material seen by participants when giving *updated* habituality estimates.

The distribution tails in the case of the empirical data are fatter, and there is a hint of bimodality around the 50% mark. Otherwise, qualitatively the habituality densities match up fairly well, and the mean habitualities are qualitatively and numerically similar; see additionally Figure 8.41.

8.5.3 Model Summary

Overall, the noisy channel hRSA model qualitatively captures all empirically validated and predicted effects, and does so utilizing only machinery that has already

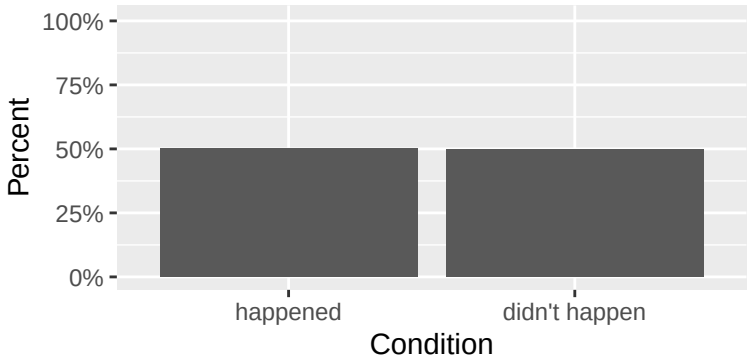

Figure 8.27: Noisy hRSA literal listener input: "(. . .)", 50% habituality (probability that event happened: 0.5; probability that it didn't happen: 0.5)

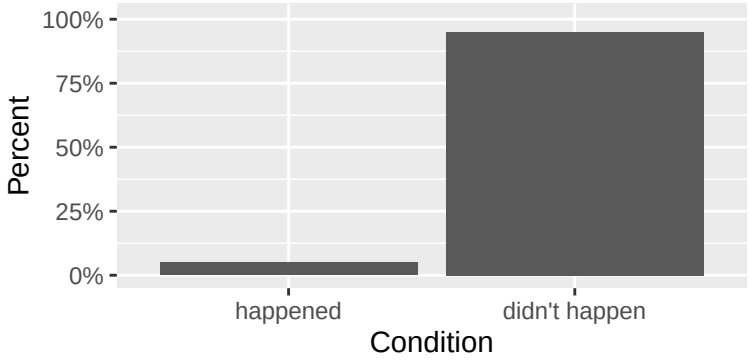

Figure 8.28: Noisy hRSA literal listener input: "(. . .)", 5% habituality (probability that event happened: 0.05; probability that it didn't happen: 0.95)

been established as necessary to account for other pragmatic phenomena. This is an arguable strength of this model, as RSA models may be criticized for including new parameters or machinery for the sake of accounting for specific pragmatic phenomena, with no justification of or generalization of the mechanism to other phenomena of interest.

8.6 Summary

The noisy channel hRSA model qualitatively captures all empirically validated and predicted effects of interest. As I've shown above, in the case of *habitual* activities, pragmatic listeners are more likely to conclude that activities which are mentioned explicitly are in fact relatively *non-habitual*. Further, more effortful utterances lead to stronger *habituality* inferences. Speakers are likewise more likely to use more effortful utterances to communicate less predictable meanings. This effect, although not empirically validated, is intuitive and straightforwardly predicted by theories such as UID.

This model confirms that joint reasoning RSA models can be successfully used to represent listeners' reasoning about the common ground, as in Degen et al. (2015). It

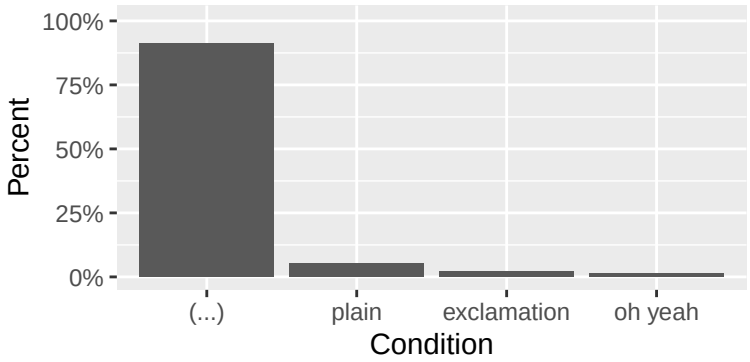

Figure 8.29: Noisy hRSA speaker input: activity happened, 95% habituality (probability of '(...)': 0.913; probability of 'plain': 0.054; probability of 'exclamation': 0.020; probability of 'oh yeah': 0.012)

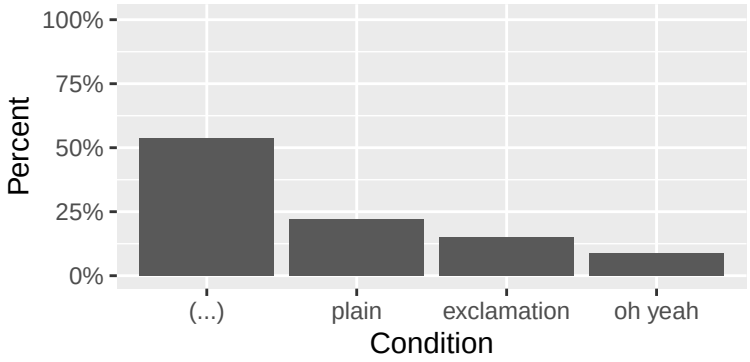

Figure 8.30: Noisy hRSA speaker input: activity happened, 50% habituality (probability of '(...)': 0.539; probability of 'plain': 0.223; probability of 'exclamation': 0.149; probability of 'oh yeah': 0.089)

also confirms and underscores the importance of not simply assuming that speakers are communicating through a clean channel – aside from being an unvalidated assumption (Levy, 2008), it fails to derive distinct pragmatic inferences, or inferences of different strengths, given truth-conditionally equivalent utterances (Bergen et al., 2016), which can be viewed as a potentially major flaw of standard RSA models. Critically, this model assumes that the noisy channel does not apply only at the level of perception, but likewise at the level of encoding and recall – what might be considered listener-internal noise. This enables the RSA model to reflect the effect that increased utterance salience should have on utterance choice or comprehension (cf. Wilson & Sperber, 2004).

Overall, unlike the UID model and hypothesis, the RSA model clearly predicts that unwarranted redundancy may distort the speaker's intended message, placing a clear limit on how redundant a speaker may be while still fulfilling their communicative goals. While the unformalized portions of the UID hypothesis have always been vague and difficult to represent or falsify, the RSA model further explicitly models the effect of redundancy on comprehension, allowing one to make clearer empirical predictions

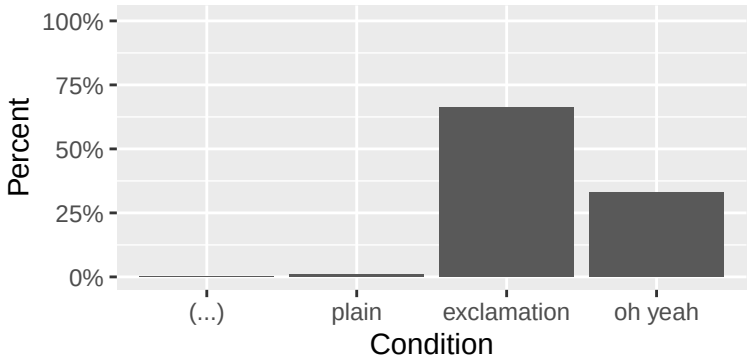

Figure 8.31: Noisy hRSA speaker input: activity happened, 5% habituality (probability of '(...)': 0.0003; probability of 'plain': 0.010; probability of 'exclamation': 0.660; probability of 'oh yeah': 0.329)

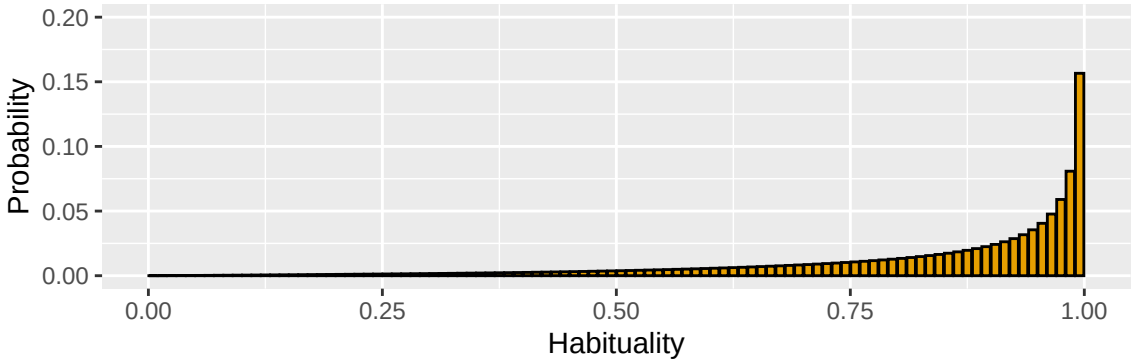

Figure 8.32: Noisy hRSA pragmatic listener input: "(. . .)"; habituality only

for exactly how informative a speaker may be when attempting to faithfully transmit their intended message to the listener. However, as I show, the standard clean channel RSA model likewise fails to account for predicted and empirically validated effects of increased effort, or attentional prominence, on utterance perception, encoding, or recall – and therefore, comprehension. In order to account for these effects, it is necessary to incorporate one of the core insights of UID – the notion of the *noisy channel* – into the RSA model.

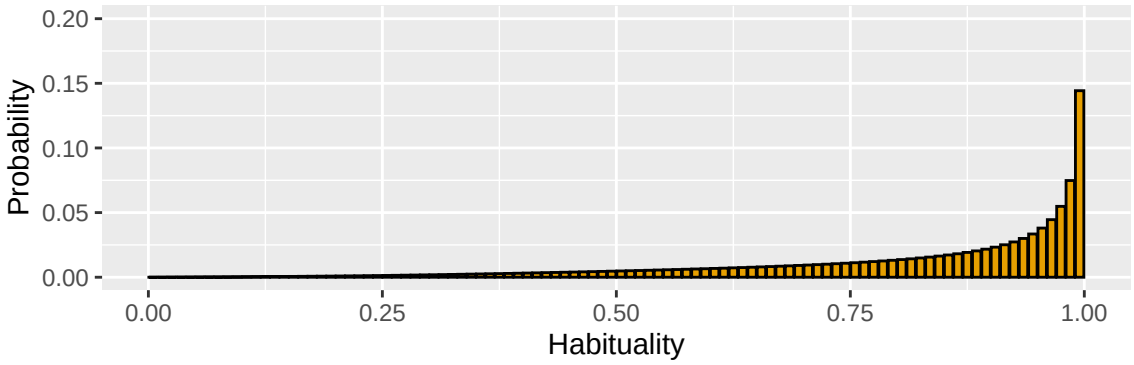

Figure 8.33: Noisy hRSA pragmatic listener input: "John paid the cashier."; habituality only

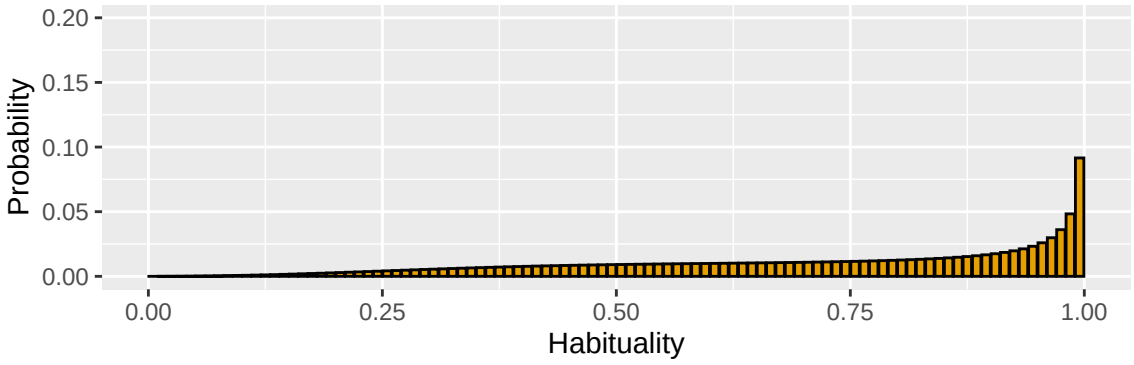

Figure 8.34: Noisy hRSA pragmatic listener input: "John paid the cashier!"; habituality only

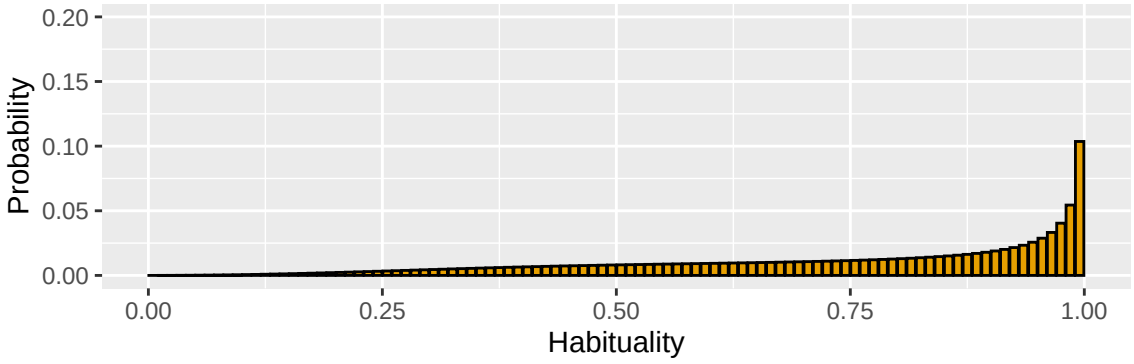

Figure 8.35: Noisy hRSA pragmatic listener input: "Oh yeah, and John paid the cashier."; habituality only

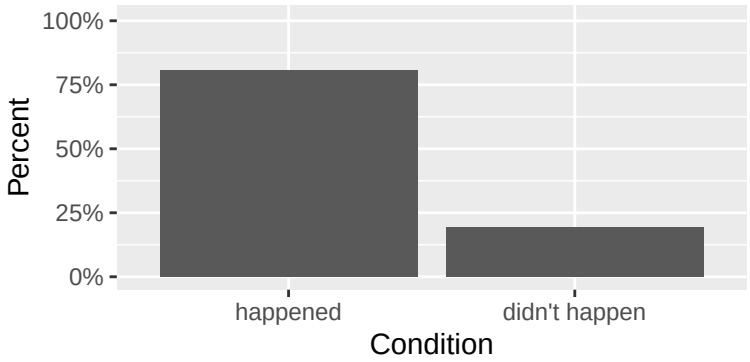

Figure 8.36: Noisy hRSA pragmatic listener input: "(. . .)"; state only (probability that event happened: 0.805; probability that it didn't happen: 0.195)

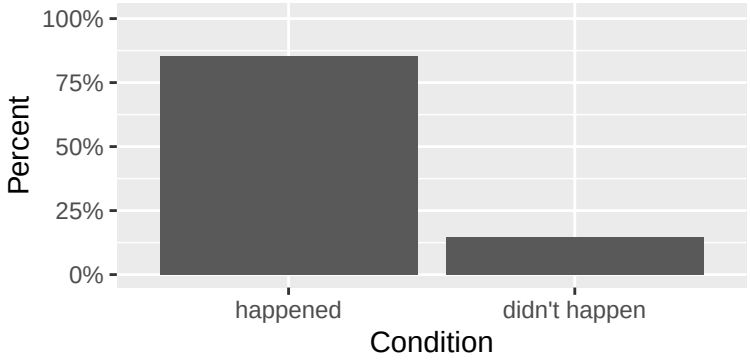

Figure 8.37: Noisy hRSA pragmatic listener input: "John paid the cashier."; state only (probability that event happened: 0.854; probability that it didn't happen: 0.146)

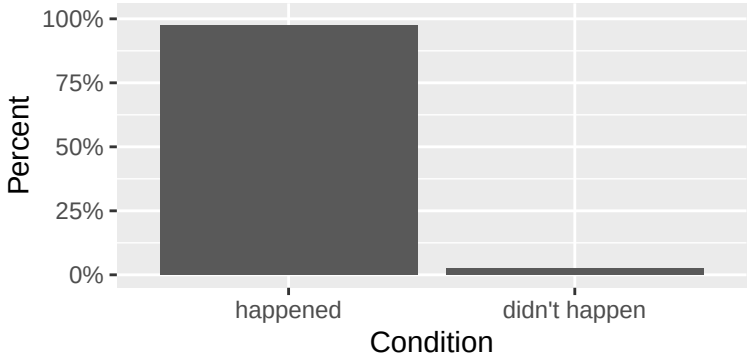

Figure 8.38: Noisy hRSA pragmatic listener input: "John paid the cashier!"; state only (probability that event happened: 0.976; probability that it didn't happen: 0.024)

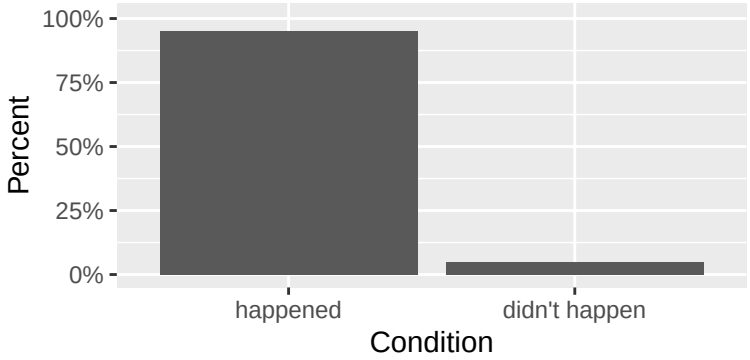

Figure 8.39: Noisy hRSA pragmatic listener input: "Oh yeah, and John paid the cashier."; state only (probability that event happened: 0.953; probability that it didn't happen: 0.047)

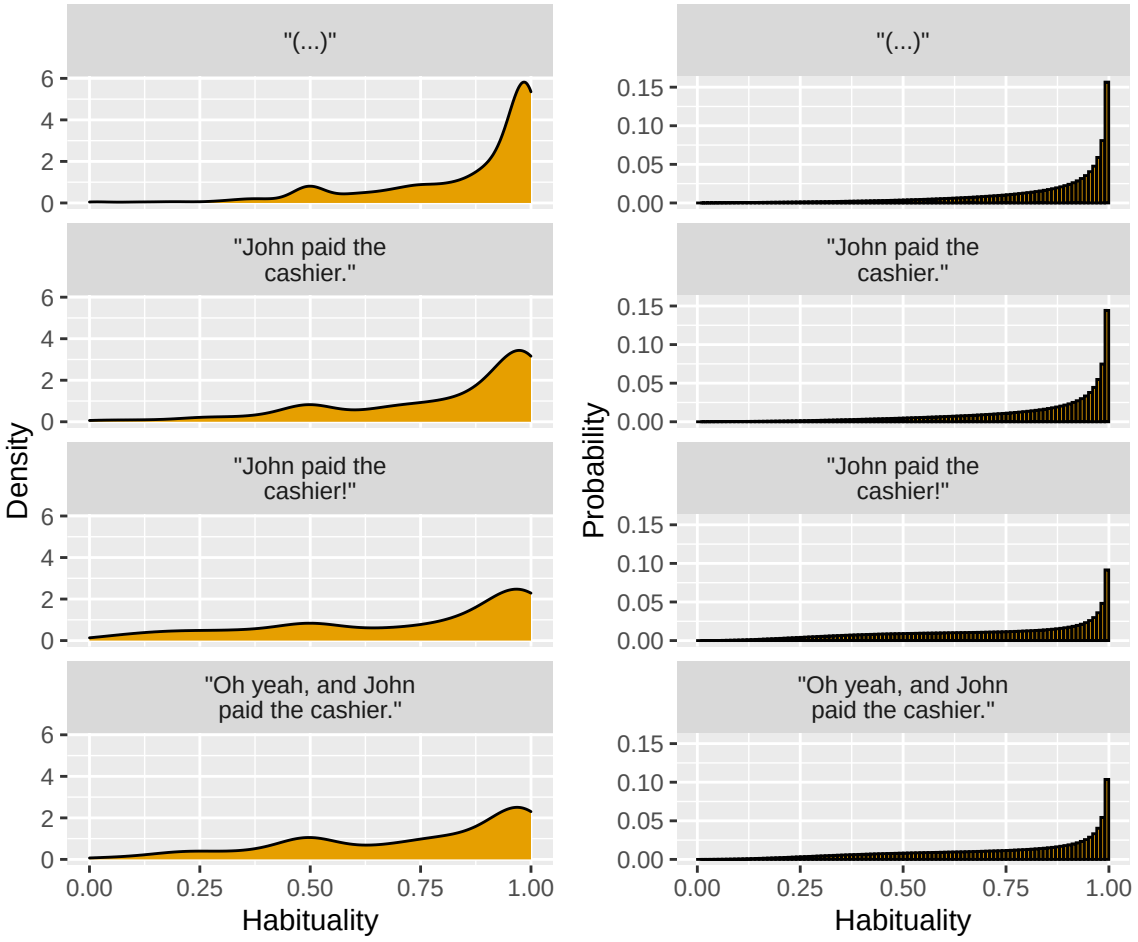

Figure 8.40: Empirical vs. predicted probability densities

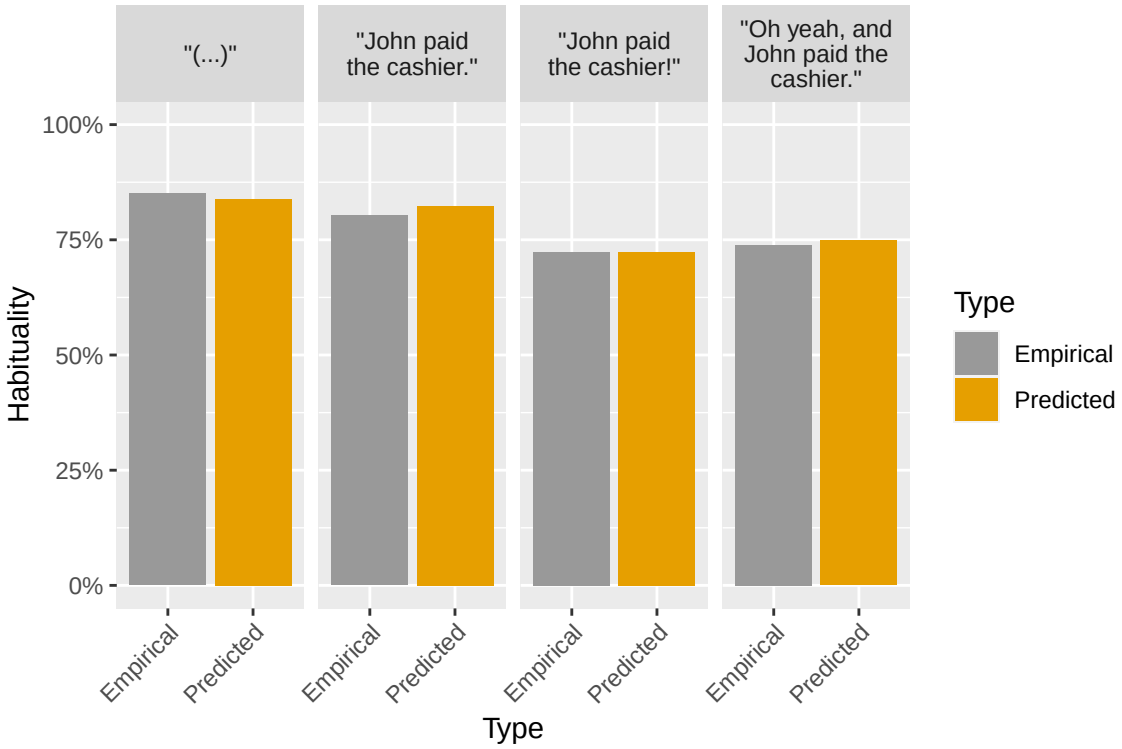

Figure 8.41: Empirical vs. predicted habituality means

Chapter 9

Rational Speech Act Model: Referring Expression Choice

In this chapter, I demonstrate that the Rational Speech Act (RSA) model is able to straightforwardly account for the observed influence of several factors, on whether referent predictability is found to influence referring expression choice (cf. Bott et al., 2018). A Rational Speech Act model of referring expression production was initially proposed by Orita et al. (2015). This model straightforwardly assumes that a speaker's choice of referring expression is determined by the expression's utility to the literal listener, as well as by the cost of the expression. However, it does not, by itself, account for cases in which prompt and task design significantly influence the likelihood of detecting an effect of predictability on referring expression choice, as is outlined in Section 9.1.2. Further, it does not account for the strong relative bias speakers exhibit towards referring to subjects with pronouns, and objects with names, irrespective of the referent likelihood, as demonstrated in all experimental work on this question to date – neither does it account for listeners taking a speaker's grammatical bias into account when interpreting reference (Rohde & Kehler, 2014).

First, in Section 9.1, I discuss the similarities and differences between the Bayesian pronoun interpretation model proposed by Rohde & Kehler (2014), and the Rational Speech Act model. I show that while both predict an relative asymmetry in utterance choice and utterance interpretation, where the same sets of contextual factors may affect production and interpretation to different degrees, the standard Rational Speech act model (which may similarly incorporate a partially grammar-based production bias) predicts a clear effect of referent predictability on referring expression choice, at least when pronoun reference is not trivially disambiguated.

In Section 9.2, I show that the Rational Speech Act model trivially confirms Bott et al. (2018)'s observation that effects of referent predictability on referring expression choice are stronger and more consistent when pronominal reference is ambiguous, than when it is unambiguous. However, this does not account for those cases in which

pronominal reference *is* unambiguous, and yet an effect of referent predictability on utterance choice is observed [cf. Rosa & Arnold (2017); Exp. 1 and 2]. Here, I argue that the *noisy channel* model introduced in the previous chapter can account for why this effect may still be detectable in some contexts, given that one third-person pronoun may easily be mistaken for another (or even mistakenly uttered).

In Section 9.3, I propose that effects of predictability on referring expression choice may be more easily detected using a *constrained choice* passage completion paradigm (as hypothesized by Bott et al., 2018) due to the fact that, in a constrained completion paradigm, participants are forced to refer back to antecedents that they do *not* consider likely completions. Although participants refer back to (on average) dispreferred antecedents even in free completion paradigms, it seems likely that they restrict those references only to those antecedents that they, individually, consider *sufficiently likely* continuations – particularly given the fact that the task instruction is specifically to write the *most likely continuation*. As the RSA model I present makes clear, if the constrained choice paradigm causes participants to refer to lower-predictability referents (whereas a free completion paradigm results in participants referring back only to somewhat higher-predictability referents), then the model does in fact make a straightforward prediction that effects of referent predictability on referring expression choice should be more easily detected using a constrained completion paradigm – a distinction that may otherwise seem unprincipled.

There remains an overarching question of why the relevant effect is detected only inconsistently in the passage completion paradigm, and remains rather weak – whereas it appears to be quite robust and replicable in the more naturalistic and interactive experiments in Rosa & Arnold (2017). While the UID model does not provide a straightforward way to discuss speakers expressing varying degrees of *rationality*, or sophistication of audience design, this is in fact something that can be approached fairly straightforwardly in the RSA framework. Although this is more speculative, I demonstrate that both modulating the degree to which the pragmatic speaker attempts to design their utterances with their listener in mind (using the α parameters), and modulating the pragmatic sophistication of the agent (with respect to level of recursion), may account for this disparity – intuitively, it appears plausible that more naturalistic and interactive tasks may prompt greater sophistication and listener-oriented *rationality* on the part of the speaker.

Finally, while the Bayesian interpretation model proposed by Rohde & Kehler (2014) is an appealing account that, at least for a subset of results, has excellent empirical coverage, it has the downside of providing no account of where, precisely the grammatical or topic-oriented bias governing referring expression choice originates. In contrast, both UID and the *Expectancy Hypothesis*, as discussed in Section 9.5, clearly argue that because the subject grammatical position correlates with increased likelihood of re-mention, 'reserving' the shorter expression for the (on average) more predictable referent makes the grammatical bias efficient, *on average*. As I show, if one makes the (I argue, reasonable) assumption that subjects are on average more

likely to be referred back to than objects, increasingly sophisticated speaker agents, potentially representing multiple generations of speakers gradually conventionalizing a certain pattern of expression, fairly quickly converge on a pronoun bias for subjects, and a name bias for objects. This requires one to posit no *a priori* grammatical bias on the part of speakers, and provides a parsimonious account of speaker utterance choice consistent with the wealth of evidence that *in general*, speaker utterance choices tend towards greater efficiency.

9.1 Bayesian Interpretation Model vs. Rational Speech Act Model

First, I present a very basic Bayesian model of referring expression production and interpretation, based on the Bayesian model of interpretation introduced by Rohde & Kehler (2014). To note, Rohde & Kehler (2014) do not present a comprehensive production model – their proposed production model can be reduced down to a grammatical speaker bias to overwhelmingly use pronouns to refer back to subjects, and names (or other longer expressions) for objects. What the Bayesian model accomplishes is accounting for the fact that interpretation biases do not straightforwardly mirror production biases, at least with respect to the relative influence of various contextual factors, or their magnitude. However, it is important to note that the Bayesian interpretation account, in itself, does not *require* that production biases are based only in the grammatical position, or topic status, of the antecedents, and is, *in itself*, agnostic about production choices.

9.1.1 Rohde & Kehler (2014) Model

Rohde & Kehler proposed the following pronoun interpretation model (omitting the normalization constant), where the probability of a referent given a particular referring expression is proportional to the probability of the expression, given the referent (which represents the speaker's production bias), modulated by the predictability of the referent (which represents the probability of the referent being mentioned in context):

$$P(\text{referent} \mid \text{expression}) \propto P(\text{expression} \mid \text{referent}) \cdot P(\text{referent}) \tag{9.1}$$

In this model, the speaker's utterance choices ($P(\text{expression} \mid \text{referent})$) are based solely on the grammatical position of the antecedent corresponding to the intended referent, with pronouns overwhelmingly preferred for subjects, and proper names overwhelmingly preferred for objects:

The production and interpretation biases, in this case, are validated experimentally – minimally, it is clear that speakers in this experiment do not base their choice

	state	prob
1	Pronoun	0.8
2	Noun	0.2

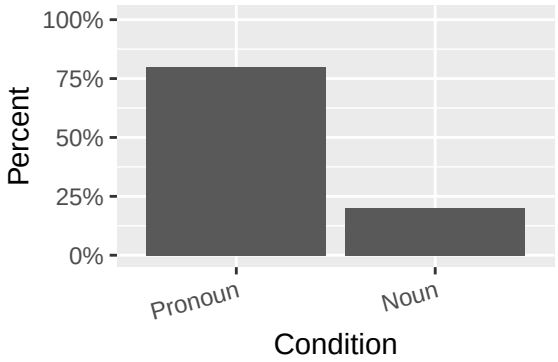

Figure 9.1: Rohde & Kehler (2014) speaker model: Referring back to subject (probability of using pronoun: 0.8; probability of using noun: 0.2).

	state	prob
1	Noun	0.8
2	Pronoun	0.2

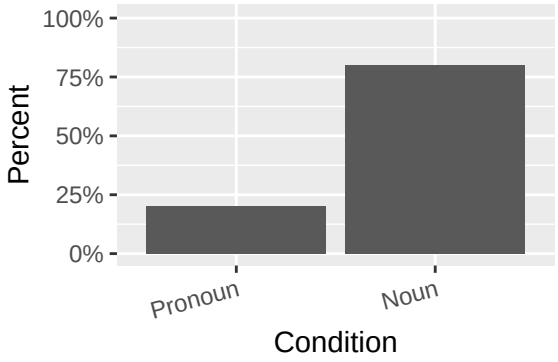

Figure 9.2: Rohde & Kehler (2014) speaker model: Referring back to object (probability of using pronoun: 0.2; probability of using noun: 0.8).

of referring expression on the referent's likelihood in context (P(referent), and that comprehenders base their interpretation of pronouns on both the speaker's production biases, and the referent's likelihood in context.

Speaker Utterance Choice

In the model below, I assume an 80% likelihood that a subject will be referred to with a pronoun, and a 20% likelihood that an object will be referred to with a pronoun.

As can be seen in Figures 9.1 and 9.2, speaker utterance choice in this model reflects the grammatical position of the antecedent, here implemented with an approximation (based on prior literature; cf. Rohde & Kehler, 2014) of pronouns being used to refer back to subjects roughly 80% of the time, and nouns being used to refer back to objects roughly 80% of the time.

Listener Interpretation

In Figures 9.3 and 9.4, it can be observed that listener pronoun interpretations do not mirror speaker referring expression choice preferences; listeners base their interpretations on *both* speaker preferences, and the likelihood of the referent being mentioned.

	state	prob
1	Subject	0.92307692
2	Object	0.07692308

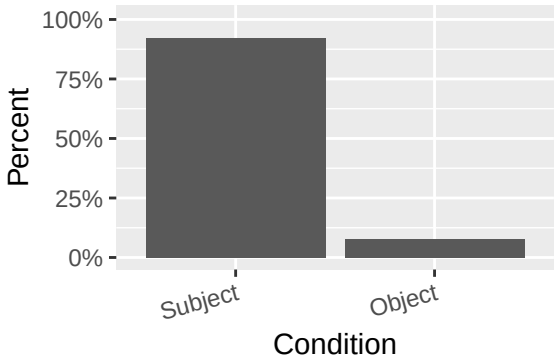

Figure 9.3: Rohde & Kehler (2014) interpretation model: Probability of interpreting an ambiguous pronoun as referring back to the subject, given a 75% prior likelihood of referring to the subject (probability of subject interpretation: 0.923; probability of object interpretation: 0.077).

	state	prob
1	Subject	0.5714286
2	Object	0.4285714

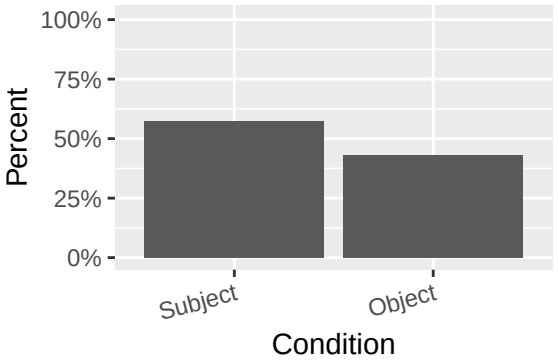

Figure 9.4: Rohde & Kehler (2014) interpretation model: Probability of interpreting an ambiguous pronoun as referring back to the subject, given a 25% prior likelihood of referring to the subject (probability of subject interpretation: 0.571; probability of object interpretation: 0.429).

Regardless of which referent is more predictable in context, listeners tend to interpret pronouns as referring back to subject antecedents, although it is clear that referent predictability partially modulates this bias.

As noted, this is, in essence, primarily an interpretation model. Unlike in the RSA framework, the speaker is not at all concerned with the utterance's utility to the listener (except, perhaps, *on average*), and bases their referring expression choice solely on the grammatical position of the last mention. This grammatical speaker bias here is not independently motivated; see Section 9.5, however, for an account of how such a bias might emerge, based on principles of efficient communication.

However, as there is no component of the interpretation model that inherently *necessitates* that the speaker not concern themselves with utterance utility, it can straightforwardly interface with the RSA model, where it would serve as a pragmatic listener model, reasoning about utterance meaning both with respect to the speaker's utterance choice biases, and the probability of the referent in question being mentioned. In the rest of this chapter, I explore the question of whether the RSA model may have better empirical coverage, and is able to account for otherwise unexplained differences in empirical results.

9.1.2 Basic RSA Model

In a standard RSA model of speaker referring expression choice, in contrast, speaker referring expression choice is based on both the likelihood of the referent being mentioned, and on utterance cost. In this case, I also introduce a grammatical bias, G, as the existence of a speaker bias towards using pronouns for subjects appears, at this point, to be universally observed.

$$P_{L_0}(\text{referent} \mid \text{expression}) \propto [\![\text{expression}]\!](\text{referent}) \cdot P(\text{referent}) \tag{9.2}$$

$$\begin{aligned} P_{S_1}(\text{expression} \mid \text{referent}; \alpha, \lambda, C, G) \propto P(\text{expression}; \lambda, C, G) \\ \exp(\alpha \log P_{L_0}(\text{referent} \mid \text{expression})) \end{aligned} \tag{9.3}$$

$$P_{L_1}(\text{referent} \mid \text{expression}) \propto P_{S_1}(\text{expression} \mid \text{referent}; \alpha, \lambda, C, G) \cdot P(\text{referent}) \tag{9.4}$$

In this model, I assume, based on estimates from passage completion literature (for ease of comparison), that a semantic bias towards mentioning a particular referent corresponds to roughly a 75% likelihood of said referent being mentioned. I further assume, given the empirical estimates in Rohde & Kehler (2014), that speakers have a baseline bias towards using pronouns for subjects roughly 80% of the time, and for objects 20% of the time. For the sake of simplicity, I modulate the cost-based production bias by the grammatical role bias. Both α and λ are set at 1.

	state	prob
1	Pronoun	0.8907682
2	Name: Subject	0.1092318

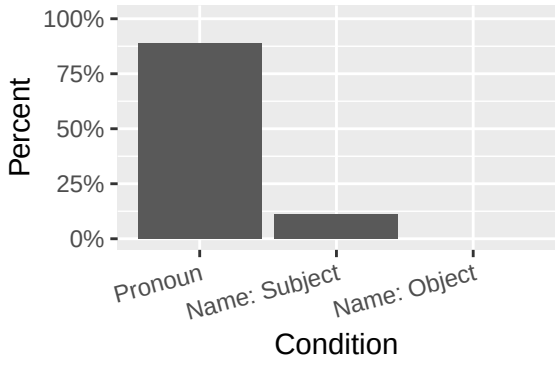

Figure 9.5: Basic RSA model: Speaker model; given reference to the subject and a subject-biased verb.

	state	prob
1	Name: Object	0.662393
2	Pronoun	0.337607

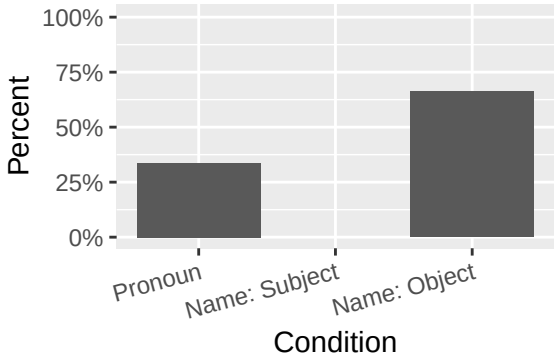

Figure 9.6: Basic RSA model: Speaker model; given reference to the object and an object-biased verb.

Speaker Utterance Choice

In Figures 9.5 and 9.6, one can see that, as in the Rohde & Kehler (2014) models, speakers predominately use pronouns to refer to subjects, and names to refer to objects. However, Figures 9.7 and 9.8 make it clear that this bias is *also* modulated by the referent's predictability in context.

Listener Interpretation

Figure 9.9 shows that pragmatic listeners overwhelmingly interpret pronouns as referring to subjects, when subjects are more likely to be mentioned. Figure 9.10 shows that given the particular parameters set in this model, listeners *are* more likely to interpret pronouns as referring to objects when objects are more likely to be mentioned, but only marginally so. As in Rohde & Kehler (2014) interpretation model, these biases do not mirror the production biases seen in Figures 9.5-9.8.

	state	prob
1	Pronoun	0.7310586
2	Name: Subject	0.2689414

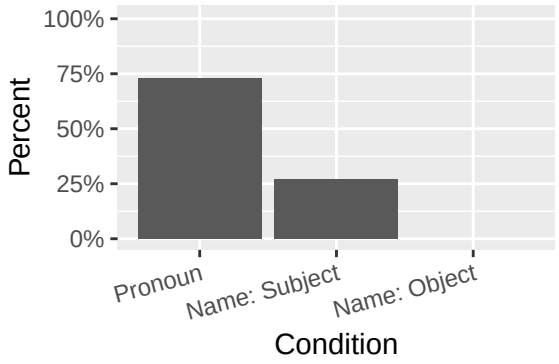

Figure 9.7: Basic RSA model: Speaker model; given reference to the subject and an object-biased verb.

	state	prob
1	Name: Object	0.8547793
2	Pronoun	0.1452207

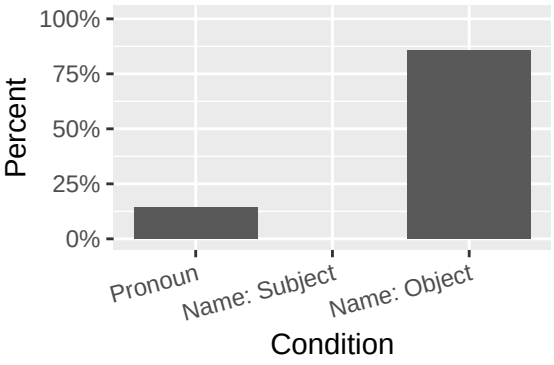

Figure 9.8: Basic RSA model: Speaker model; given reference to the object and a subject-biased verb.

	state	prob
1	Subject	0.94845807
2	Object	0.05154193

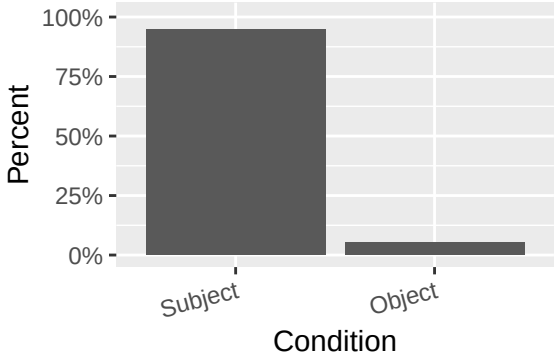

Figure 9.9: Basic RSA model: Listener model; given an ambiguous pronoun and a subject-biased verb.

	state	prob
1	Object	0.5807861
2	Subject	0.4192139

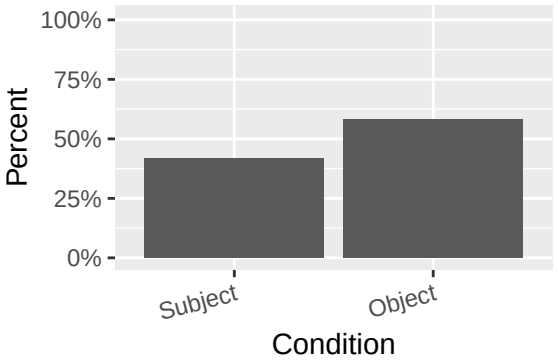

Figure 9.10: Basic RSA model: Listener model; given an ambiguous pronoun and an object-biased verb.

9.2 Impact of Ambiguous vs. Non-Ambiguous Reference

Bott et al. (2018) propose that an effect of predictability on referring expression choice is consistently detectable only when passage completion prompts contain same-gender antecedents, or when pronominal reference is ambiguous. Rohde & Kehler (2014) speculate that speakers may engage in audience design (i.e., consider which referent the listener expects to be referred to next) only when there is some question of whether the comprehender will be able to deduce the intended referent. However, this is more an intuition than a formal prediction.

The Rational Speech Act model considers utterance choice as a function of both utterance cost, and the relative utility of the utterance to the listener (with respect to their ability to recover the intended message). The standard clean channel RSA model indeed predicts that speakers will not take message predictability into account if the listener's ability to recover the intended meaning is not in question – making clear empirical predictions that are consistent with Bott et al. (2018)'s hypothesis that referent predictability plays a larger role in utterance choice when pronominal reference is ambiguous.

However, this runs into some conflict with the fact that Rosa & Arnold (2017) detect an effect of referent predictability on referring expression choice, even in the context of opposite-gender prompts, in two of their experiments (although not the third, which showed an effect only in the context of same-gender prompts). Here, I argue that a noisy channel RSA model, which introduces the possibility of one pronoun being mistaken for another (or being mistakenly uttered), straightforwardly accounts both for why effects of predictability are stronger and more consistent when reference is ambiguous, and for why effects of predictability, *nevertheless*, are still detectable when pronominal reference is unambiguous, even if they are weaker or less consistent. As the unambiguous pronoun and name in this scenario are meaning-equivalent, the model is otherwise unable to account for speaker preferences based on

	state	prob
1	Pronoun: Subject	0.91577619
2	Name: Subject	0.08422381

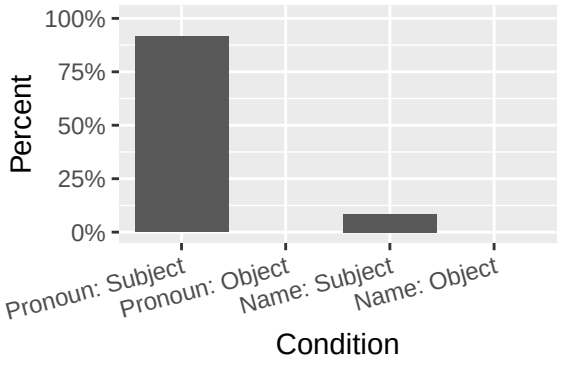

Figure 9.11: Non-ambiguous RSA model: Speaker model; reference to the subject, given a subject-biased verb.

	state	prob
1	Pronoun: Subject	0.91577619
2	Name: Subject	0.08422381

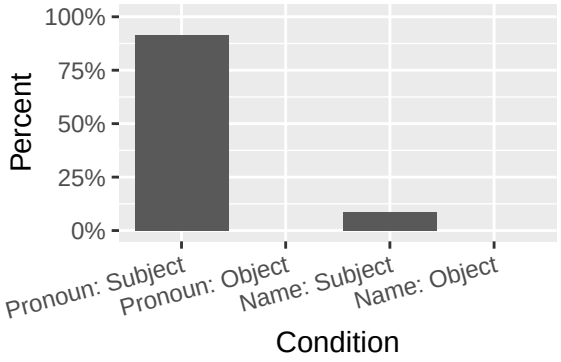

Figure 9.12: Non-ambiguous RSA model: Speaker model; reference to the subject, given an object-biased verb.

utterance cost, or effort expended (see Sections 7.2.3 for discussion).

RSA Model for Non-Ambiguous Pronouns (Speaker Choice)

In the model below, I assume a 75% likelihood that speakers will refer to the subject following a verb that biases towards a subject continuation, and a 25% likelihood that speakers will refer to the subject following a verb that biases towards an object continuation. α and λ are both set to 1. Pronouns are assumed to have a cost of 1, and Names of 2. I again assume 80% likelihoods of using pronouns to refer to subjects, and names to refer to objects.

Figures 9.11 and 9.12 show that when there is no additional utility in using a name (i.e., pronominal reference is unambiguous), speakers are as likely to use pronouns to refer to subjects in object-biased contexts, as in subject-biased contexts. Similarly, Figures 9.13 and 9.14 show the same for references to objects.

	state	prob
1	Name: Object	0.5953903
2	Pronoun: Object	0.4046097

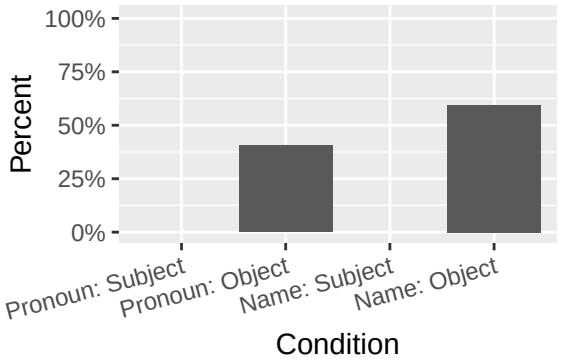

Figure 9.13: Non-ambiguous RSA model: Speaker model; reference to the object, given a subject-biased verb

	state	prob
1	Name: Object	0.5953903
2	Pronoun: Object	0.4046097

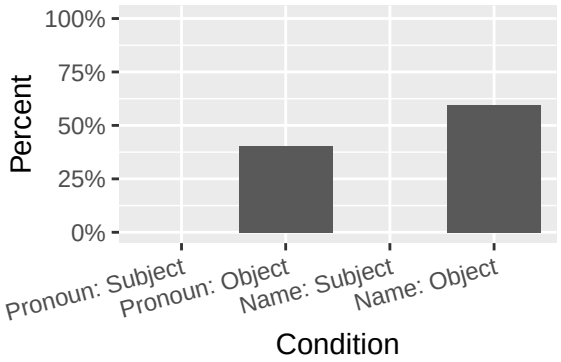

Figure 9.14: Non-ambiguous RSA model: Speaker model; reference to the object, given an object-biased verb.

Noisy Channel RSA Model for Non-Ambiguous Pronouns (Speaker Choice)

The noisy channel RSA model of referring expression production assumes that all utterances have, minimally, a chance of being misheard – particularly in the case of pronouns, which are minimal pairs. Speakers take account of the possibility that their utterances will be misperceived when selecting an utterance, and listeners similarly take account of the possibility that the utterance they perceived may not be the utterance intended. To note, this does introduce the possibility that speakers will use the incorrect pronoun to refer to an entity periodically, but as speech errors are common, particularly with minimal pair lexical items, I do not view this as conceptually problematic. For more discussion of the noisy channel RSA model, see Section 8.5.

$$P_{L_0}(\text{referent} \mid \text{expression}_p) \propto \sum_{\text{expression}_i : [\![\text{expression}_i]\!](\text{referent})=1} P(\text{expression}_p \mid \text{expression}_i) \\ P(\text{expression}_i) \cdot P(\text{referent}) \tag{9.5}$$

$$P_{S_1}(\text{expression}_i \mid \text{referent}; \alpha, \lambda, C, G) \propto P(\text{expression}_i; \lambda, C, G) \cdot \\ \exp(\alpha \sum_{\text{expression}_p} P(\text{expression}_p \mid \text{expression}_i) \cdot \\ \log P_{L_0}(\text{referent} \mid \text{expression}_p)) \tag{9.6}$$

$$P_{L_1}(\text{referent} \mid \text{expression}_p) \propto \sum_{\text{expression}_i} P_{S_1}(\text{expression}_i \mid \text{referent}; \alpha, \lambda, C, G) \cdot \\ P(\text{expression}_p \mid \text{expression}_i) P(\text{expression}_i) \cdot P(\text{referent}) \tag{9.7}$$

In Figures 9.15 through 9.18, one can see that once the model accounts for the influence of noise on perception, a small but detectable effect of predictability on pronominalization rates emerges. I assume the same parameters as in the previous model, as well as the confusion matrix in Table 9.1, with respect to what a given input is likely to be perceived as.

Table 9.1: Confusion matrix showing the estimated likelihood of one referring expression being mistakenly perceived as another. In this case it's assumed that the pronouns that would be used to refer to the subject and object are not identical.

	Pronoun (Sub.)	*Name (Sub.)*	*Pronoun (Obj.)*	*Name (Obj.)*
Pronoun (Sub.)	0.9800	0.0001	0.0200	0.0001
Name (Sub.)	0.0001	0.9990	0.0001	0.0001
Pronoun (Obj.)	0.0200	0.0001	0.9800	0.0001
Name (Obj.)	0.0001	0.0001	0.0001	0.9990

	state	prob
1	Pronoun: Subject	0.8595028488
2	Name: Subject	0.0842215151
3	Pronoun: Object	0.0559754083
4	Name: Object	0.0003002278

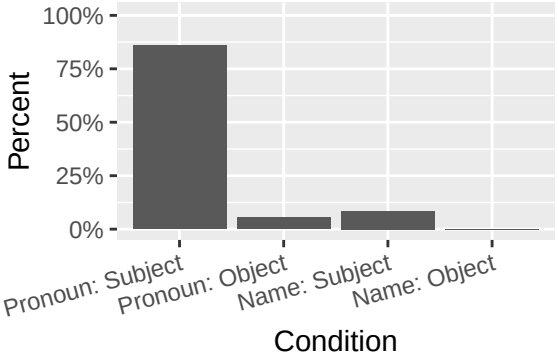

Figure 9.15: Noisy channel non-ambiguous RSA model: Speaker model, given a subject-biased verb and a subject reference.

	state	prob
1	Pronoun: Subject	8.956947e−01
2	Name: Subject	9.643442e−02
3	Pronoun: Object	7.832431e−03
4	Name: Object	3.842641e−05

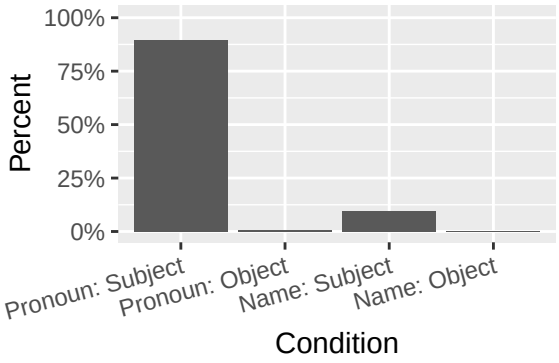

Figure 9.16: Noisy channel non-ambiguous RSA model: Speaker model, given an object-biased verb and a subject reference.

In this case, the assumption is that pronouns (which are shorter and phonologically very similar) are significantly more likely to be misperceived than nouns (which are longer and more distinct from one another, as well a pronouns).

Next, I look at whether the RSA framework can account for an increased effect of predictability on referring expression choice, in the context of constrained completion tasks, in comparison to free completion tasks.

9.3 Impact of Free vs. Constrained Completion

In this section, I show how one may straightforwardly derive a greater effect of predictability on referring expression choice when using a *constrained* passage completion paradigm, compared to the case of a *free* completion paradigm. In a *free completion* experiment, participants are instructed to write the *most likely* continuation. Given this, one may presume that any continuation they *do* write meets, from their reading of the prompt and judgment of the context, a certain *predictability threshold* – for example, minimally a subjective 45% likelihood of mention.

If a speaker agent is only minimally *rational* with respect to utility maximization, or is a particularly pragmatically unsophisticated agent (see Section 9.4), then it

	state	prob
1	Name: Object	0.5950750580
2	Pronoun: Object	0.3797469938
3	Pronoun: Subject	0.0248751460
4	Name: Subject	0.0003028023

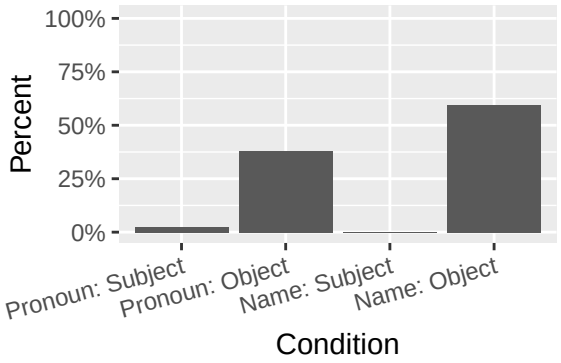

Figure 9.17: Noisy channel non-ambiguous RSA model: Speaker model, given an object-biased verb and a object reference.

	state	prob
1	Name: Object	6.298243e−01
2	Pronoun: Object	3.669127e−01
3	Pronoun: Subject	3.227162e−03
4	Name: Subject	3.582421e−05

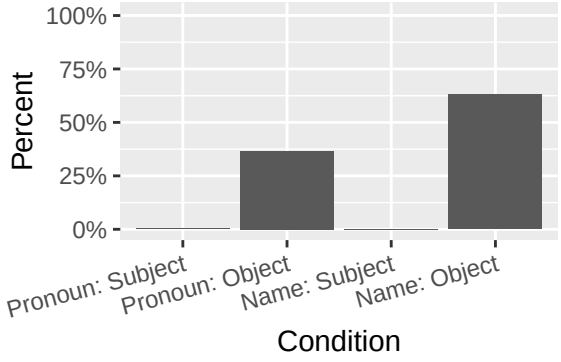

Figure 9.18: Noisy channel non-ambiguous RSA model: Speaker model, given a subject-biased verb and an object reference.

	state	prob
1	Pronoun	0.8717242
2	Name: Subject	0.1282758

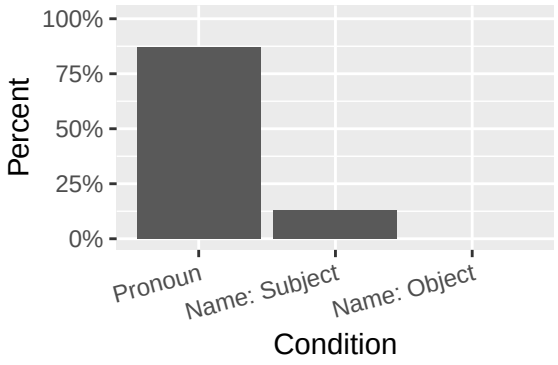

Figure 9.19: Free completion RSA model: Speaker model; given a subject reference and a subject-biased verb.

	state	prob
1	Pronoun	0.8030497
2	Name: Subject	0.1969503

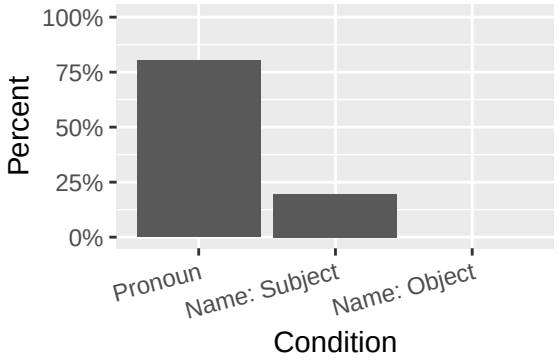

Figure 9.20: Free completion RSA model: Speaker model; given a subject reference and an object-biased verb.

would be expected that relatively minor differences in how relatively *unpredictable* those referents mentioned can be may cause any effect of predictability on production to wash out, in the case of the free completion paradigm, and to be relatively amplified in the case of the constrained completion paradigm. In other words, if all referents that *are* mentioned are relatively similar in their subjective predictability, from the participant's point of view, then one should not expect to see a great difference in predictability on referring expression choice.

Free Completion

Figures 9.19 through 9.22 show that when the difference between "predictable" and "unpredictable" referents is constrained, the effect of referent predictability on referring expression choice decreases. In an experimental context, particularly given relatively unsophisticated speaker agents, or agents that make only minimal attempts to maximize an utterance's utility, it is likely that small effects may wash out. In the following model, I assume a higher likelihood of referencing an object following a subject-biasing verb (45% vs. 25%), as well as for referencing a subject following an object-biasing verb.

	state	prob
1	Name: Object	0.7018866
2	Pronoun	0.2981134

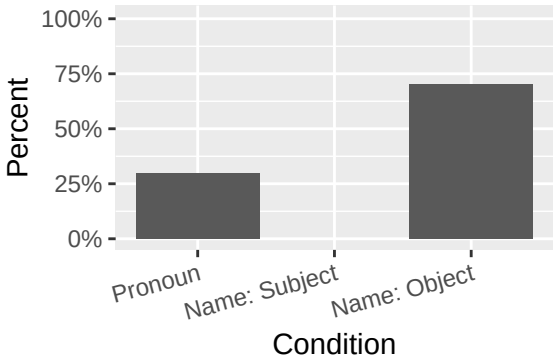

Figure 9.21: Free completion RSA model: Speaker model; given an object reference and an object-biased verb.

	state	prob
1	Name: Object	0.796915
2	Pronoun	0.203085

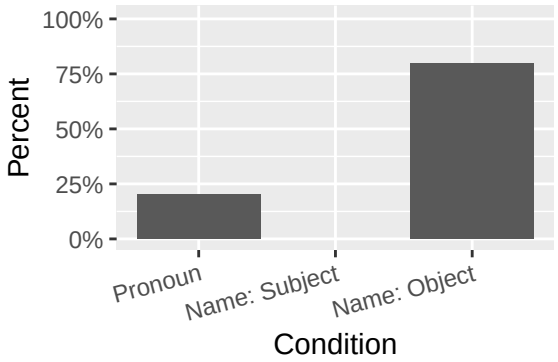

Figure 9.22: Free completion RSA model: Speaker model; given an object reference and a subject-biased verb.

	state	prob
1	Pronoun	0.8907682
2	Name: Subject	0.1092318

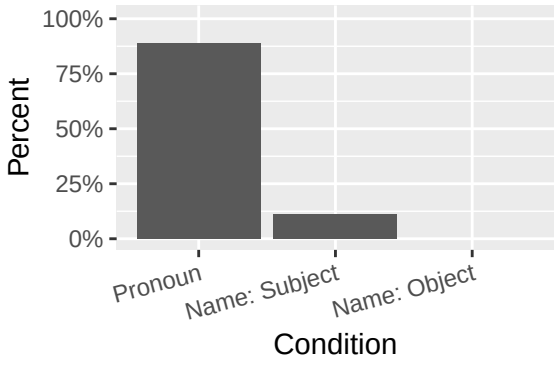

Figure 9.23: Constrained completion RSA model: Speaker model; given a subject reference and a subject-biased verb.

	state	prob
1	Pronoun	0.7310586
2	Name: Subject	0.2689414

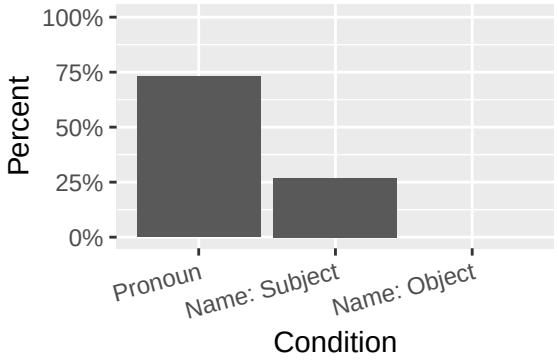

Figure 9.24: Constrained completion RSA model: Speaker model; given a subject reference and an object-biased verb.

Constrained Completion

Figures 9.23 to 9.26 show that, given a greater distinction between the contextual predictabilities of "predictable" and "unpredictable" referents, the effects of predictability on referring expression choice are understandably significantly greater, and therefore more likely to be detected in an experimental context. In the following model, I assume the same relevant parameters as in previous models.

In the following section, I consider the fact that, even if the above distinctions in prompt and task design are taken into account, it remains difficult to explain the empirical results. For instance, in Section 6.4.2, I do not detect an effect of referent predictability on referring expression choice, despite using same-gender antecedents, and a constrained completion task design. In contrast, Rosa & Arnold (2017) detect robust effects of referent predictability on production in two of their experiments, even when using opposite-gender antecedents in completion prompts. I argue that these differences may be accounted for by critical distinctions in the naturalness and interactivity of the experimental setups used.

	state	prob
1	Name: Object	0.662393
2	Pronoun	0.337607

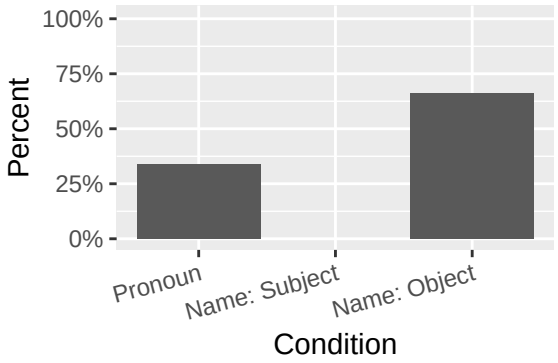

Figure 9.25: Constrained completion RSA model: Speaker model; given an object reference and an object-biased verb.

	state	prob
1	Name: Object	0.8547793
2	Pronoun	0.1452207

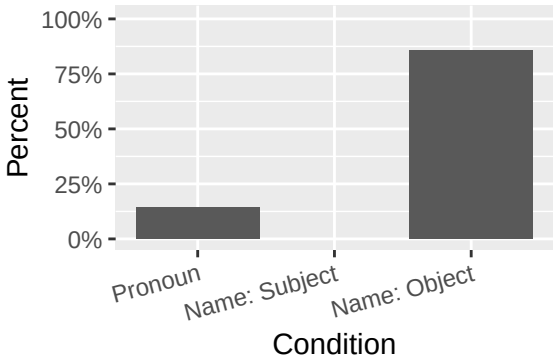

Figure 9.26: Constrained completion RSA model: Speaker model; given an object reference and a subject-biased verb.

9.4 Rationality and Pragmatic Sophistication of Agents

In this section, I argue that the small to non-detectable effect of referent predictability on referring expression choice, in the case of written passage completion paradigms, may be accounted for by the paradigm being less likely to evoke a degree of *audience design*, or pragmatic sophistication, on the part of speakers. In a non-interactive, non-naturalistic setting, where there does not appear to be a likely or plausible audience, speakers may not be prompted to make particular effort to optimize the utility of their utterances – and may rather prefer to rely on coarse heuristics (e.g., if subjects are on average more predictable, then they may be referred to with pronouns across the board, even if they are not predictable *in context*).

In contrast, more interactive, naturalistic, or discourse-rich paradigms may be more likely to prompt speakers to engage in audience design (i.e., to maximize the utterance's utility to the listener), or to engage in more sophisticated pragmatic reasoning. In Section 6.5, I discuss related empirical work which suggests that speakers may engage in notably more audience design when tasks are interactive, and confederates are more believable.

Ultimately, I assume that speakers may choose to engage in more, or less audience design, or maximization of utterance utility. In the RSA framework, this translates to modulation of the α parameter. In the following section, I show that at very low levels of α, effects of referent predictability on speaker referring expression choice virtually disappear. In contrast, if α is higher, presumably in contexts where the speaker has an audience they must communicate with, then effects begin to emerge.

Another way of representing speaker sophistication is by varying the level of recursive pragmatic reasoning that the speaker engages in. Franke & Degen (2016) look at individual differences in how participants respond using a more complex version of a Frank & Goodman (2012) style reference game. Typically in RSA models it's presumed that speakers reason about literal listeners only. Franke & Degen (2016) however found that approximately 15% of speakers engaged in more sophisticated pragmatic reasoning in their experimental paradigm - i.e., speakers reasoning about pragmatic listeners reasoning about speakers (reasoning about literal listeners). I thus consider the possibility that different experimental paradigms may push agents to be either more, or less sophisticated in their pragmatic reasoning.

Reduced vs. Increased Listener Utility

I compare two clean channel RSA models assuming ambiguous reference, and differing only in the degree of audience design the listener engages in. In the case of the reduced listener utility model, α is set to 0.01, whereas in the increased listener utility model, α is set to 1.25.

	state	prob
1	Pronoun	0.91555404
2	Name: Subject	0.08444596

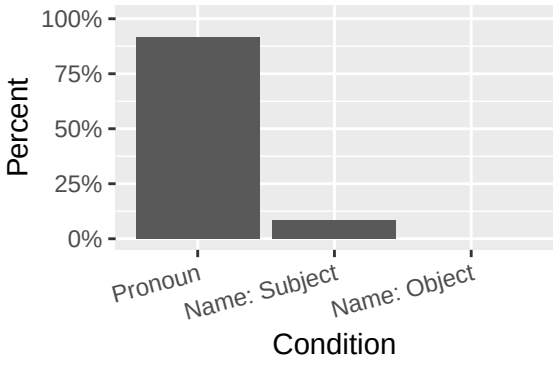

Figure 9.27: Reduced listener utility RSA model: Speaker model; given a subject reference and a subject-biased verb.

	state	prob
1	Pronoun	0.91470076
2	Name: Subject	0.08529924

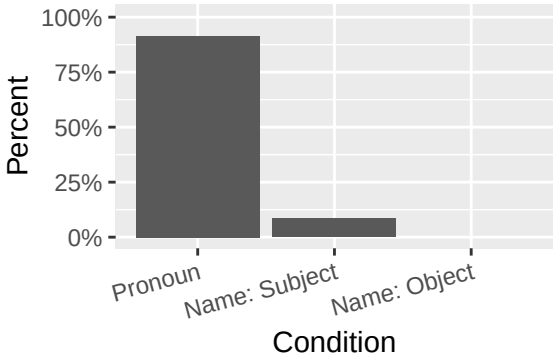

Figure 9.28: Reduced listener utility RSA model: Speaker model; given a subject reference and an object-biased verb.

Reduced Listener Utility Figures 9.27 through 9.30 show that when agent *rationality* (with respect to α) is arbitrarily reduced, the effect of predictability on referring expression choice is nearly undetectable.

Increased Listener Utility In contrast, Figures 9.31 through 9.34 show that when agent *rationality* is increased, the effect of predictability on referring expression choice correspondingly increases.

Variation in Agent Sophistication

Here, I again compare two clean channel RSA models assuming ambiguous reference, and differing in the degree of speaker agent sophistication. As in Franke & Degen (2016), I also consider the possibility of literal speakers who are only concerned with producing true descriptions without any further pragmatic reasoning (S_0). I compare this speaker model to one of a more sophisticated speaker (S_2), who reasons about the pragmatic listener L_1.

	state	prob
1	Name: Object	0.5960832
2	Pronoun	0.4039168

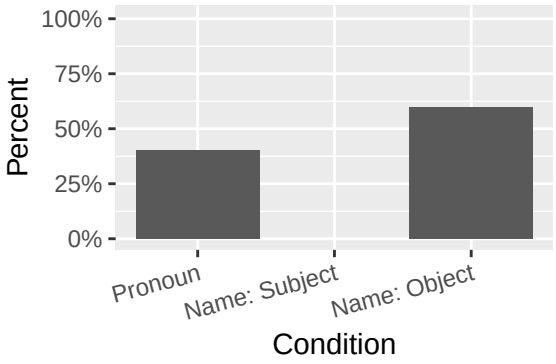

Figure 9.29: Reduced listener utility RSA model: Speaker model; given an object reference and an object-biased verb.

	state	prob
1	Name: Object	0.5987255
2	Pronoun	0.4012745

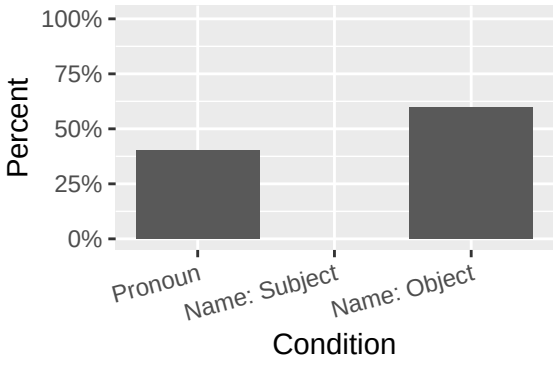

Figure 9.30: Reduced listener utility RSA model: Speaker model; given an object reference and a subject-biased verb.

	state	prob
1	Pronoun	0.8835712
2	Name: Subject	0.1164288

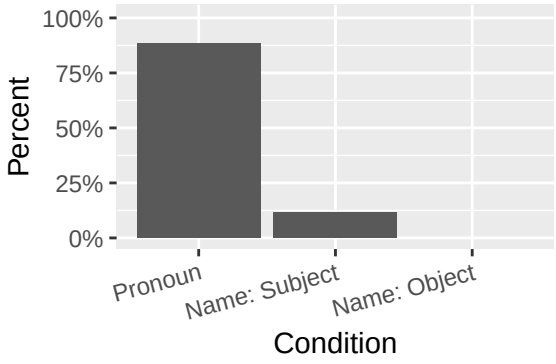

Figure 9.31: Increased listener utility RSA model: Speaker model; given a subject reference and a subject-biased verb.

	state	prob
1	Pronoun	0.6577822
2	Name: Subject	0.3422178

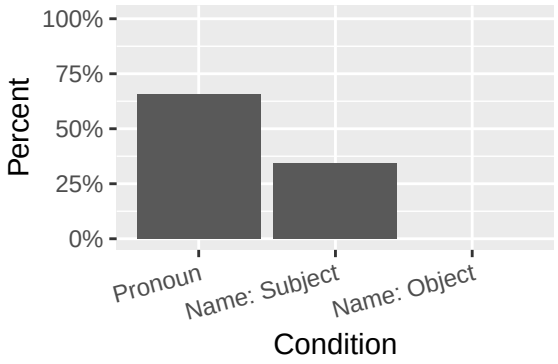

Figure 9.32: Increased listener utility RSA model: Speaker model; given a subject reference and an object-biased verb.

	state	prob
1	Name: Object	0.678284
2	Pronoun	0.321716

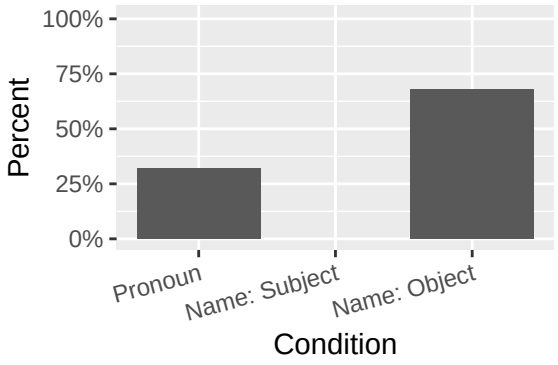

Figure 9.33: Increased listener utility RSA model: Speaker model; given an object reference and an object-biased verb.

	state	prob
1	Name: Object	0.8927518
2	Pronoun	0.1072482

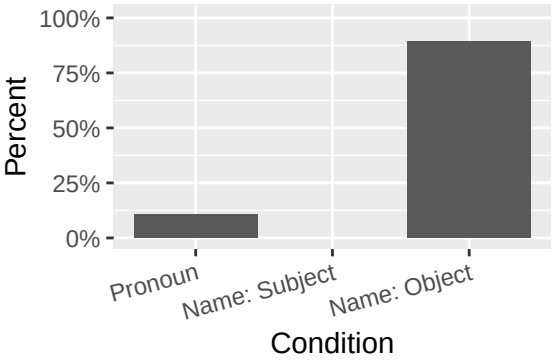

Figure 9.34: Increased listener utility RSA model: Speaker model; given an object reference and a subject-biased verb.

	state	prob
1	Pronoun	0.91577619
2	Name: Subject	0.08422381

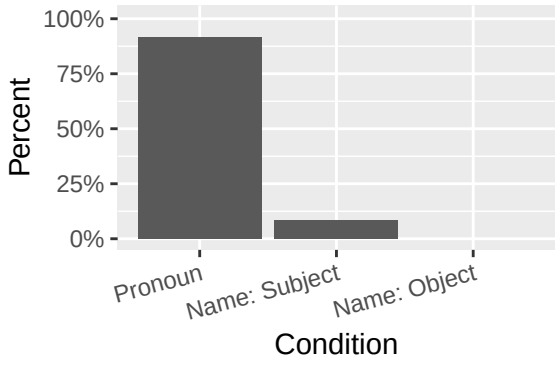

Figure 9.35: Unsophisticated speaker agent RSA model: Speaker model; given a subject reference and a subject-biased verb.

	state	prob
1	Pronoun	0.91577619
2	Name: Subject	0.08422381

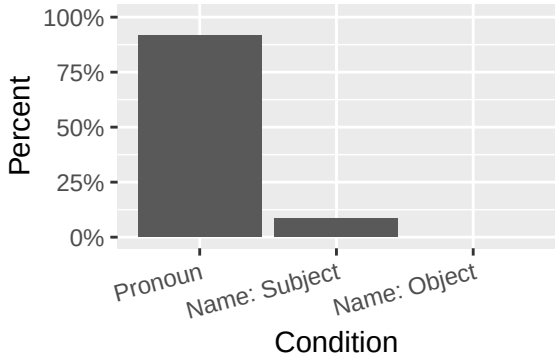

Figure 9.36: Unsophisticated speaker agent RSA model: Speaker model; given a subject reference and an object-biased verb.

Unsophisticated Speaker Agents Figures 9.35 through 9.38 show that pragmatically unsophisticated, or 'literal' speakers, who are concerned only with producing utterances which are literally true, do not base their choice of referring expression on referent predictability.

Sophisticated Speaker Agents In Figures 9.39 through 9.42, it can be seen that more pragmatically sophisticated speakers – those presumably more likely to be seen in more naturalistic and interactive settings – are, in contrast, significantly more likely to base their choice of referring expression on referent predictability. This is particularly so in the case of objects, which corresponds to Bott et al. (2018)'s observations that object reference appears to be more strongly affected by predictability than subject reference.

In the following section, I explore the possibility that the speaker grammatical bias noted by Rohde & Kehler (2014), rather than being an arbitrary feature of subject/topic reference, can be explained as a natural consequence of speech habits that are *on average* efficient developing over successive generations.

	state	prob
1	Name: Object	0.5953903
2	Pronoun	0.4046097

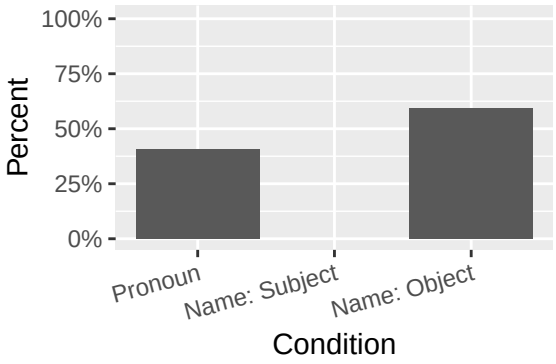

Figure 9.37: Unsophisticated speaker agent RSA model: Speaker model; given an object reference and an object-biased verb.

	state	prob
1	Name: Object	0.5953903
2	Pronoun	0.4046097

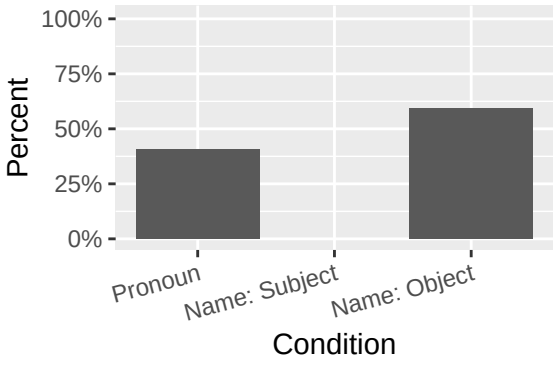

Figure 9.38: Unsophisticated speaker agent RSA model: Speaker model; given an object reference and a subject-biased verb.

	state	prob
1	Pronoun	0.91160381
2	Name: Subject	0.08839619

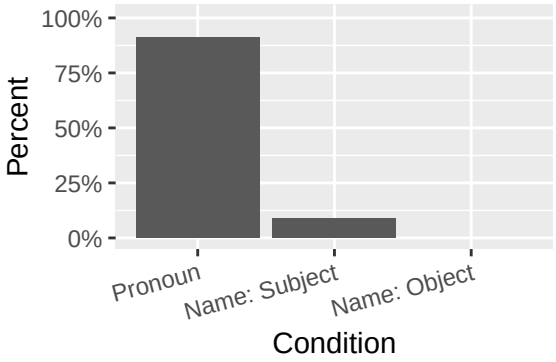

Figure 9.39: Sophisticated speaker agent RSA model: Speaker model; given a subject reference and a subject-biased verb.

	state	prob
1	Pronoun	0.8200845
2	Name: Subject	0.1799155

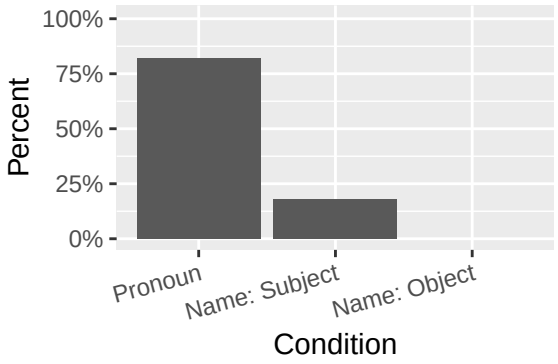

Figure 9.40: Sophisticated speaker agent RSA model: Speaker model; given a subject reference and an object-biased verb.

	state	prob
1	Name: Object	0.7170077
2	Pronoun	0.2829923

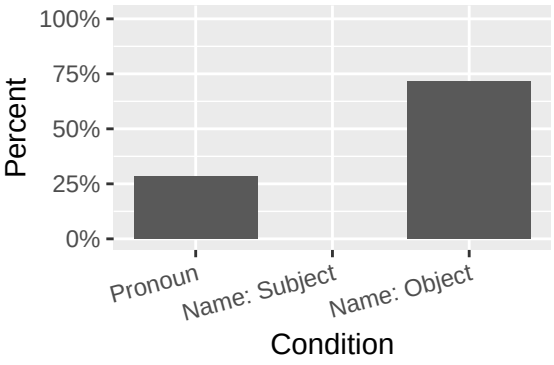

Figure 9.41: Sophisticated speaker agent RSA model: Speaker model; given an object reference and an object-biased verb.

	state	prob
1	Name: Object	0.96615896
2	Pronoun	0.03384104

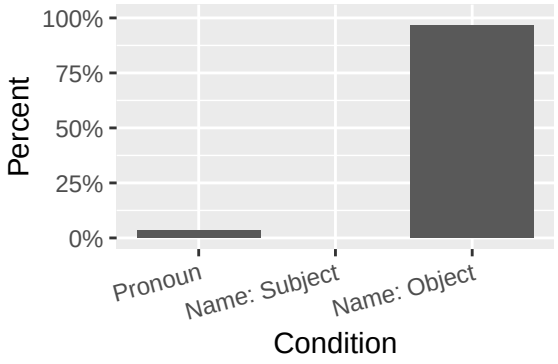

Figure 9.42: Sophisticated speaker agent RSA model: Speaker model; given an object reference and a subject-biased verb.

9.5 Emergence of a Grammatical Bias

While many accounts leave the effect of antecedent grammatical position on referring expression choice unexplained, the emergence of such a bias can be relatively straightforwardly accounted for within a probabilistic framework which assumes a speaker drive towards making utterances choices which are robust, yet efficient, *on average* (cf. Seyfarth, 2014). If one assumes that subjects, on average, are more likely to be referred to (cf. Arnold, 2008), then this would indicate that subjects are, on average, the *more predictable* referents. If speakers are primarily concerned, in this context, with *good-enough* efficiency, rather than tailoring their utterance choices to the local context, then one would expect that subjects, which are more likely to be re-mentioned, would be preferentially pronominalized – and that objects, which are *less* likely to be re-mentioned, would be pronominalized relatively infrequently.

As Arnold (2008) points out, increased likelihood of future mention correlates with exactly those properties which have been observed to influence pronominalization: length to last mention, frequency of mention, subjecthood of the antecedent, and so forth. As discussed in Section 2.1.1, there is compelling evidence that online tendencies to reduce more predictable elements may eventually become conventionalized, so that utterance choice preferences emerge which are efficient *on average*, even if not necessarily in the local context.

In the case of referring expressions, a sizable number of sentences contain *only* a subject. Separately, there is evidence that subjects are on average more likely than non-subjects to be referred back to (Arnold, 1998) – and under an efficiency-based account, this should result in subjects on average being more likely to be referred to with pronouns. If such a tendency were to become conventionalized, one would expect exactly the pattern currently observed – which is that subjects are overwhelmingly associated with pronouns, and non-subjects with longer forms.

To model how such a pattern may arise, perhaps through successive generations of speakers, I present the following model.

Model

In this model, I assume that subjects are more likely to be referred back to than objects (60% vs. 40%). α and λ are both set at 1.5:

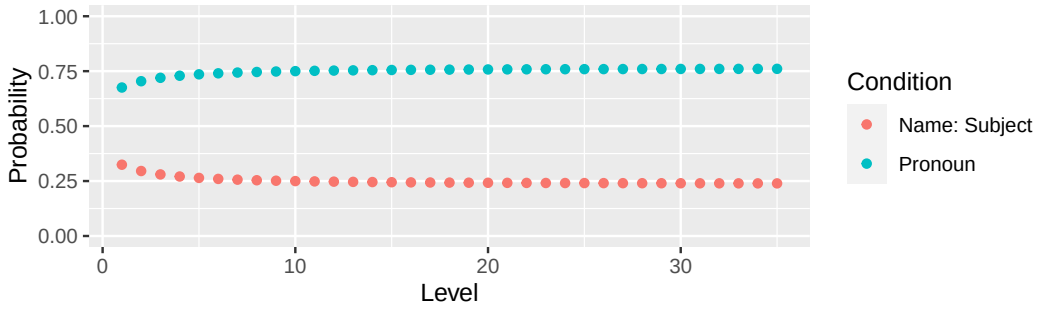

Figure 9.43: Grammatical bias RSA model: Speaker model; given a subject reference.

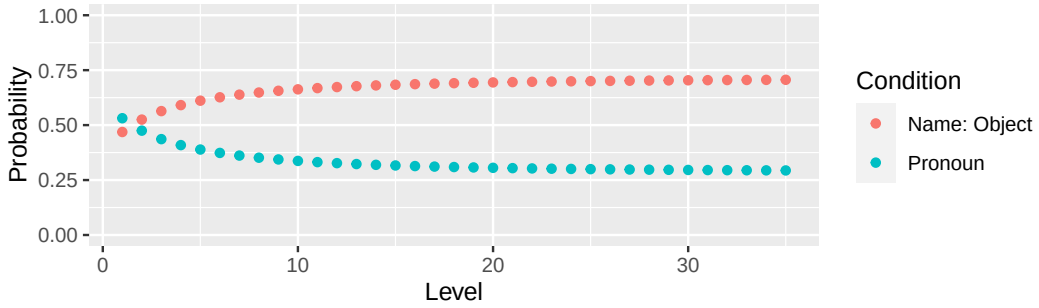

Figure 9.44: Grammatical bias RSA model: Speaker model; given an object reference.

$$P_{L_0}(\text{referent} \mid \text{expression}) \propto [\![\text{expression}]\!](\text{referent}) \cdot P(\text{referent} \mid \text{verb}) \cdot P(\text{verb}) \tag{9.8}$$

$$P_{S_1}(\text{expression} \mid \text{referent}; \alpha, \lambda, C) \propto P(\text{expression}; C, \lambda) \cdot \exp(\alpha \cdot \log P_{L_0}(\text{referent} \mid \text{expression})) \tag{9.9}$$

$$P_{L_1}(\text{referent} \mid \text{expression}) \propto P_{S_1}(\text{expression} \mid \text{referent}, \text{verb}; \alpha, \lambda, C) \cdot P(\text{referent} \mid \text{verb}) \tag{9.10}$$

$$P_{S_n}(\text{expression} \mid \text{referent}; \alpha, \lambda, C) \propto P(\text{expression}; C, \lambda) \cdot \exp(\alpha \cdot \log P_{L_{n-1}}(\text{referent} \mid \text{expression})) \tag{9.11}$$

$$P_{L_n}(\text{referent} \mid \text{expression}) \propto P_{S_n}(\text{expression} \mid \text{referent}, \text{verb}; \alpha, \lambda, C) \cdot P(\text{referent} \mid \text{verb}) \tag{9.12}$$

In Figures 9.43 and 9.44, one can see that, given a starting assumption that subjects are marginally more likely to be referred to than objects, and increasingly more sophisticated agents (or later generations), a stable pattern emerges of using predominately pronouns for subjects, and names or longer expressions for objects (even when, initially, objects are more likely to be referred to with pronouns, due to their reduced cost).

This model demonstrates how the grammatical speaker preference observed by

Rohde & Kehler (2014), among others, may emerge naturally under an efficiency-based account. Most accounts, in contrast, do not explain the emergence of such a bias, but rather seem to regard it as a relatively arbitrary feature of language use.

9.6 Summary

In summary, the Rational Speech Act framework can, at least in principle, account for all observed and postulated patterns in the relevant set of empirical results, using established machinery. Although this remains speculative, given that in some cases results remain equivocal, or may require replication, the RSA models in this chapter provide a clear predictive and explanatory framework, which may go a long way towards accounting for why certain task designs and paradigms result in stronger effects of referent predictability on referring expression choice – or, in contrast, weaker ones.

In this respect, the RSA framework appears clearly superior to the UID model, which lacks the machinery to account for the presence of detectable results in some paradigms, but not others. Similarly, as UID does not explicitly model listener utility in terms of utterance ambiguity, it is, for instance, unable to account for why effects of predictability on referring expression choice may be greater when the intended reference is more ambiguous. However, as in the previous chapter, the standard assumption of a clean channel in the RSA framework leaves some effects unaccounted for, suggesting that more routine integration of noisy channel assumptions into the RSA framework may be warranted.

Chapter 10

Conclusion

In conclusion, the data I've presented here suggests that UID in itself has unsatisfying explanatory and predictive value at the discourse level of production, and is unable to adequately predict or represent constraints on speaker redundancy, largely due to its failure to model meaning and incorporate pragmatic reasoning. UID, additionally, has largely concerned itself with alternative ways of expressing the same meaning in context – failing to account for the fact that the assumption of meaning-equivalence tends to break down, particularly at the discourse level, or with the use (or omission) of multi-word utterances.

In the last part of this thesis, I have demonstrated that in many contexts (particularly at the discourse level), it may be preferable to model speaker utterance choice using a more clearly defined probabilistic model of pragmatic reasoning and utterance interpretation, such as the Rational Speech Act (RSA) model. I have shown, following Bergen et al. (2016), that the RSA model can easily incorporate the notion of the *noisy channel* (which UID, in contrast, fails to formalize), which allows the model to derive many of the same utterance-choice predictions that UID makes[1]. This model is also, in critical contrast to UID, able to represent speakers' and listeners' reasoning about each others' beliefs and intentions, as well as background world states that may not be known to both parties. These additional factors inform both speaker utterance choice, and the message that comprehenders ultimately receive. The RSA model further introduces a clear and straightforward constraint on speaker redundancy, principally lacking in UID – by demonstrating the potential thereof to distort the intended message. Unlike UID, it is also able to clearly represent the relationship between production and comprehension. Critically, this can account for asymmetries in the relative effects of the predictability of some linguistic elements on production and comprehension – something clearly observed by Rohde & Kehler (2014) for example, but which cannot be accounted for within the UID framework.

[1]To note, some of these predictions can be derived without the additional noisy channel machinery - based on the penalty for using more costly meaning-equivalent utterances.

Building on this, I have also further demonstrated that the RSA framework benefits greatly from explicitly incorporating the information-theoretic notion of the *noisy channel*. Bergen & Goodman (2015) previously showed that the interpretation of pragmatic focus could be successfully modeled as pragmatic inference, so long as one added a formalization of the notion of a noisy channel to a standard joint reasoning RSA model. In Chapters 8-9, I show that the notion of the *noisy channel* can similarly generalize to higher-level encoding and recall phenomena – which, similarly to low-level perceptual phenomena, are likewise subject to noisy comprehender-internal processes. These processes are also presumably sensitive to how perceptually and attentionally prominent a received utterance is.

All in all, incorporating the notion of the noisy channel into a general framework of utterance comprehension and interpretation accounts straightforwardly for the fact that more attentionally and perceptually prominent utterances generally generate stronger inferences (cf. Wilson & Sperber, 2004). It further enables the RSA model to more fully account for those utterance choice phenomena that UID has been used to model – where utterances differ in cost (or the amount of speaker effort required), but the meaning that they signal is identical (cf. Chapter 4). Finally and more generally, there is increasing consensus that human speakers communicate through a noisy channel, and that this influences both production and comprehension processes. This makes a strong argument that some version of the *noisy channel* at minimum needs to be made a part of the standard RSA toolkit (similar to the notion of utterance cost, or that of joint reasoning). There is a valid criticism that the more this toolkit is (arbitrarily) expanded to account for individual phenomena, the less its explanatory and predictive value. However, in this case, I argue that the addition of an independently motivated, verified, and empirically clearly relevant modeling element is unquestionably warranted. If the standard RSA toolkit is unable to account for an intuition as basic as that of more attentionally or perceptually prominent stimuli producing stronger inferences – or even the far more general observation that intended speaker output and received listener input are not necessarily identical – then it is clearly insufficient.

10.1 Summary of Results and Implications

The empirical results and models I've introduced in this thesis cover a number of linguistic phenomena, including the interpretation of informationally redundant utterances and utterance prosody, as well as speaker choice of referring expressions. They similarly include suggestions of more explicitly expanding the notion of the noisy channel to higher-level recall phenomena, as well as brief speculative accounts (and models) of how use of referring expressions may have evolved over time. Overall, the Rational Speech Act (RSA) models presented are able to account for speaker utterance choice patterns, and listener pragmatic inferences, that are unaccounted for by both standard RSA models, and the base UID model of utterance choice/pref-

erence. Overall, I show that UID has somewhat limited predictive and explanatory power beyond lower-level production phenomena, and that higher-level phenomena are better modeled by the RSA or RSA-like models.

More specifically, in Part 1 of the thesis, I introduce informationally redundant utterances, such as *John went shopping.* ***He paid the cashier***. Such utterances, in a normal context, are pragmatically at odds with a typical comprehender's assumptions about the background world state. If *John* is a habitual cashier-payer, as is normally presumed, then there would be no need to explicitly mention his having *paid the cashier*. I show that comprehenders resolve this pragmatic anomaly by revising their beliefs about the world to accommodate such utterances – determining that *John* may not be a habitual cashier-payer, after all. I further show that modulating the attentional and perceptual prominence of the redundant utterances, with greater speaker effort likely reflecting increased speaker *intentionality* in conveying a given bit of information, likewise modulates the degree to which comprehenders are willing to revise their initial beliefs.

Contrary to the lack of penalty that UID places on speaker redundancy, the pragmatic inferences triggered by informationally redundant utterances show that unnecessary redundancy has the potential to substantially distort the speaker's intended message. This places a clear limit on how redundant a speaker may be without harming comprehension – in contrast to the idea that redundancy is nothing but helpful to the listener, and at worst is suboptimal from the perspective of reducing speaker effort. A balance between utility and conciseness is clearly important not just from the speaker's perspective, but also the listener's.

In Part 2, I test several emerging hypotheses regarding the role of referent predictability in choice of referring expression. In contrast to previous work, I test a wider variety of verbs that modulate referent predictability, further controlling for task design and experiment population. Contrary to UID predictions, I find that the verb type used to manipulate referent predictability does *not* modulate the presence or lack of an effect of predictability on referring expression choice, Rosa & Arnold (2017). I further test the hypothesis that task design and referential ambiguity are critical determinants of the presence or lack of an effect of predictability, as argued by Bott et al. (2018). I likewise find no support for this hypothesis, although I argue that a weaker version of it remains sufficiently compelling given previous findings, and further that it is principally motivated in the context of RSA-like models of pragmatic reasoning. I explore this further in Chapter 9.

Additionally, unlike previous work on the topic, I look at whether using lengthier definite descriptions as antecedents, in place of short proper names, modulates the effect of predictability on referring expression choice. UID, as well as similar theories, would predict that given a choice between a short pronoun, and a lengthy multiword expression, speakers should be even more likely to use reduced expressions for more predictable (i.e., recoverable) referents, in order to conserve effort. However, I find no evidence that referent predictability affects referring expression choice in this

case, either. I *do*, however, find evidence that speakers are overall more likely to use pronouns when referring back to lengthier antecedents - indicating that speakers *are*, in general, motivated to conserve effort. The previous result is at odds with UID predictions, and I speculate in Chapters 5 and 6 as to why this may be. I conclude that the effects of predictability on referring expression choice are at best minor, and often undetectable, at least in the task paradigms commonly used to evoke them. I further discuss the fact that UID is not equipped to account for the existing pattern of empirical findings, independently of my own results.

Finally, in Part 3, I show that unlike the UID model, the Rational Speech Act model, with the addition of the assumption of a *noisy channel*, is able to account for the pattern of empirical results in both the case of informationally redundant utterance interpretation, and in the case of referring expression choice. A standard joint reasoning RSA model is able to straightforwardly account for the emergence of *habituality* inferences in response to utterances which are redundant with respect to world knowledge. A joint reasoning model incorporating the assumption of the noisy channel is further able to account for the modulating effect of attentional or perceptual utterance prominence on inference strength (cf. Wilson & Sperber, 2004).

Similarly, the RSA model, in contrast to both the UID model, and a simple Bayesian model of referring expression interpretation (Rohde & Kehler, 2014), is largely able to account for both existing and hypothesized patterns of empirical finding, with respect to whether referent predictability affects referring expression choice. Unlike UID, the RSA model predicts, in agreement with a subset of previous empirical findings, that referential ambiguity should modulate the effect of referent predictability on referring expression choice. Similarly to the Bayesian model proposed by Rohde & Kehler (2014), it also accounts for an asymmetry in the influence of referent predictability on referring expression choice vs. interpretation (which UID offers no explanation for, and no way to model). In contrast to Rohde & Kehler (2014), however, the RSA model is able to further account both for the emergence of a conventionalized grammatical bias in referring expression choice, and for the emergence of an effect of referent predictability on expression choice in the contexts of a subset of experimental paradigms and stimulus designs. Finally, the RSA model postulates the possibility that different speakers, in different contexts, may be variably "sophisticated" with respect to choosing their utterances, in considering how well listeners may be able to interpret their intended message. This may account for the fact that more interactive and naturalistic paradigms appear to produce more robust and consistent effects of referent predictability on speaker choice of referring expression (cf. Lockridge & Brennan, 2002).

In sum, the RSA model appears to be a highly attractive alternative for modeling utterance choice, particularly above the phonetic/phonological level, or simple cases of lexical choice. It provides greater explanatory value for speaker behavior in some cases, as well as coverage of comprehension phenomena (which may in turn influence production). A critical component that has been missing in the standard RSA toolkit,

I argue, is the assumption of a noisy channel (which is assumed, but not formalized, in the UID model). At the level of lexical or phonological choice, however, where there is little to no semantic or pragmatic difference between linguistic alternatives, the UID model may still adequately account for the data.

10.2 Future Directions

The findings presented in this thesis open an avenue for future research. First, there is an empirical question of just how well the RSA model can account for the same phenomena as UID. Conceptually, the model incorporates UID assumptions. However, its adequacy in making the correct qualitative and quantitative predictions should be confirmed empirically. It would also be fruitful to explore different methods of representing the assumptions of the UID model in an RSA framework (cf. Levy, 2018).

Second, it would be highly useful to search for more production phenomena involving higher-level instances of linguistic choice (e.g., involving multi-word utterances), particularly where pragmatic reasoning may be involved, to see how well the UID and/or RSA models may account for them. Last, this work has already demonstrated, following Degen et al. (2015), the utility of models which incorporate reasoning about background world knowledge, demonstrating that pragmatically odd inferences may substantially alter a comprehender's beliefs about the world. Empirical work in pragmatics frequently fails to consider the possibility of this occurring, which may lead to inaccurate predictions about which inferences comprehenders should draw, or inadequate accounting for empirical data. This suggests that the predictions made by such models should be more widely considered.

Specifically with respect to informationally redundant utterances, it would be highly useful to run more focused production studies, in order to empirically determine exactly whether, and when, speakers may choose to be redundant, as well as the apparent purposes of speaker redundancy. More work should be done to determine whether comprehenders make similar inferences (regarding world states, for instance) when encountering other varieties of speaker redundancy, as well as the degree to which such inferences may alter the received message. It would similarly be interesting to see whether markers of increased utterance perceptual or attentional prominence, or of increased speaker effort in producing utterances, significantly influences inference strength or type in other contexts, as might be predicted by Relevance Theory (cf. Wilson & Sperber, 2004).

With respect to the use of referring expressions, there should be far more, and more varied, investigation of what might account for the current pattern of empirical results. Without further replication, or attention to differences in task and stimulus design (as well as experimental population), any attempts to account for the phenomena in question remain speculative. Primarily, it is important to run more studies using highly interactive and naturalistic designs, and rich discourse contexts (cf. Rosa &

Arnold, 2017). Other alternatives to the written passage completion paradigm, which is relatively unlikely to promote particularly sophisticated reasoning and audience design on the part of speakers, should be utilized. Studies should also be carried out in multiple languages, and by using different catalogues of referring expressions (as in German; cf. Bott et al., 2018), rather than largely restricting investigation to the respective use of pronouns and proper names in English, as is currently typical.

A lack of any observable influence of predictability on referring expression choice, in any context, would be difficult to account for in both the UID and the RSA framework – and while they may be accounted for by Rohde & Kehler (2014)'s Bayesian model, there is no principled explanation for *why* referring expression choice should not be sensitive to referent predictability. Further, the actual pattern of results may be used to either confirm the predictive validity of the RSA model, or question some of its assumptions. As demonstrated in Chapter 9, the apparent current pattern of results requires for one to make assumptions about the pragmatic sophistication of speakers (represented by various parameters and/or depth or pragmatic reasoning), which has to date received relatively little attention, as well as the influences of ambiguity and utterance cost on utterance choice. This is therefore, arguably, a rich area to mine in refining assumptions of the RSA model.

Further, I argue that more studies of large, coreference-annotated corpora should be carried out, using preferentially natural, everyday text, in order to determine whether, and to what degree, referent predictability is significantly correlated with referring expression choice. To date, two of three studies have arguably used texts that do not fully match these criteria. Tily & Piantadosi (2009) used newspaper fragments, and Modi et al. (2017) used crowdsourced descriptions of schemas that were instructed to be written in a very specific, in some respects purposefully redundant prose that would normally be considered unnatural in conversation (cf. Bower et al., 1979) – which further do not refer much to animate objects using third-person pronouns (the primary context in which one expects speakers to make a choice between 'full' and 'reduced' expressions). A further investigation that I believe would be useful in this context is an exploration of how reasonable it is to factor out those features of the text (e.g., length to last mention) that correlate with and presumably themselves determine predictability. As predictability estimates, on the part of speakers and comprehenders, do not materialize out of thin air, but are rather based on preceding dues in the discourse, factoring out those cues indiscriminately may, in effect, factor out the effect of predictability itself.

Appendix A

IRU Appendix: Stimuli

	Common Ground			Utterance
1	John often goes to the grocery store around the corner from his apartment. *ordinary*	Recently, he came home from the store with groceries. When he came in, he saw his roommate Susan in the hallway, and started talking to her about his trip to the store. As he went to the kitchen to put his groceries away, Susan went to the living room, where their roommate Peter was watching TV.	Susan said to Peter:	"John just came back from the grocery store. He paid the cashier." *habitual*
	John is typically broke, and doesn't usually pay when he goes to the grocery store. *wonky*			"John just came back from the grocery store. He got some apples." *non−habitual*
2	Mary is a journalist who often goes to restaurants after her interviews. *ordinary*	Yesterday, she went to a popular Chinese place. As she was leaving, she ran into her friend David, and they started talking about the restaurant. After they parted, David continued on his way when he suddenly ran into Sally, a mutual friend of him and Mary.	David said to Sally:	"I ran into Mary leaving that Chinese place. She ate there." *habitual*
	Mary is a journalist who often interviews restaurant waiters, but doesn't like eating out. *wonky*			"I ran into Mary leaving that Chinese place. She got to see their kitchen." *non−habitual*

	Common Ground			Utterance
3	Jim lives in a shared apartment, where it's his job to feed the dog in the evenings. *ordinary*	The other day he was feeding the dog some canned food, as his roommate Lucy came into the kitchen, and made herself a snack while chatting with him. Later in the evening, she settled down to watch TV alone with their roommate Carl.	Lucy said to Carl:	"Jim was feeding the dog earlier. He threw the can away." *habitual*
	Jim lives in a shared apartment with his dog, but can't seem to keep the common areas clean. *wonky*			"Jim was feeding the dog earlier. He had to add some medication to the food." *non−habitual*
4	Jane takes the subway all the time to get around the city. *ordinary*	Today she was entering a subway station when she ran into her friend Don, and they took the train together as they were heading in the same direction. Later that day, Don ran into Beth, Jane's sister, on the street.	Don said to Beth:	"I took a train with Jane today. She bought a subway ticket." *habitual*
	Jane takes the subway all the time to get around the city, but rarely pays for her rides. *wonky*			"I took a train with Jane today. She almost fell off the platform." *non−habitual*
5	Rick prefers to go to the local gas station to fuel up, though they overcharge him. *ordinary*	Yesterday he was fueling up there when he saw his friend Annie. They talked until he was finished, then said goodbye. After he was gone Annie went inside to buy some cigarettes, and ran into Sean, a mutual friend of theirs.	Annie said to Sean:	"Rick just fueled up here. He closed the fuel cap." *habitual*
	Rick often doesn't pay attention to what he's doing when he fuels up at the local gas station. *wonky*			"Rick just fueled up here. He got some discounted gas." *non−habitual*

	Common Ground			Utterance
6	Lisa likes to go swimming at a nearby pool after work. *ordinary*	A couple days ago she was at the pool when she saw Harvey, another regular member, and they stopped to chat. After Harvey changed and went out into the pool area, he ran into Jen, another swimmer and a friend of Lisa's.	Harvey said to Jen:	"Lisa's here to swim, too. She brought her swimsuit." *habitual*
	Lisa likes to go swimming regularly, but often forgets to bring the things she needs. *wonky*			"Lisa's here to swim, too. She came with her children." *non−habitual*
7	Brian takes the train most mornings, though the commute takes a long time. *ordinary*	Last week when he was getting on the train, he ran into his old colleague Rachel, and they chatted until Brian got off. When Rachel got to work, she saw Oliver, who also used to work with Brian.	Rachel said to Oliver:	"I saw Brian on the train this morning. He got off at his stop." *habitual*
	Brian takes the train most mornings, though he has a terrible habit of getting distracted and missing his stop. *wonky*			"I saw Brian on the train this morning. He was running late to work." *non−habitual*
8	Laura works as a software engineer at a large company. *ordinary*	A couple of days ago she was getting ready to leave for work together with her husband Dustin. After they both left the house, he ran to catch his bus, and met up with Courtney, an acquaintance who took the same bus with him every day.	Dustin said to Courtney:	"Laura was just getting ready for work with me. She grabbed her house keys." *habitual*
	Laura works as a software engineer at a large company, but is scatter-brained in everyday life. *wonky*			"Laura was just getting ready for work with me. She put on several layers of clothing." *non−habitual*

	Common Ground			Utterance
9	Bruce goes to his local medical practice every few years. *ordinary*	Yesterday after leaving the practice he ran into his friend Sarah on the street, and they stopped to catch up. After they parted, Sarah walked on and soon saw Bruce's brother Drake on the street. She stopped to say Hi.	Sarah said to Drake:	"Bruce was just leaving the medical practice. He got examined by the doctor." *habitual*
	Bruce goes to his local medical practice every few years, but usually only sees the nurse. *wonky*			"Bruce was just leaving the medical practice. He was wearing a heart rate monitor." *non−habitual*
10	Olivia has beautiful hair, and pays a lot of attention to it. *ordinary*	Today, when she was leaving the bathroom after showering, she ran into her roommate and best friend Thomas. She talked to him briefly about her hair, as she tends to do. Later that day, when their housemate Jill came home, she and Thomas started talking about Olivia.	Thomas said to Jill:	"Olivia was talking to me about washing her hair. She used shampoo." *habitual*
	Olivia has beautiful hair, although she uses a cleansing conditioner only. *wonky*			"Olivia was talking to me about washing her hair. She found some split ends." *non−habitual*
11	Jared takes skydiving courses at the local airfield, when he has free time. *ordinary*	Last week he was at the skydiving center, with his friend Stella in the same group as him. They spent the day together, and when Stella went home in the evening, she texted Jared's brother Don, who was also a good friend of hers.	Stella said to Don:	"Jared was in the skydiving course today. He jumped out of the plane." *habitual*
	Jared takes skydiving courses at the local airfield, although he is still terrified of heights. *wonky*			"Jared was in the skydiving course today. He was the first to jump." *non−habitual*

	Common Ground			Utterance
12	Amy enjoys writing letters to people she is close to, especially around holidays. *ordinary*	About two days ago, she wrote a letter to her cousin Michelle, and today she talked about it with her brother Steve. In the evening, Steve got a call from Michelle, and they started talking about family.	Steve said to Michelle:	"Amy wrote you a letter. She mailed it." *habitual*
	Amy enjoys writing letters to people she is close to, but prefers to keep them to herself rather than mailing them. *wonky*			"Amy wrote you a letter. She used really expensive stationery." *non−habitual*
13	Adam usually takes the bus to work, as the stop is a few blocks from his house. *ordinary*	Last week, after he got off the bus, he ran into Virginia, his ex-girlfriend. They stopped for a little while to catch up.	Adam said to Virginia:	"I took the bus this morning. I walked to the bus stop." *habitual*
	Adam usually takes the bus to work, but bikes to the stop although it's only a few blocks from his house. *wonky*			"I took the bus this morning. I barely had room to stand." *non−habitual*

	Common Ground			Utterance
14	Esther often goes along with her friends when they go clothes shopping, as it's something she also enjoys. *ordinary*	Today, when she was walking out of a mall after spending time with her friends, she ran into George, another old friend of hers. They decided to catch up while walking to the bus stop.	Esther said to George:	"I was out clothes shopping. I tried something on." *habitual*
	Esther often goes along with her friends when they go clothes shopping, although it bores her, and she just reads as they browse. *wonky*			"I was out clothes shopping. I came across a big sale." *non−habitual*
15	Nick enjoys making pasta dishes for his roommates, as it's an easy way to contribute to the household. *ordinary*	Yesterday he was preparing pasta in the kitchen, to sit in the fridge until a party tomorrow. When he was done and cleaning up, his roommate Clara came into the kitchen, and they started talking about his dish.	Nick said to Clara:	"I made some pasta for the meal. I boiled it in water." *habitual*
	Nick enjoys making pasta dishes for his roommates, but prefers to bake fresh pasta that doesn't need to be pre-boiled. *wonky*			"I made some pasta for the meal. I added some vegetables." *non−habitual*

	Common Ground			Utterance
16	Grace enjoys baking, as it's a great way to make new friends. *ordinary*	A few days ago she was baking a cake in her kitchen. After she had put it in the oven, her roommate Kyle came into the kitchen to make a salad for himself. They started chatting about food.	Grace said to Kyle:	"I'm baking a cake right now. I preheated the oven." *habitual*
	Grace enjoys baking, although she's terrible at following basic directions in recipes. *wonky*			"I'm baking a cake right now. I added chocolate chips to the recipe." *non−habitual*
17	Greg frequently travels by air, to see family and attend conferences. *ordinary*	Last week he flew to a conference, and met up with Helen, an old colleague he occasionally traveled with. They went to breakfast together, and started talking about their travel.	Greg said to Helen:	"I flew here. I took my cell phone on board with me." *habitual*
	Greg frequently travels by air, but hates carrying things around with him, and checks absolutely everything into the hold. *wonky*			"I flew here. I got into business class." *non−habitual*
18	Sandy usually cuts her own hair, although she has no training. *ordinary*	Two days ago, after she gave herself another haircut, she went for a walk along her street. She quickly ran into her ex, Patrick, and they stopped to catch up for a few minutes.	Sandy said to Patrick:	"I just cut my hair. I used scissors." *habitual*
	Sandy usually cuts her own hair, simply by taking a buzzer to it. *wonky*			"I just cut my hair. I cut it a bit shorter than intended." *non−habitual*

	Common Ground			Utterance
19	Henry often goes to art exhibitions, as there's an art museum a short walk from his place. *ordinary*	Last week, after going to a new photography exhibition, he encountered his friend Max on his way home. They paused on the street and chatted for a while.	Henry said to Max:	"I just went to the new photo exhibit. I looked at the photographs." *habitual*
	Henry often goes to art exhibitions, but only because his girlfriend drags him. *wonky*			"I just went to the new photo exhibit. I decided to buy a photograph." *non−habitual*
20	Helen works hard at her job, and enjoys the challenges she's given at work. *ordinary*	Today, after driving her car to work as usual, she ran into her office-mate Peter while walking into the building. They stopped briefly to say hello.	Helen said to Peter:	"I just parked my car. I locked it." *habitual*
	Helen works hard at her job, although she is incredibly scatterbrained. *wonky*			"I just parked my car. One of my tail lights has gone out." *non−habitual*
21	Gary often orders pizza at work, from a famous pizzeria nearby. *ordinary*	A few days ago, after he placed an order, his colleague Stephanie walked over to his cubicle to chat.	Gary said to Stephanie:	"I just ordered pizza. I picked the toppings." *habitual*
	Gary often orders pizza at work, but doesn't usually get to choose which type of pizza to get. *wonky*			"I just ordered pizza. I used a gift certificate." *non−habitual*

	Common Ground			Utterance
22	Julia always tries to wash the dishes after eating, to avoid annoying her roommates. *ordinary*	A few days ago, she was getting ready to go out after doing the dishes. She ran into her roommate Justin on her way out, and started talking to him.	Julia said to Justin:	"I just did the dishes. I rinsed them." *habitual*
	Julia always tries to wash the dishes after eating, but doesn't always bother to rinse them. *wonky*			"I just did the dishes. I polished them." *non−habitual*
23	Emma often borrows books from the library, as she doesn't have much spare cash to spend. *ordinary*	Last week, after going to the library, she was heading home with several books, and ran into her best friend Tim on the street. They stopped to quickly say hello.	Emma said to Tim:	"I just got some books at the library. I checked them out." *habitual*
	Emma often steals books from the library, as she doesn't have money to buy her own copies. *wonky*			"I just got some books at the library. I looked at the library's exhibit." *non−habitual*
24	Logan recently started doing his own laundry, after moving out of his parents' house. *ordinary*	Yesterday, after doing a load, he went to the living room to watch some TV. Soon his roommate Sophia came home, and asked about his day while taking off her coat.	Logan said to Sophia:	"I just did the laundry. I used detergent." *habitual*
	Logan recently started doing his own laundry, but can't get a handle even on the basics. *wonky*			"I just did the laundry. I added some softener to the wash." *non−habitual*

Appendix B

IRU Appendix: Conventionally *non-habitual* activities

Here, I present the results of my cross-experiment analysis of *non-habitual* activities.

I used the maximal converging model, with by-subject random intercepts and slopes for common ground context (*ordinary* / *wonky*) and belief measure (*pre-utterance* / *post-utterance*), by-item random intercepts and slopes for both factors and their interaction, and a by-item random slope for experiment. By-subject random slopes for the interaction were not included in the model due to lack of within-subject repeated measures in my data for the interaction. The random slope for the full by-item experiment by common ground by belief measure interaction was not included due to non-convergence.

The results are shown in Table B.1 and Figure B.1.

Table B.1: Experiment 1-3: *conventionally non-habitual* (*apple-buying*) activities analysis.

	β	SE(β)	**t**	**p**
Intercept	41.18	1.89	21.84	**<.001**
'!' vs. 'Oh yeah...'	0.84	0.72	1.17	0.2
'.' vs. Relevance Markers	1.50	0.62	2.41	**<.05**
Common Ground: Ordinary	2.00	1.97	1.02	0.3
Belief: Post-utterance	4.51	1.70	2.65	**<.05**
'!' vs. 'Oh yeah' * Common Ground	−2.11	1.05	−2.02	**<.05**
'.' vs. Relevance Markers * Common Ground	0.34	0.91	0.38	0.7
'!' vs. 'Oh yeah' * Belief	3.71	1.11	3.34	**<.001**
'.' vs. Relevance Markers * Belief	3.76	0.96	3.90	**<.001**
Common Ground * Belief	−0.73	1.43	−0.51	0.6
'!' vs. 'Oh yeah' * CG * Belief	−1.34	2.07	−0.65	0.5
'.' vs. Relevance Markers * CG * Belief	−0.64	1.79	−0.35	0.7

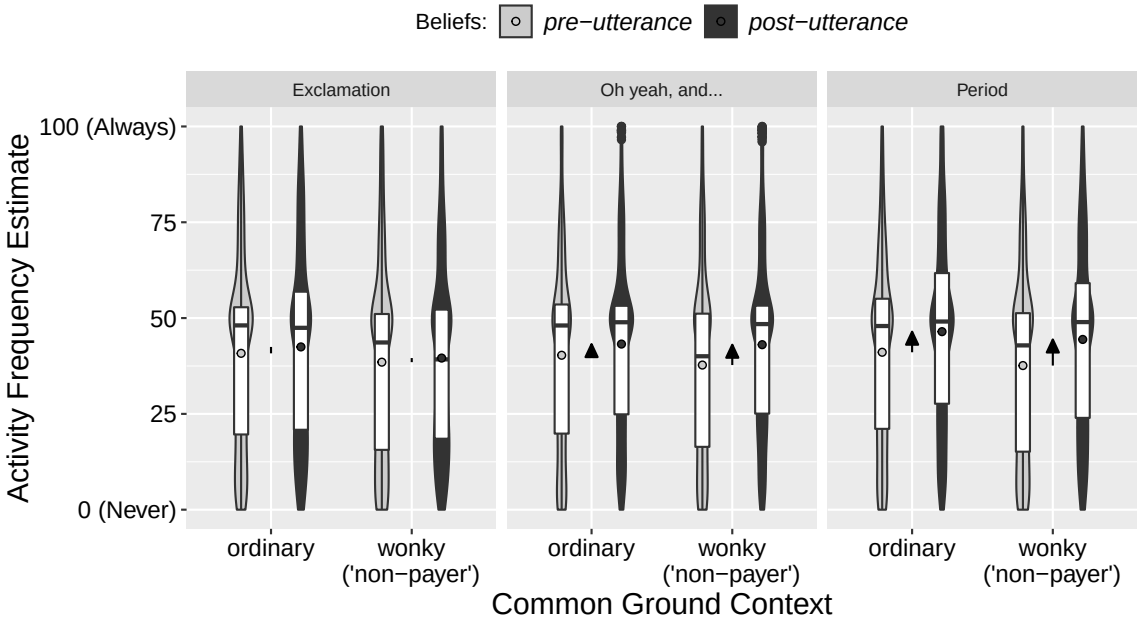

Figure B.1: Experiment 1-3: *conventionally non-habitual* (*apple-buying*) activities analysis.

Appendix C

IRU Appendix: Replicated Experiments

Here I present a previous iteration of this series of experiments, using the same design as that reported in Chapter 4, but run on separate populations (as opposed to concurrently), and with a slightly different set of stimuli. I include these results here as evidence that the effects I report are robust, replicating closely despite being run on a different population, substantial revision of the stimuli to improve naturalness, addition of filler stimuli, and a larger amount of data being collected to improve power for all relevant comparisons.

C.1 Methods

C.1.1 Participants

1200 eligible participants (1242 total), 400 per experiment, were recruited on Amazon Mechanical Turk, with the task only open to workers located in the US, and with an approval rating of $\geq 95\%$. Participants who did not report their native language, or reported their native language as other than English, were excluded (42; 3.38%), with additional participants recruited to replace them.

C.1.2 Materials

The design was identical to that reported in Chapter 4, aside from the inclusion of fillers, as each participant saw only 6 stimuli and no condition more than once, with all stimuli differing across multiple non-critical dimensions. I therefore reasoned that there was little likelihood of learning the purpose of the experiment in the course of the task, and there was risk of increased task length/tedium decreasing the likelihood of participants reading passages closely enough to pick up on relatively subtle effects.

The stimuli in the replicated experiments were constructed to minimize variation in syntactic and information structure, as well as length, between stimuli. However, this came at the cost of naturalness. Here I present a stimulus example:

(1) Original Stimulus

[1a] John often *goes to his local supermarket, as it's close by*$_{ordinary}$.	[1b] John often *doesn't pay at the supermarket, as he's typically broke*$_{wonky}$.
[2] Today he entered the apartment with his shopping bags flowing over. He ran into Susan, his best friend, and talked to her about his trip. Susan then wandered over to Peter, their roommate, who was in a different room.	
[3] Recently, he came home from the store with groceries. When he came in, he saw his roommate Susan in the hallway, and started talking to her about his trip to the store. As he went to the kitchen to put his groceries away, Susan went to the living room, where their roommate Peter was watching TV.	
[4] She commented: "John went shopping.	
[5a] He **paid the cashier**$_{\textbf{habitual}}$!	[5b] He **got some apples**$_{\textbf{non-habitual}}$!
[6] I just saw him in the living room."	

C.1.3 Procedure

The procedure was identical to that of the other experiments.

C.1.4 Measures

The same response measures as in the other experiments were used to estimate *pre-utterance beliefs* and *post-utterance beliefs*.

C.2 Results

As in the experiments reported in Chapter 4, I modeled the difference between *pre-utterance* and *post-utterance* beliefs. *Conventionally habitual* and conventionally *non-habitual* activities were modeled separately. All binary factors were effect/sum coded, and the experiment factor was Helmert coded.

C.2.1 *Conventionally habitual* activities

The regression analysis showed a significant three-way interaction between discourse marker presence, common ground context, and belief measure: there was a significantly smaller *atypicality* effect in Exp. 3 than in Experiments 1 and 2 (β =6.16,

Table C.2: Replicated Experiment 1-3: *conventionally habitual* (*cashier-paying*) activities analysis.

	β	SE(β)	**t**	**p**
Intercept	61.22	2.11	29.01	**<.001**
'!' vs. 'Oh yeah...'	1.30	1.02	1.28	0.2
'.' vs. Relevance Markers	4.34	0.88	4.92	**<.001**
Common Ground: Ordinary	38.04	3.72	10.22	**<.001**
Belief: Post-utterance	−0.46	1.42	−0.32	0.8
'!' vs. 'Oh yeah' * Common Ground	−0.87	1.82	−0.48	0.6
'.' vs. Relevance Markers * Common Ground	2.44	1.57	1.55	0.1
'!' vs. 'Oh yeah' * Belief	0.61	1.76	0.35	0.7
'.' vs. Relevance Markers * Belief	6.68	1.52	4.39	**<.001**
Common Ground * Belief	−12.66	1.27	−9.97	**<.001**
'!' vs. 'Oh yeah' * CG * Belief	0.42	3.11	0.14	0.9
'.' vs. Relevance Markers * CG * Belief	6.16	2.69	2.29	**<.05**

p <0.05), and no significant difference between Experiments 1 and 2 (β =0.42, p =0.89).

I used the maximal converging model, with by-subject random intercepts and slopes for common ground context (*ordinary / wonky*) and belief measure (*pre-utterance / post-utterance*), as well as by-item random intercepts and slopes for all factors. By-subject random slopes for the interaction were not included in the model, because I did not have any repeated measures for subjects for the interaction. By-item random slopes for the interactions were not included in the model due to nonconvergence. The model summary can be found in Table C.2. A plot illustrating the higher-order experiment by common ground by belief measure interaction can be seen in Figure C.1.

A similar analysis of the conventionally *non-habitual* activities can be found below.

C.2.2 *Conventionally non-habitual* activities

I used the maximal converging model, with by-subject random intercepts and slopes for common ground context (*ordinary / wonky*) and belief measure (*pre-utterance / post-utterance*), and by-item random intercepts and slopes for both factors and their interaction. By-subject random slopes for the interaction were not included in the model due to lack of within-subject repeated measures in my data for the interaction. The by-item random slope experiment was not included due to non-convergence.

The results are shown in Table C.3 and Figure C.2.

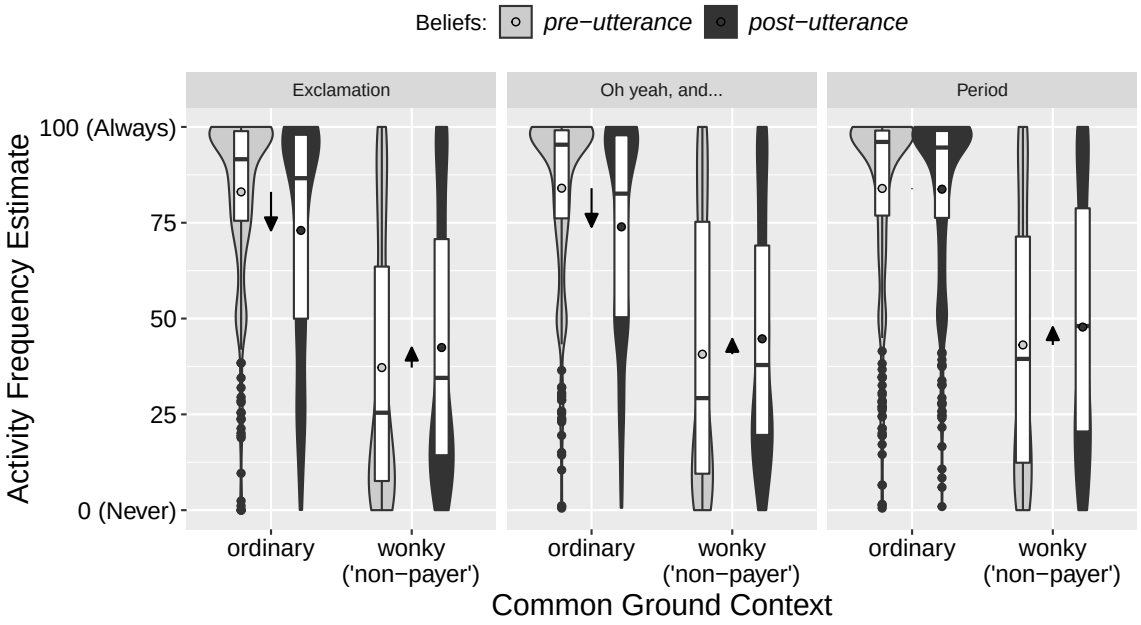

Figure C.1: Replicated Experiments 1-3: *conventionally habitual* (*cashier-paying*) activities analysis.

Table C.3: Replicated Experiments 1-3: conventionally *non-habitual* (*apple-buying*) activities analysis.

	β	SE(β)	**t**	**p**
Intercept	39.80	2.47	16.09	**<.001**
'!' vs. 'Oh yeah...'	2.25	1.01	2.22	**<.05**
'.' vs. Relevance Markers	2.33	1.02	2.28	**<.05**
Common Ground: Ordinary	2.98	2.01	1.49	0.2
Belief: Post-utterance	6.20	1.98	3.14	**<.01**
'!' vs. 'Oh yeah' * Common Ground	−0.02	1.40	−0.02	1
'.' vs. Relevance Markers * Common Ground	1.05	1.22	0.87	0.4
'!' vs. 'Oh yeah' * Belief	4.37	1.53	2.85	**<.01**
'.' vs. Relevance Markers * Belief	5.52	1.33	4.15	**<.001**
Common Ground * Belief	−4.65	1.14	−4.08	**<.001**
'!' vs. 'Oh yeah' * CG * Belief	−3.38	2.80	−1.21	0.2
'.' vs. Relevance Markers * CG * Belief	−4.28	2.43	−1.76	0.1

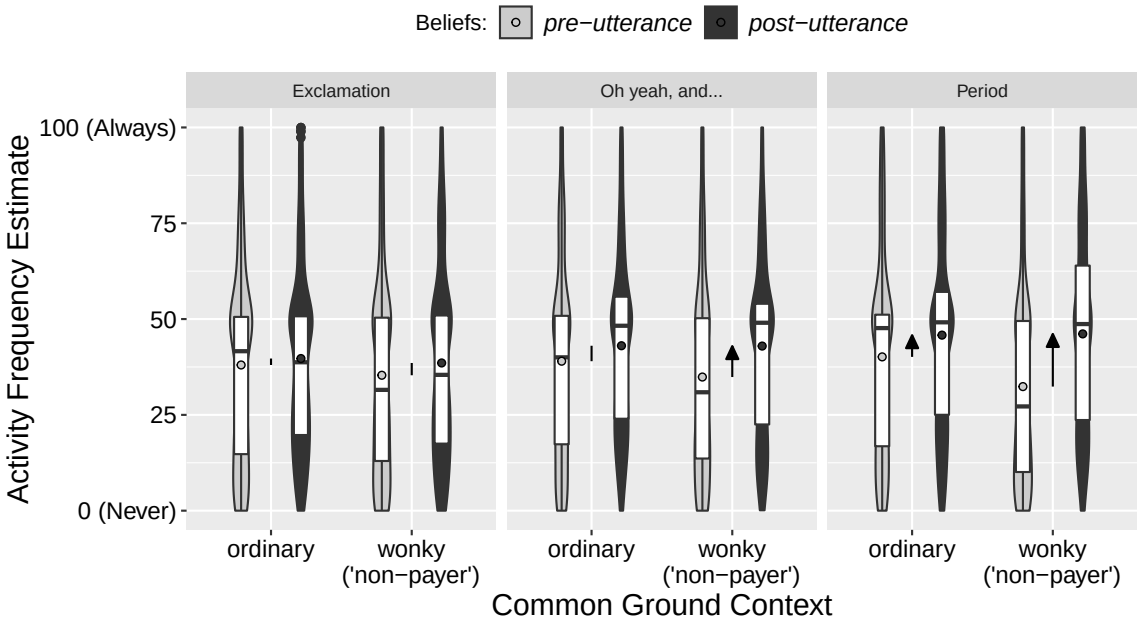

Figure C.2: Replicated Experiments 1-3: *conventionally non-habitual* (*apple-buying*) activities analysis.

C.3 Discussion

Overall, the results of these experiments were broadly replicated by those reported in Chapter 4. The only salient difference is that in the original iteration of Exp. 3, there was no measurable effect of informational redundancy on perceptions of activity typicality, while in the 'new' Exp. 3, there was a significant, but diminished effect, as I had originally predicted. The absence of a significant effect in the first iteration surprised me, and I attribute it to either chance (possibly due to fewer subjects run) or to increased prominence of the utterance in the revised stimuli. To compare:

(2) REVISED: "John just came back from the grocery store. **He paid the cashier**."

(3) ORIGINAL: "John went shopping. **He paid the cashier**. I just saw him in the living room."

The utterance in question appears more discourse-prominent in the revised version of the stimuli, as it is utterance-final (i.e., I removed the last sentence), and in general competes with fewer adjacent utterances for attention. I leave it to future work to definitively answer whether the minor change in utterance prominence does indeed eliminate the effect entirely.

Appendix D

IRU Appendix: Power Analysis

D.1 Power Analysis

```
# Power by simulation for a normally distributed continuous outcome with
# subjects, items, and residual variability

# Population parameters
# mu: underlying mean of the outcome in the reference group
# betaN: effect size of predictor or interaction
# sdItem: sd of random effect at the item level
# sdSubject: sd of random effect at the subject level
# sdResid: sd of residual error

# Design parameters
# nSubjects: number of subjects in simulation
# nIterations: number of iterations in simulation

rm(list=ls())
setwd("~/Shared/informationally-redundant-utterances/code/")
library(Hmisc)
library(rms)
library(lme4)
library(lmerTest)
fnPower <-
  function(mu,beta1,beta2,beta3,beta4,beta5,beta6,beta7,beta8,beta9,
           beta10,beta11,sdItem,sdSubject,sdResid,nSubjects,nIterations,
           dots=TRUE){
    start.time <- Sys.time()
    if(dots) cat("Simulations (",nIterations,
```

```
          ") \n---|-- 1 --|-- 2 --|-- 3 --|-- 4 --| -- 5 \n",
          sep="")
# objects to store pvalue, beta, and standard error from each iteration
# of simulation
    pVals <- betaVals <- seVals <- matrix(NA,nrow=nIterations,ncol=11)
# build design matrices
    m <- matrix(NA,nrow=nSubjects*4,ncol=7)
    colnames(m) <- c("worker","exp.alike","exp.diff","story","condition",
                     "context","slider")
    m[,1] <- rep(1:nSubjects,each=4)
    m[,2] <- rep(c(-0.5,0.5,0),each=4,length.out=length(m[,2]))
    m[,3] <- rep(c(-0.3333333,-0.3333333,0.6666667),each=4,
                 length.out=length(m[,3]))
    i <- 1
    while(i < (length(m[,4]))){ m[i:(i+3),4] <- sample(1:24,size=4,
                                                       replace=FALSE)
    i <- i+4 }
    m[,5] <- rep(c(-0.5, 0.5))
    m[,6] <- rep(c(-0.5, 0.5),each=2)
    # i <- 1 # (when testing for-loop)
    for(i in 1:nIterations){

  # draw random effects
      itemRE <- rnorm(24,0,sdItem)
      subjRE <- rnorm(nSubjects,0,sdSubject)
      residRE <- rnorm(nrow(m),0,sdResid)

  # create outcome
      y <- mu + beta1*m[,2] + beta2*m[,3] + beta3*m[,6] + beta4*m[,5] +
        beta5*m[,2]*m[,6] + beta6*m[,3]*m[,6] + beta7*m[,2]*m[,5] +
        beta8*m[,3]*m[,5] + beta9*m[,6]*m[,5] + beta10*m[,2]*m[,6]*m[,5] +
        beta11*m[,3]*m[,6]*m[,5] + itemRE[m[,4]] + subjRE[m[,1]] + residRE
      m[,7] <- y
      dm <- as.data.frame(m)
      dm$worker <- as.factor(dm$worker)
      dm$story <- as.factor(dm$story)
  # fit model, store p-value, beta and standard error
      o <- lmer(slider ~ exp.alike + exp.diff + context + condition +
                  exp.alike:context + exp.diff:context +
                  exp.alike:condition + exp.diff:condition +
                  context:condition + exp.alike:context:condition +
                  exp.diff:context:condition + (1|story) + (1|worker),dm)
      pVals[i,] <- coef(summary(o))[2:12,5]
      betaVals[i,] <- coef(summary(o))[2:12,1]
```

```
      seVals[i,] <- coef(summary(o))[2:12,2]

      if(dots) cat(".",sep="")
      if(dots && i %% 50 == 0) cat(i,"\n")

    }

    if(dots) cat("\nSimulation Run Time:",round(difftime(Sys.time(),
                start.time,units="hours"),3)," Hours \n")

# calculate power
    powerOut <- apply(pVals,2,function(x) length(x[x<0.05])/length(x))
    return(list(power=powerOut,p=pVals,beta=betaVals,se=seVals))
  }


# calibrate by setting betas = 0; histograms should all look level,
# with about 5% of results significant by chance
outCalibrate <- fnPower(mu=61.0444,beta1=0,beta2=0,beta3=0,beta4=0,
                        beta5=0,beta6=0,beta7=0,beta8=0,beta9=0,beta10=0,
                        beta11=0,sdItem=10.138,sdSubject=9.285,
                        sdResid=21.839,nSubjects=1200,nIterations=10000,
                        dots=TRUE)
outCalibrate$power
hist(outCalibrate$p[,1],main="beta1",xlab="p value")
hist(outCalibrate$p[,2],main="beta2",xlab="p value")
hist(outCalibrate$p[,3],main="beta3",xlab="p value")
hist(outCalibrate$p[,4],main="beta4",xlab="p value")
hist(outCalibrate$p[,5],main="beta5",xlab="p value")
hist(outCalibrate$p[,6],main="beta6",xlab="p value")
hist(outCalibrate$p[,7],main="beta7",xlab="p value")
hist(outCalibrate$p[,8],main="beta8",xlab="p value")
hist(outCalibrate$p[,9],main="beta9",xlab="p value")
hist(outCalibrate$p[,10],main="beta10",xlab="p value")
hist(outCalibrate$p[,11],main="beta11",xlab="p value")

nSubjects <- seq(1200,2400,100)

st <- c()
for(i in 1:length(nSubjects)){
  st[[paste("subj",nSubjects[i])]] <-
    fnPower(mu=61.2216,beta1=1.3018,beta2=4.3355,beta3=38.0383,
            beta4=-0.4559,beta5=-0.8669,beta6=2.4416,beta7=0.6141,
            beta8=6.6776,beta9=-12.6572,beta10=0.4207,beta11=6.1601,
```

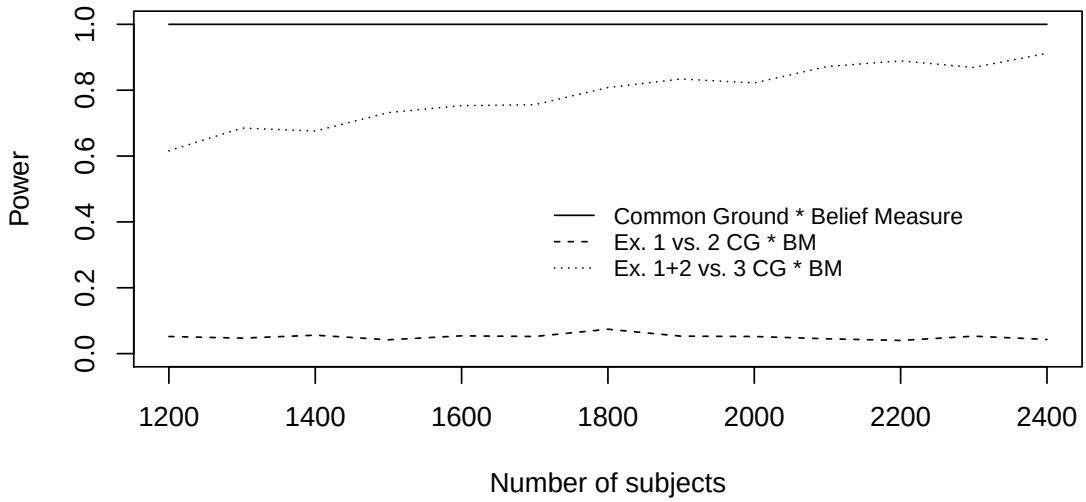

Figure D.1: Power curves for effects of interest

```
            sdItem=10.138,sdSubject=9.285,sdResid=21.839,
            nSubjects=nSubjects[i],nIterations=1000,dots=TRUE)
}
```

D.2 Plot

Figure D.1 shows a plot of the power curves for the critical common ground by belief measure interaction, a comparison of that effect across Experiments 1 and 2, and another comparison across Experiment 3 and Experiments 1/2 (as a group). I expect a robustly replicable common ground by belief measure interaction (power = 1.00 at all sample sizes), no significant difference between Experiments 1 and 2 (power $\leq$ 0.1 at all sample sizes), and a significant difference between Experiment 3 and Experiments 1/2 (power $\geq$ 0.85 at a sample size of 2100 or more).

Appendix E

RE Appendix: Stimuli

Every experimental list contained 10 IC stimuli, 12 ToP stimuli, and 10 fillers. Half of each set would take the form of definite descriptions, and half would take the form of proper names. The particular entities used as subjects or objects in the stimuli varied from list to list, to ensure that the particular entity type minimally influenced item response. A sample NP1/NP2 stimuli skeleton (to be used with any given IC verbs/ToP verb & possession pairs) is provided here, as well as a list of relevant IC, ToP, and filler verbs.

E.1 Sample Stimuli Skeleton

All definite descriptions below were preceded by the definite article "the" in the stimuli lists. Which NP was subject/object was varied between lists, and no NP was consistently associated with any particular verb or verb-object pair.

item	type	NP1 / NP2	NP2 / NP1
1	IC	short man with gray hair	woman wearing a red jacket
2	IC	woman from work	tall man wearing glasses
3	IC	man wearing an orange shirt	woman with curly hair
4	IC	thin woman with dyed hair	man from a few houses down
5	IC	man from class	large woman with bleached hair
6	IC	Heather	Zack
7	IC	Roy	Ellen
8	IC	Emily	Anthony
9	IC	Bill	Kate
10	IC	Ashley	Paul
11	ToP	man wearing sunglasses	short woman from the store
12	ToP	tall woman with short hair	man wearing a white shirt

item	type	NP1 / NP2	NP2 / NP1
13	ToP	man from the office	woman wearing shorts
14	ToP	woman wearing a purple shirt	thin man with blond hair
15	ToP	large man with styled hair	woman from school
16	ToP	woman from upstairs	man with brown hair
17	ToP	Laura	Nick
18	ToP	Josh	Tiffany
19	ToP	Crystal	Brian
20	ToP	Bob	Kelly
21	ToP	Alice	Mike
22	ToP	Rob	Debbie
23	filler	Scott	short woman wearing a green jacket
24	filler	woman from across town	Joe
25	filler	tall man wearing a yellow shirt	Kristen
26	filler	Stephanie	man with red hair
27	filler	Ben	thin woman with long hair
28	filler	woman with black hair	Eric
29	filler	large man from the market	Beth
30	filler	Rebecca	man from the gym
31	filler	Ken	woman wearing a blue jacket
32	filler	woman wearing a hat	Craig

E.2 Verbs

E.2.1 Implicit Causality

IC-1 Verbs:

1. aggravated
2. amused
3. apologized to
4. charmed
5. confessed to
6. amazed
7. annoyed
8. bored

9. offended
10. deceived

IC-2 Verbs:

1. blames
2. congratulated
3. detests
4. envies
5. helped
6. assisted
7. comforted
8. corrected
9. fears
10. hates

E.2.2 Transfer-of-Possession Verbs:

1. transferred a payment to / accepted a payment from
2. lent a bike to / borrowed a bike from
3. threw a ball to / caught a ball from
4. paid a week's salary to / earned a week's salary from
5. gave a gift to / got a gift from
6. bequeathed the family treasures to / inherited the family treasures from
7. traded some Magic cards to / acquired some Magic cards from
8. sold some furniture to / bought some furniture from
9. returned the amount due to / collected the amount due from
10. extended insurance to / gained insurance from
11. handed some cake to / grabbed some cake from
12. rented an apartment to / leased an apartment from

13. supplied a fake ID to / obtained a fake ID from
14. shipped some clothes to / ordered some clothes from
15. loaned a book to / procured a book from
16. offered a ticket to / purchased a ticket from
17. sent a letter to / received a letter from
18. leased a car to / rented a car from
19. ceded the plot of land to / reclaimed the plot of land from
20. passed the DVD to / seized the DVD from
21. smuggled some contraband to / smuggled some contraband from
22. presented the package to / snatched the package from
23. slipped an envelope to / took an envelope from
24. yielded the trophy to / wrested the trophy from

Item 21 was changed in Experiment 2 as it was judged insufficiently felicitous:

21. provided some information to / gathered some information from

E.2.3 Filler

1. chatted with
2. saw
3. worked with
4. watched
5. studied with
6. ran into
7. stood next to
8. waited to see
9. went to visit
10. split some fries with

Appendix F

IRU Model Appendix: Code

F.1 Base RSA

```
// Current activity state
// the activity being described at this point in time either took
// place, or didn't
var state = ["happened","didn't happen"]

// State priors
// assume highly predictable/habitual activity
// with a 90% chance of occurring, for purpose of demonstration
var statePrior = function() {
  categorical([0.9, 0.1], state)
}

// Utterances
// choice of 4 utterances; prosody not modeled separately as affects
// only one variant
var utterance = ['oh yeah','exclamation','plain','(...)']

// Utterance cost
// (rough estimate of number of constituents + extra for
// articulatory effort)
var cost = {
  "oh yeah": 4.5,
  "exclamation": 4,
  "plain": 3,
```

```
  "(...)": 0
}

// Meaning
// literal meaning of all overt utterances is that activity happened.
// literal meaning of null "utterance" is consistent with all activity
// states
var meaning = function(utt,state) {
  utt === "oh yeah" ? state === "happened" :
  utt === "exclamation" ? state === "happened" :
  utt === "plain" ? state === "happened" :
  utt === "(...)" ? true :
  true
}

// Speaker optimality (maximizing utility)
var alpha = 4

// Speaker optimality (minimizing cost)
var lambda = 1

// Utterance prior
// utterance prior determined by utterance cost, as defined above
var utterancePrior = function() {
  var uttProbs = map(function(u) {return Math.exp(-lambda * cost[u])},
   utterance)
  return categorical(uttProbs, utterance)
}

// Literal listener
var literalListener = mem(function(utterance) {
  return Infer({method: 'enumerate', model: function() {
    var state = statePrior()
    condition(meaning(utterance,state))
    return state
  }})
})

// Speaker
var speaker = mem(function(state) {
  return Infer({method: 'enumerate', model: function() {
```

```
    var utterance = utterancePrior()
    factor(alpha * literalListener(utterance).score(state))
    return utterance
  }})
})

// Pragmatic listener
var pragmaticListener = function(utterance) {
  return Infer({method: 'enumerate', model: function() {
    var state = statePrior()
    observe(speaker(state),utterance)
    return state
  }})
}
```

F.2 hRSA

```
// Function to create discrete beta distribution
var roundTo3 = function(x){
  return Math.round(x * 1000) / 1000
}

var granularity = 100
var midBins = map(function(x) {roundTo3(x/granularity +
                  1/(2*granularity))}, _.range(0,granularity))

var DiscreteBeta = cache(function(a, b){
  Infer({model: function(){
    categorical({
      vs:midBins,
      ps:map(function(x){
        Math.exp(Beta({a, b}).score(x))
      }, midBins)
    })
  }})
})

// Is this a world in which the conventionally habitual activity is
// habitual (presumed cashier-payer) or non-habitual (presumed
```

```
// non-payer)? (mostly for demonstration)
var world = ["wonky","ordinary"]

// Assume uniform likelihood
var worldPrior = function() {
  categorical([0.5, 0.5], world)
}

// Habituality priors
// beta distributions fit to empirical priors
var habitualityPrior = function(world) {
  world === "ordinary" ? sample(DiscreteBeta(beta_high_a, beta_high_b)) :
  world === "wonky" ? sample(DiscreteBeta(beta_low_a, beta_low_b)) :
  true
}

// Current activity state
// the activity being described at this point in time either took place,
// or didn't
var state = ["happened","didn't happen"]

// State priors
// whether the activity took place is dependent on prior likelihood
var statePrior = function(habituality) {
  flip(habituality) ? state[0] : state[1]
}

// Utterances
// choice of 4 utterances; prosody not modeled separately as affects
// only one variant
var utterance = ['oh yeah','exclamation','plain','(...)']

// Utterance cost
// (rough estimate of number of constituents + extra for articulatory )
// effort)
var cost = {
  "oh yeah": 4.5,
  "exclamation": 4,
  "plain": 3,
  "(...)": 0
}
```

```
// Meaning
// literal meaning of all overt utterances is that activity happened.
// literal meaning of null "utterance" is consistent with all activity
// states
var meaning = function(utt,state) {
  utt === "oh yeah" ? state === "happened" :
  utt === "exclamation" ? state === "happened" :
  utt === "plain" ? state === "happened" :
  utt === "(...)" ? true :
  true
}

// Speaker optimality (maximizing utility)
var alpha = 4

// Speaker optimality (minimizing cost)
var lambda = 1

// Utterance prior
// utterance prior determined by utterance cost, as defined above
var utterancePrior = function() {
  var uttProbs = map(function(u) {return Math.exp(-lambda * cost[u])},
   utterance)
  return categorical(uttProbs, utterance)
}

// Literal listener
var literalListener = mem(function(utterance, habituality) {
  return Infer({method: 'enumerate', model: function() {
    var state = statePrior(habituality)
    condition(meaning(utterance,state))
    return state
  }})
})

// Speaker
var speaker = mem(function(state, habituality) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior()
    factor(alpha * literalListener(utterance, habituality).score(state))
```

```
    return utterance
  }})
})

// Pragmatic listener
// assume high-habit world for demonstration
var pragmaticListener = function(utterance, info) {
  return Infer({method: 'enumerate', model: function() {
    var world = "ordinary"
    var habituality = habitualityPrior(world)
    var state = statePrior(habituality)
    observe(speaker(state, habituality),utterance)
    info === "both" ? {state: state, habituality: habituality} :
    info === "state" ? state :
    info === "habituality" ? habituality :
    true
  }})
}
```

F.3 Noisy Channel hRSA

```
// Function to create discrete beta distribution
var roundTo3 = function(x){
  return Math.round(x * 1000) / 1000
}

var granularity = 100
var midBins = map(function(x) {roundTo3(x/granularity +
                  1/(2*granularity))}, _.range(0,granularity))

var DiscreteBeta = cache(function(a, b){
  Infer({model: function(){
    categorical({
      vs:midBins,
      ps:map(function(x){
        Math.exp(Beta({a, b}).score(x))
      }, midBins)
    })
  }})
```

```
})

// Is this a world in which the conventionally habitual activity is
// habitual (presumed cashier-payer) or non-habitual (presumed
// non-payer)? (mostly for demonstration)
var world = ["wonky","ordinary"]

// Assume uniform likelihood
var worldPrior = function() {
  categorical([0.5, 0.5], world)
}

// Habituality priors
// beta distributions fit to empirical priors
var habitualityPrior = function(world) {
  world === "ordinary" ? sample(DiscreteBeta(beta_high_a, beta_high_b)) :
  world === "wonky" ? sample(DiscreteBeta(beta_low_a, beta_low_b)) :
  true
}

// Current activity state
// the activity being described at this point in time either took place,
// or didn't
var state = ["happened","didn't happen"]

// State priors
// whether the activity took place is dependent on prior likelihood
var statePrior = function(habituality) {
  flip(habituality) ? state[0] : state[1]
}

// Utterances (intended)
// choice of 4 utterances; prosody not modeled separately as affects
// only one variant
var utterance = ['oh yeah','exclamation','plain','(...)']

// Utterance cost
// rough estimate of relative costs given number of constituents +
// articulatory effort
var cost = {
  "oh yeah": 4.5,
```

```
  "exclamation": 4,
  "plain": 3,
  "(...)": 0
}

// Utterances (recalled/attended to)
// assume that utterance most likely to be recalled as itself, but also
// has non-trivial likelihood of being recalled as 'neighboring'
// utterance (with markers for plain utterance; vice versa; no
// utterance for "plain" utterance; and vice versa).
// alternately, this can be conceptualized as listener's belief of what
// the speaker *intended* to say - but unclear if below is best way to
// represent that
var oh_yeah = [0.97,0.01,0.02,0.0001]
var exclamation = [0.01,0.97,0.02,0.0001]
var plain = [0.02,0.02,0.95,0.01]
var zero = [0.0001,0.0001,0.01,0.99]

var utterance_r = function(u_i) {
  u_i === "oh yeah" ? categorical(oh_yeah, utterance) :
  u_i === "exclamation" ? categorical(exclamation, utterance) :
  u_i === "plain" ? categorical(plain, utterance) :
  u_i === "(...)" ? categorical(zero, utterance) :
  true
}

// Confusion matrix for purpose of summing up probabilities
var utterance_r_prob = function(u_i, u_r) {
  u_i === "oh yeah" ? oh_yeah[_.indexOf(utterance, u_r)] :
  u_i === "exclamation" ? exclamation[_.indexOf(utterance, u_r)] :
  u_i === "plain" ? plain[_.indexOf(utterance, u_r)] :
  u_i === "(...)" ? zero[_.indexOf(utterance, u_r)] :
  true
}

// Meaning
// literal meaning of all overt utterances is that activity happened.
// literal meaning of null "utterance" is consistent with all activity
// states
var meaning = function(utterance,state) {
  utterance === "oh yeah" ? state === "happened" :
```

```
  utterance === "exclamation" ? state === "happened" :
  utterance === "plain" ? state === "happened" :
  utterance === "(...)" ? true :
  true
}

// Speaker optimality (maximizing utility)
var alpha = 4

// Speaker optimality (minimizing cost)
var lambda = 1

// Utterance prior
// utterance prior determined by utterance cost, as defined above
var utterancePrior = function() {
  var uttProbs = map(function(u) {return Math.exp(-lambda * cost[u])},
   utterance)
  return categorical(uttProbs, utterance)
}

// Utterance posterior P(u_r | u_i)
var utterancePosterior = mem(function(u_r) {
  Infer({method: 'enumerate', model: function() {
    var u_i = utterancePrior()
    condition(u_r === utterance_r(u_i))
    return u_i
  }})
})

// Literal listener
var literalListener = mem(function(u_r, habituality) {
  return Infer({method: 'enumerate', model: function() {
    var state = statePrior(habituality)
    var u_i = sample(utterancePosterior(u_r))
    condition(meaning(u_i, state))
    return state
  }})
})

// Expected utilities
var get_EUs = function(u_i, state, habituality){
```

```
  var EUs = sum(map(function(u_r) {
    utterance_r_prob(u_i, u_r) *
      literalListener(u_r, habituality).score(state)
    }, utterance))
  return EUs
}

// Speaker
var speaker = mem(function(state, habituality) {
  return Infer({method: 'enumerate', model: function() {
    var u_i = utterancePrior()
    var EUs = get_EUs(u_i, state, habituality)
    factor(alpha * EUs)
    return u_i
  }})
})

// Pragmatic listener
// assume particular world for demonstration
var pragmaticListener = function(u_r, info) {
  return Infer({method: 'enumerate', model: function() {
    var world = world_type
    var habituality = habitualityPrior(world)
    var state = statePrior(habituality)
    var u_i = sample(utterancePosterior(u_r))
    observe(speaker(state, habituality),u_i)
    info === "both" ? {state: state, habituality: habituality} :
    info === "state" ? state :
    info === "habituality" ? habituality :
    true
  }})
}
```

Appendix G

RE Model Appendix: Code

G.1 Rohde & Kehler (2014) Bayesian model

```
var speakerUtterancePrior = function(referent) {
  referent == "Subject" ? (flip(0.8) ? "Pronoun" : "Noun") :
  referent == "Object" ? (flip(0.2) ? "Pronoun" : "Noun") :
  true
}

var speaker = function(referent) {
        return Infer({method: 'enumerate', model:function() {
                var utterance = speakerUtterancePrior(referent)
                return utterance
        }})}

var listener = function(utterance, subjectPrior) {
        return Infer({method: 'enumerate', model:function() {
                var referent = flip(subjectPrior) ? "Subject" : "Object"
                observe(speaker(referent), utterance)
                return referent
        }})}
```

G.2 RSA Model: Ambiguous

```
var referent = ["Subject", "Object"]

var referentPrior = function(verb) {
  verb == "S-Bias" ? categorical([0.75, 0.25], referent) :
  verb == "O-Bias" ? categorical([0.25, 0.75], referent) :
  true
}

var alpha = 1

var lambda = 1

var utterance = ["Pronoun", "Name: Subject", "Name: Object"]

var cost = {
  "Pronoun": 1,
  "Name: Subject": 2,
  "Name: Object": 2
}

var utterancePrior = function(referent) {
  var uttProbs = map(function(u) {return Math.exp(-lambda * cost[u])},
   utterance)
  var subjBias = map2(function(x, y) {return x * y}, [0.8, 0.2, 0.2],
   uttProbs)
  var objBias = map2(function(x, y) {return x * y}, [0.2, 0.8, 0.8],
   uttProbs)
  referent == "Subject" ? (categorical(subjBias, utterance)) :
  referent == "Object" ? (categorical(objBias, utterance)) :
  true
}

var meaning = function(utterance, referent) {
  utterance == "Name: Subject" ? referent == "Subject" :
  utterance == "Name: Object" ? referent == "Object" :
  utterance == "Pronoun" ? true :
  true
}
```

```
var literalListener = function(utterance, verb) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior(verb)
    condition(meaning(utterance, referent))
    return referent
  }})
}

var speaker = function(referent, verb) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * literalListener(utterance, verb).score(referent))
    return utterance
  }})
}

var pragmaticListener = function(utterance, verb) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior(verb)
    observe(speaker(referent, verb), utterance)
    return referent
  }})
}
```

G.3 RSA Model: Non-Ambiguous

```
var referent = ["Subject", "Object"]

var referentPrior = function(verb) {
  verb == "S-Bias" ? categorical([0.75, 0.25], referent) :
  verb == "O-Bias" ? categorical([0.25, 0.75], referent) :
  true
}

var alpha = 1

var lambda = 1
```

```
var utterance = ["Pronoun: Subject", "Pronoun: Object",
 "Name: Subject", "Name: Object"]

var cost = {
  "Pronoun: Subject": 1,
  "Pronoun: Object": 1,
  "Name: Subject": 2,
  "Name: Object": 2
}

var utterancePrior = function(referent) {
  var uttProbs = map(function(u) {return Math.exp(-lambda * cost[u])},
   utterance)
  var subjBias = map2(function(x, y) {return x * y}, [0.8,0.8, 0.2,0.2],
   uttProbs)
  var objBias = map2(function(x, y) {return x * y}, [0.2,0.2, 0.8,0.8],
   uttProbs)
  referent == "Subject" ? (categorical(subjBias, utterance)) :
  referent == "Object" ? (categorical(objBias, utterance)) :
  true
}

var meaning = function(utterance, referent) {
  (utterance == "Name: Subject" || utterance == "Pronoun: Subject") ?
    referent == "Subject" :
  (utterance == "Name: Object" || utterance == "Pronoun: Object") ?
    referent == "Object" :
  true
}

var literalListener = function(utterance, verb) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior(verb)
    condition(meaning(utterance, referent))
    return referent
  }})
}

var speaker = function(referent, verb) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
```

```
    factor(alpha * literalListener(utterance, verb).score(referent))
    return utterance
  }})
}

var pragmaticListener = function(utterance, verb) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior(verb)
    observe(speaker(referent, verb), utterance)
    return referent
  }})
}
```

G.4 RSA Model: Non-Ambigous Noisy Channel

```
var referent = ["Subject", "Object"]

var referentPrior = function(verb) {
  verb == "S-Bias" ? categorical([0.75, 0.25], referent) :
  verb == "O-Bias" ? categorical([0.25, 0.75], referent) :
  true
}

var alpha = 1

var lambda = 1

var utterance = ["Pronoun: Subject", "Pronoun: Object",
"Name: Subject", "Name: Object"]

var cost = {
  "Pronoun: Subject": 1,
  "Pronoun: Object": 1,
  "Name: Subject": 2,
  "Name: Object": 2
}

var utterancePrior = function(referent) {
  var uttProbs = map(function(u) {return Math.exp(-lambda * cost[u])},
```

```
    utterance)
  var subjBias = map2(function(x, y) {return x * y}, [0.8,0.8, 0.2,0.2],
    uttProbs)
  var objBias = map2(function(x, y) {return x * y}, [0.2,0.2, 0.8,0.8],
    uttProbs)
  referent == "Subject" ? (categorical(subjBias, utterance)) :
  referent == "Object" ? (categorical(objBias, utterance)) :
  true
}

// Confusion matrix
var pronounS = [0.98,0.0001,0.02,0.0001]
var nameS = [0.0001,0.999,0.0001,0.0001]
var pronounO = [0.02,0.0001,0.98,0.0001]
var nameO = [0.0001,0.0001,0.0001,0.999]

// Perceived utterance
var utterance_p = function(u_i) {
  u_i === "Pronoun: Subject" ? categorical(pronounS, utterance) :
  u_i === "Name: Subject" ? categorical(nameS, utterance) :
  u_i === "Pronoun: Object" ? categorical(pronounO, utterance) :
  u_i === "Name: Object" ? categorical(nameO, utterance) :
  true
}

// Confusion matrix for purpose of summing up probabilities
var utterance_p_prob = function(u_i, u_p) {
  u_i === "Pronoun: Subject" ? pronounS[_.indexOf(utterance, u_p)] :
  u_i === "Name: Subject" ? nameS[_.indexOf(utterance, u_p)] :
  u_i === "Pronoun: Object" ? pronounO[_.indexOf(utterance, u_p)] :
  u_i === "Name: Object" ? nameO[_.indexOf(utterance, u_p)] :
  true
}

// Utterance posterior P(u_p | u_i)
var utterancePosterior = mem(function(u_p, referent) {
  Infer({method: 'enumerate', model: function() {
    var u_i = utterancePrior(referent)
    condition(u_p === utterance_p(u_i))
    return u_i
  }})
```

```
})

var meaning = function(utterance, referent) {
  (utterance == "Name: Subject" || utterance == "Pronoun: Subject") ?
 referent == "Subject" :
  (utterance == "Name: Object" || utterance == "Pronoun: Object") ?
 referent == "Object" :
  true
}

var literalListener = function(u_p, verb) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior(verb)
    var u_i = sample(utterancePosterior(u_p, referent))
    condition(meaning(u_i, referent))
    return referent
  }})
}

// Expected utilities
var get_EUs = function(u_i, referent, verb){
  var EUs = sum(map(function(u_p) {
    utterance_p_prob(u_i, u_p) *
      literalListener(u_p, verb).score(referent)
    }, utterance))
  return EUs
}

var speaker = function(referent, verb) {
  return Infer({method: 'enumerate', model: function() {
    var u_i = utterancePrior(referent)
    var EUs = get_EUs(u_i, referent, verb)
    factor(alpha * EUs)
    return u_i
  }})
}

var pragmaticListener = function(u_p, verb) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior(verb)
    var u_i = sample(utterancePosterior(u_p, referent))
```

```
    observe(speaker(referent, verb), u_i)
    return referent
  }})
}
```

G.5 RSA Model: Free Completion

```
var referent = ["Subject", "Object"]

var referentPrior = function(verb) {
  verb == "S-Bias" ? categorical([0.55, 0.45], referent) :
  verb == "O-Bias" ? categorical([0.45, 0.55], referent) :
  true
}

var alpha = 1

var lambda = 1

var utterance = ["Pronoun", "Name: Subject", "Name: Object"]

var cost = {
  "Pronoun": 1,
  "Name: Subject": 2,
  "Name: Object": 2
}

var utterancePrior = function(referent) {
  var uttProbs = map(function(u) {return Math.exp(-lambda * cost[u])},
   utterance)
  var subjBias = map2(function(x, y) {return x * y}, [0.8, 0.2, 0.2],
   uttProbs)
  var objBias = map2(function(x, y) {return x * y}, [0.2, 0.8, 0.8],
   uttProbs)
  referent == "Subject" ? (categorical(subjBias, utterance)) :
  referent == "Object" ? (categorical(objBias, utterance)) :
  true
}
```

```
var meaning = function(utterance, referent) {
  utterance == "Name: Subject" ? referent == "Subject" :
  utterance == "Name: Object" ? referent == "Object" :
  utterance == "Pronoun" ? true :
  true
}

var literalListener = function(utterance, verb) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior(verb)
    condition(meaning(utterance, referent))
    return referent
  }})
}

var speaker = function(referent, verb) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * literalListener(utterance, verb).score(referent))
    return utterance
  }})
}

var pragmaticListener = function(utterance, verb) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior(verb)
    observe(speaker(referent, verb), utterance)
    return referent
  }})
}
```

G.6 RSA Model: Constrained Completion

```
var referent = ["Subject", "Object"]

var referentPrior = function(verb) {
  verb == "S-Bias" ? categorical([0.75, 0.25], referent) :
  verb == "O-Bias" ? categorical([0.25, 0.75], referent) :
  true
```

```
}

var alpha = 1

var lambda = 1

var utterance = ["Pronoun", "Name: Subject", "Name: Object"]

var cost = {
  "Pronoun": 1,
  "Name: Subject": 2,
  "Name: Object": 2
}

var utterancePrior = function(referent) {
  var uttProbs = map(function(u) {return Math.exp(-lambda * cost[u])},
   utterance)
  var subjBias = map2(function(x, y) {return x * y}, [0.8, 0.2, 0.2],
   uttProbs)
  var objBias = map2(function(x, y) {return x * y}, [0.2, 0.8, 0.8],
   uttProbs)
  referent == "Subject" ? (categorical(subjBias, utterance)) :
  referent == "Object" ? (categorical(objBias, utterance)) :
  true
}

var meaning = function(utterance, referent) {
  utterance == "Name: Subject" ? referent == "Subject" :
  utterance == "Name: Object" ? referent == "Object" :
  utterance == "Pronoun" ? true :
  true
}

var literalListener = function(utterance, verb) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior(verb)
    condition(meaning(utterance, referent))
    return referent
  }})
}
```

```
var speaker = function(referent, verb) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * literalListener(utterance, verb).score(referent))
    return utterance
  }})
}

var pragmaticListener = function(utterance, verb) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior(verb)
    observe(speaker(referent, verb), utterance)
    return referent
  }})
}
```

G.7 RSA Model: Reduced Audience Design

```
var referent = ["Subject", "Object"]

var referentPrior = function(verb) {
  verb == "S-Bias" ? categorical([0.75, 0.25], referent) :
  verb == "O-Bias" ? categorical([0.25, 0.75], referent) :
  true
}

var alpha = 0.01

var lambda = 1

var utterance = ["Pronoun", "Name: Subject", "Name: Object"]

var cost = {
  "Pronoun": 1,
  "Name: Subject": 2,
  "Name: Object": 2
}

var utterancePrior = function(referent) {
```

```
  var uttProbs = map(function(u) {return Math.exp(-lambda * cost[u])},
   utterance)
  var subjBias = map2(function(x, y) {return x * y}, [0.8, 0.2, 0.2],
   uttProbs)
  var objBias = map2(function(x, y) {return x * y}, [0.2, 0.8, 0.8],
   uttProbs)
  referent == "Subject" ? (categorical(subjBias, utterance)) :
  referent == "Object" ? (categorical(objBias, utterance)) :
  true
}

var meaning = function(utterance, referent) {
  utterance == "Name: Subject" ? referent == "Subject" :
  utterance == "Name: Object" ? referent == "Object" :
  utterance == "Pronoun" ? true :
  true
}

var literalListener = function(utterance, verb) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior(verb)
    condition(meaning(utterance, referent))
    return referent
  }})
}

var speaker = function(referent, verb) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * literalListener(utterance, verb).score(referent))
    return utterance
  }})
}

var pragmaticListener = function(utterance, verb) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior(verb)
    observe(speaker(referent, verb), utterance)
    return referent
  }})
}
```

G.8 RSA Model: Increased Audience Design

```
var referent = ["Subject", "Object"]

var referentPrior = function(verb) {
  verb == "S-Bias" ? categorical([0.75, 0.25], referent) :
  verb == "O-Bias" ? categorical([0.25, 0.75], referent) :
  true
}

var alpha = 1.25

var lambda = 1

var utterance = ["Pronoun", "Name: Subject", "Name: Object"]

var cost = {
  "Pronoun": 1,
  "Name: Subject": 2,
  "Name: Object": 2
}

var utterancePrior = function(referent) {
  var uttProbs = map(function(u) {return Math.exp(-lambda * cost[u])},
   utterance)
  var subjBias = map2(function(x, y) {return x * y}, [0.8, 0.2, 0.2],
   uttProbs)
  var objBias = map2(function(x, y) {return x * y}, [0.2, 0.8, 0.8],
   uttProbs)
  referent == "Subject" ? (categorical(subjBias, utterance)) :
  referent == "Object" ? (categorical(objBias, utterance)) :
  true
}

var meaning = function(utterance, referent) {
  utterance == "Name: Subject" ? referent == "Subject" :
  utterance == "Name: Object" ? referent == "Object" :
  utterance == "Pronoun" ? true :
  true
}
```

```
var literalListener = function(utterance, verb) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior(verb)
    condition(meaning(utterance, referent))
    return referent
  }})
}

var speaker = function(referent, verb) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * literalListener(utterance, verb).score(referent))
    return utterance
  }})
}

var pragmaticListener = function(utterance, verb) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior(verb)
    observe(speaker(referent, verb), utterance)
    return referent
  }})
}
```

G.9 RSA Model: Reduced Agent Sophistication

```
var referent = ["Subject", "Object"]

%var referentPrior = function(verb) {
  verb == "S-Bias" ? categorical([0.75, 0.25], referent) :
  verb == "O-Bias" ? categorical([0.25, 0.75], referent) :
  true
}

var alpha = 1

var lambda = 1
```

```
var utterance = ["Pronoun", "Name: Subject", "Name: Object"]

var cost = {
  "Pronoun": 1,
  "Name: Subject": 2,
  "Name: Object": 2
}

var utterancePrior = function(referent) {
  var uttProbs = map(function(u) {return Math.exp(-lambda * cost[u])},
   utterance)
  var subjBias = map2(function(x, y) {return x * y}, [0.8, 0.2, 0.2],
   uttProbs)
  var objBias = map2(function(x, y) {return x * y}, [0.2, 0.8, 0.8],
   uttProbs)
  referent == "Subject" ? (categorical(subjBias, utterance)) :
  referent == "Object" ? (categorical(objBias, utterance)) :
  true
}

var meaning = function(utterance, referent) {
  utterance == "Name: Subject" ? referent == "Subject" :
  utterance == "Name: Object" ? referent == "Object" :
  utterance == "Pronoun" ? true :
  true
}

var speaker_0 = function(referent, verb) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    condition(meaning(utterance, referent))
    return utterance
  }})
}

var literalListener = function(utterance, verb) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior(verb)
    condition(meaning(utterance, referent))
    return referent
  }})
```

```
}

var speaker_1 = function(referent, verb) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * literalListener(utterance, verb).score(referent))
    return utterance
  }})
}

var pragmaticListener = function(utterance, verb) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior(verb)
    observe(speaker(referent, verb), utterance)
    return referent
  }})
}
```

G.10 RSA Model: Increased Agent Sophistication

```
var referent = ["Subject", "Object"]

var referentPrior = function(verb) {
  verb == "S-Bias" ? categorical([0.75, 0.25], referent) :
  verb == "O-Bias" ? categorical([0.25, 0.75], referent) :
  true
}

var alpha = 1

var lambda = 1

var utterance = ["Pronoun", "Name: Subject", "Name: Object"]

var cost = {
  "Pronoun": 1,
  "Name: Subject": 2,
  "Name: Object": 2
}
```

```
var utterancePrior = function(referent) {
  var uttProbs = map(function(u) {return Math.exp(-lambda * cost[u])},
   utterance)
  var subjBias = map2(function(x, y) {return x * y}, [0.8, 0.2, 0.2],
   uttProbs)
  var objBias = map2(function(x, y) {return x * y}, [0.2, 0.8, 0.8],
   uttProbs)
  referent == "Subject" ? (categorical(subjBias, utterance)) :
  referent == "Object" ? (categorical(objBias, utterance)) :
  true
}

var meaning = function(utterance, referent) {
  utterance == "Name: Subject" ? referent == "Subject" :
  utterance == "Name: Object" ? referent == "Object" :
  utterance == "Pronoun" ? true :
  true
}

var literalListener = function(utterance, verb) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior(verb)
    condition(meaning(utterance, referent))
    return referent
  }})
}

var speaker = function(referent, verb) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * literalListener(utterance, verb).score(referent))
    return utterance
  }})
}

var pragmaticListener = function(utterance, verb) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior(verb)
    observe(speaker(referent, verb), utterance)
    return referent
```

```
  }})
}

var speaker_2 = function(referent, verb) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * pragmaticListener(utterance, verb).score(referent))
    return utterance
  }})
}

var pragmaticListener_2 = function(utterance, verb) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior(verb)
    observe(speaker_2(referent, verb), utterance)
    return referent
  }})
}
```

G.11 RSA Model: Grammatical Bias

```
// possible referents: subject, object
var referent = ["Subject", "Object"]

// if subject bias, subject more likely to be referred to; same for
// object
var referentPrior = function() {
  categorical([0.6, 0.4], referent)
}

// utility optimization parameter
var alpha = 1.5

// cost optimization parameter
var lambda = 1.5

// possible utterances: ambiguous pronoun, names referring to subject
// or object
var utterance = ["Pronoun", "Name: Subject", "Name: Object"]
```

```
var cost = {
  "Pronoun": 1,
  "Name: Subject": 2,
  "Name: Object": 2
}

// utterance prior based on cost
var utterancePrior = function() {
  var uttProbs = map(function(u) {return Math.exp(-lambda * cost[u])},
   utterance)
  return categorical(uttProbs, utterance)
}

// nameS/O can only be used to refer to subject/object; pronoun can
// refer to both
var meaning = function(utterance, referent) {
  utterance == "Name: Subject" ? referent == "Subject" :
  utterance == "Name: Object" ? referent == "Object" :
  utterance == "Pronoun" ? true :
  true
}

// P_L0(referent|utterance) \propto
// [[utterance]](referent) * P(referent|verb) * P(verb)
var literalListener = mem(function(utterance) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior()
    condition(meaning(utterance, referent))
    return referent
  }})
})

// P_S1(utterance|referent; alpha, lambda, C) \propto
// P(utterance; C, lambda) * exp(alpha * logL0(referent|utterance))
var speaker_1 = mem(function(referent) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * literalListener(utterance).score(referent))
    return utterance
  }})
```

```
})

// P_L1(referent|utterance) \propto
// P_S1(utterance|referent, verb; alpha, lambda, C) * P(referent|verb)
var pragmaticListener_1 = mem(function(utterance) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior()
    observe(speaker_1(referent), utterance)
    return referent
  }})
})

var speaker_2 = mem(function(referent) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * pragmaticListener_1(utterance).score(referent))
    return utterance
  }})
})

var pragmaticListener_2 = mem(function(utterance) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior()
    observe(speaker_2(referent), utterance)
    return referent
  }})
})

var speaker_3 = mem(function(referent) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * pragmaticListener_2(utterance).score(referent))
    return utterance
  }})
})

var pragmaticListener_3 = mem(function(utterance) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior()
    observe(speaker_3(referent), utterance)
    return referent
```

```
  }})
})

var speaker_4 = mem(function(referent) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * pragmaticListener_3(utterance).score(referent))
    return utterance
  }})
})

var pragmaticListener_4 = mem(function(utterance) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior()
    observe(speaker_4(referent), utterance)
    return referent
  }})
})

var speaker_5 = mem(function(referent) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * pragmaticListener_4(utterance).score(referent))
    return utterance
  }})
})

var pragmaticListener_5 = mem(function(utterance) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior()
    observe(speaker_5(referent), utterance)
    return referent
  }})
})

var speaker_6 = mem(function(referent) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * pragmaticListener_5(utterance).score(referent))
    return utterance
  }})
```

```
})

var pragmaticListener_6 = mem(function(utterance) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior()
    observe(speaker_6(referent), utterance)
    return referent
  }})
})

var speaker_7 = mem(function(referent) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * pragmaticListener_6(utterance).score(referent))
    return utterance
  }})
})

var pragmaticListener_7 = mem(function(utterance) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior()
    observe(speaker_7(referent), utterance)
    return referent
  }})
})

var speaker_8 = mem(function(referent) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * pragmaticListener_7(utterance).score(referent))
    return utterance
  }})
})

var pragmaticListener_8 = mem(function(utterance) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior()
    observe(speaker_8(referent), utterance)
    return referent
  }})
})
```

```
var speaker_9 = mem(function(referent) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * pragmaticListener_8(utterance).score(referent))
    return utterance
  }})
})

var pragmaticListener_9 = mem(function(utterance) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior()
    observe(speaker_9(referent), utterance)
    return referent
  }})
})

var speaker_10 = mem(function(referent) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * pragmaticListener_9(utterance).score(referent))
    return utterance
  }})
})

var pragmaticListener_10 = mem(function(utterance) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior()
    observe(speaker_10(referent), utterance)
    return referent
  }})
})

var speaker_11 = mem(function(referent) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * pragmaticListener_10(utterance).score(referent))
    return utterance
  }})
})
```

```
var pragmaticListener_11 = mem(function(utterance) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior()
    observe(speaker_11(referent), utterance)
    return referent
  }})
})

var speaker_12 = mem(function(referent) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * pragmaticListener_11(utterance).score(referent))
    return utterance
  }})
})

var pragmaticListener_12 = mem(function(utterance) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior()
    observe(speaker_12(referent), utterance)
    return referent
  }})
})

var speaker_13 = mem(function(referent) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * pragmaticListener_12(utterance).score(referent))
    return utterance
  }})
})

var pragmaticListener_13 = mem(function(utterance) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior()
    observe(speaker_13(referent), utterance)
    return referent
  }})
})

var speaker_14 = mem(function(referent) {
```

```
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * pragmaticListener_13(utterance).score(referent))
    return utterance
  }})
})

var pragmaticListener_14 = mem(function(utterance) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior()
    observe(speaker_14(referent), utterance)
    return referent
  }})
})

var speaker_15 = mem(function(referent) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * pragmaticListener_14(utterance).score(referent))
    return utterance
  }})
})

var pragmaticListener_15 = mem(function(utterance) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior()
    observe(speaker_15(referent), utterance)
    return referent
  }})
})

var speaker_16 = mem(function(referent) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * pragmaticListener_15(utterance).score(referent))
    return utterance
  }})
})

var pragmaticListener_16 = mem(function(utterance) {
  return Infer({method: 'enumerate', model: function() {
```

```
    var referent = referentPrior()
    observe(speaker_16(referent), utterance)
    return referent
  }})
})

var speaker_17 = mem(function(referent) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * pragmaticListener_16(utterance).score(referent))
    return utterance
  }})
})

var pragmaticListener_17 = mem(function(utterance) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior()
    observe(speaker_17(referent), utterance)
    return referent
  }})
})

var speaker_18 = mem(function(referent) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * pragmaticListener_17(utterance).score(referent))
    return utterance
  }})
})

var pragmaticListener_18 = mem(function(utterance) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior()
    observe(speaker_18(referent), utterance)
    return referent
  }})
})

var speaker_19 = mem(function(referent) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
```

```
    factor(alpha * pragmaticListener_18(utterance).score(referent))
    return utterance
  }})
})

var pragmaticListener_19 = mem(function(utterance) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior()
    observe(speaker_19(referent), utterance)
    return referent
  }})
})

var speaker_20 = mem(function(referent) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * pragmaticListener_19(utterance).score(referent))
    return utterance
  }})
})

var pragmaticListener_20 = mem(function(utterance) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior()
    observe(speaker_20(referent), utterance)
    return referent
  }})
})

var speaker_21 = mem(function(referent) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * pragmaticListener_20(utterance).score(referent))
    return utterance
  }})
})

var pragmaticListener_21 = mem(function(utterance) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior()
    observe(speaker_21(referent), utterance)
```

```
    return referent
  }})
})

var speaker_22 = mem(function(referent) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * pragmaticListener_21(utterance).score(referent))
    return utterance
  }})
})

var pragmaticListener_22 = mem(function(utterance) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior()
    observe(speaker_22(referent), utterance)
    return referent
  }})
})

var speaker_23 = mem(function(referent) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * pragmaticListener_22(utterance).score(referent))
    return utterance
  }})
})

var pragmaticListener_23 = mem(function(utterance) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior()
    observe(speaker_23(referent), utterance)
    return referent
  }})
})

var speaker_24 = mem(function(referent) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * pragmaticListener_23(utterance).score(referent))
    return utterance
```

```
  }})
})

var pragmaticListener_24 = mem(function(utterance) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior()
    observe(speaker_24(referent), utterance)
    return referent
  }})
})

var speaker_25 = mem(function(referent) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * pragmaticListener_24(utterance).score(referent))
    return utterance
  }})
})

var pragmaticListener_25 = mem(function(utterance) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior()
    observe(speaker_25(referent), utterance)
    return referent
  }})
})

var speaker_26 = mem(function(referent) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * pragmaticListener_25(utterance).score(referent))
    return utterance
  }})
})

var pragmaticListener_26 = mem(function(utterance) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior()
    observe(speaker_26(referent), utterance)
    return referent
  }})
```

```
})

var speaker_27 = mem(function(referent) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * pragmaticListener_26(utterance).score(referent))
    return utterance
  }})
})

var pragmaticListener_27 = mem(function(utterance) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior()
    observe(speaker_27(referent), utterance)
    return referent
  }})
})

var speaker_28 = mem(function(referent) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * pragmaticListener_27(utterance).score(referent))
    return utterance
  }})
})

var pragmaticListener_28 = mem(function(utterance) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior()
    observe(speaker_28(referent), utterance)
    return referent
  }})
})

var speaker_29 = mem(function(referent) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * pragmaticListener_28(utterance).score(referent))
    return utterance
  }})
})
```

```
var pragmaticListener_29 = mem(function(utterance) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior()
    observe(speaker_29(referent), utterance)
    return referent
  }})
})

var speaker_30 = mem(function(referent) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * pragmaticListener_29(utterance).score(referent))
    return utterance
  }})
})

var pragmaticListener_30 = mem(function(utterance) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior()
    observe(speaker_30(referent), utterance)
    return referent
  }})
})

var speaker_31 = mem(function(referent) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * pragmaticListener_30(utterance).score(referent))
    return utterance
  }})
})

var pragmaticListener_31 = mem(function(utterance) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior()
    observe(speaker_31(referent), utterance)
    return referent
  }})
})
```

```
var speaker_32 = mem(function(referent) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * pragmaticListener_31(utterance).score(referent))
    return utterance
  }})
})

var pragmaticListener_32 = mem(function(utterance) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior()
    observe(speaker_32(referent), utterance)
    return referent
  }})
})

var speaker_33 = mem(function(referent) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * pragmaticListener_32(utterance).score(referent))
    return utterance
  }})
})

var pragmaticListener_33 = mem(function(utterance) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior()
    observe(speaker_33(referent), utterance)
    return referent
  }})
})

var speaker_34 = mem(function(referent) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * pragmaticListener_33(utterance).score(referent))
    return utterance
  }})
})

var pragmaticListener_34 = mem(function(utterance) {
```

```
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior()
    observe(speaker_34(referent), utterance)
    return referent
  }})
})

var speaker_35 = mem(function(referent) {
  return Infer({method: 'enumerate', model: function() {
    var utterance = utterancePrior(referent)
    factor(alpha * pragmaticListener_34(utterance).score(referent))
    return utterance
  }})
})

var pragmaticListener_35 = mem(function(utterance) {
  return Infer({method: 'enumerate', model: function() {
    var referent = referentPrior()
    observe(speaker_35(referent), utterance)
    return referent
  }})
})
```

List of Figures

List of Tables

Bibliography

Ariel, M. (1990). *Accessing Noun-Phrase Antecedents*. Routledge.

Arnold, B. F., Hogan, D. R., Colford, J. M., & Hubbard, A. E. (2011). Simulation methods to estimate design power: An overview for applied research. *BMC Medical Research Methodology*, *11*.

Arnold, J. E. (1998). *Reference Form and Discourse Patterns. Doctoral dissertation, Stanford University*.

Arnold, J. E. (2001). The effect of thematic roles on pronoun use and frequency of reference continuation. *Discourse Processes*, *31*, 137–162.

Arnold, J. E. (2008). Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes*, *23*, 495–527.

Arnold, J. E., & Griffin, Z. M. (2007). The effect of additional characters on choice of referring expression: Everyone counts. *Journal of Memory and Language*, *56*, 521–536.

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*, 31–56.

Aylett, M., & Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *The Journal of the Acoustical Society of America*, *119*, 3048–3058.

Baker, R., Gill, A., & Cassell, J. (2008). Reactive redundancy and listener comprehension in direction-giving. In *Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue* (pp. 37–45). Columbus, Ohio.

Bard, E., & Aylett, M. (1999). The dissociation of deaccenting, givenness, and syntactic role in spontaneous speech. *Proceedings of the XIVth international congress of phonetic science*, (p. 1753–1756).

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2020). *lme4: Linear Mixed-Effects Models using 'Eigen' and S4*. URL: `https://CRAN.R-project.org/package=lme4` r package version 1.1-23.

Bell, A., Brenier, J., Gregory, M. L., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational english. *Journal of Memory and Language*, *60*, 92–111.

Benz, A., & Rooij, R. (2007). Optimal assertions, and what they implicate. a uniform game theoretic approach. *Topoi*, *26*, 63–78.

Bergen, L., & Goodman, N. D. (2015). The Strategic Use of Noise in Pragmatic Reasoning. *Topics in Cognitive Science*, *7*, 336–350.

Bergen, L., Goodman, N. D., & Levy, R. (2012). That's what she (could have) said: How alternative utterances affect language use. *Proceedings of the Thirty-Fourth Annual Conference of the Cognitive Science Society*, (pp. 120–125).

Bergen, L., Levy, R., & Goodman, N. (2016). Pragmatic reasoning through semantic inference. *Semantics and Pragmatics*, *9*.

Bott, O., Solstad, T., & Prysłopsk, A. (2018). Implicit causality affects the choice of anaphoric form. *24th Architectures and Mechanisms for Language Processing Conference*, .

Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology*, *11*, 177–220.

Brown, P. M., & Dell, G. S. (1987). Adapting production to comprehension: The explicit mention of instruments. *Cognitive Psychology*, *19*, 441–472.

Brown-Schmidt, S., & Tanenhaus, M. K. (2008). Real-time investigation of referential domains in unscripted conversation: A targeted language game approach. *Cognitive Science*, *32*, 643–684.

Chafe, W. L. (1994). *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. University of Chicago Press.

Cohen, P. R. (1978). *On knowing what to say: planning speech acts*. Doctoral Dissertation, University of Toronto.

Cohen Priva, U. (2008). Using information content to predict phone deletion. In *Proceedings of the 27th West Coast Conference on Formal Linguistics*.

Cohen Priva, U. (2015). Informativity affects consonant duration and deletion rates. *Laboratory Phonology*, *6*, 243–278.

Corbett, A., & Chang, F. (1983). Pronoun disambiguation: Accessing potential antecedents. *Memory & Cognition*, *11*, 283–294.

Dahl, Ö., & Fraurud, K. (1996). Animacy in grammar and discourse. *Pragmatics and beyond. New series*, *38*, 47–64.

Davidson, D. (1974). Belief and the basis of meaning. *Synthese*, *27*, 309–323.

Davies, C., & Katsos, N. (2010). Over-informative children: production/comprehension asymmetry or tolerance to pragmatic violations? *Lingua*, *120*, 1956–1972.

Davies, C., & Katsos, N. (2013). Are speakers and listeners 'only moderately gricean'? an empirical response to engelhardt et al. (2006). *Journal of Pragmatics*, *49*, 78–106.

Degen, J., Franke, M., & Gerhard, J. (2012). Optimal reasoning about referential expressions. In *Proceedings of SemDIAL*.

Degen, J., Franke, M., & Jäger, G. (2013). Cost-based pragmatic inference about referential expressions. *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, (pp. 376–381).

Degen, J., & Tanenhaus, M. K. (2015). Processing scalar implicature: A constraint-based approach. *Cognitive Science*, *39*, 667–710.

Degen, J., & Tanenhaus, M. K. (2016). Availability of Alternatives and the Processing of Scalar Implicatures: A Visual World Eye-Tracking Study. *Cognitive Science*, *40*, 172–201.

Degen, J., Tessler, M. H., & Goodman, N. D. (2015). Wonky worlds: Listeners revise world knowledge when utterances are odd. *Proceedings of the 37th Annual Conference of the Cognitive Science Society*, (pp. 548–553).

Delignette-Muller, M.-L., Dutang, C., & Siberchicot, A. (2020). *fitdistrplus: Help to Fit of a Parametric Distribution to Non-Censored or Censored Data*. URL: `https://CRAN.R-project.org/package=fitdistrplus` r package version 1.1-1.

Dell, G. S., & Brown, P. M. (1991). Bridges between psychology and linguistics. chapter Mechanisms for listener-adaptation in language production: Limiting the role of the"model of the listener.". (pp. 105–129). San Diego: Academic Press.

Engelhardt, P. E., Bailey, K. G. D., & Ferreira, F. (2006). Do speakers and listeners observe the gricean maxim of quantity? *Journal of Memory and Language*, *54*, 554–573.

Fillmore, C. J. (2006). Frame semantics. In *Encyclopedia of Language and Linguistics* (pp. 613–620).

Fowler, C., & Housum, J. (1987). Talkers signaling of 'new' and 'old' words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*, *26*, 489–504.

Fowler, C. A., Levy, E. T., & Brown, J. M. (1997). Reductions of spoken words in certain discourse contexts. *Journal of Memory and Language*, *37*, 24–40.

Frank, A., & Jaeger, T. F. (2008). Speaking rationally: Uniform information density as an optimal strategy for language production. In *The 30th Annual Meeting of the Cognitive Science Society (CogSci08)* (pp. 939–944). Washington, D.C.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*, 998.

Franke, M. (2009). Signal to act: Game theory in pragmatics. doctoral dissertation, university of amsterdam., .

Franke, M., & Degen, J. (2016). Reasoning in reference games: Individual- vs. population-level probabilistic modeling. *PLOS ONE*, *11*, 1–25.

Fukumura, K., & van Gompel, P. G. (2010). Choosing anaphoric expressions: Do people take into account likelihood of reference? *Journal of Memory and Language*, (pp. 52–66).

Fukumura, K., & van Gompel, R. P. G. (2011). The effect of animacy on the choice of referring expression. *Language and Cognitive Processes*, *26*, 1472–1504.

Gahl, S., Jurafsky, D., & Roland, D. (2004). Verb subcategorization frequencies: American english corpus data, methodological studies, and cross-corpus comparisons. *Behavior Research Methods, Instruments, & Computers*, *36*, 432–443.

Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language*, *66*, 789–806.

Genzel, D., & Charniak, E. (2002). Entropy rate constancy in text. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics* (pp. 199–206). Philadelphia, Pennsylvania, USA: Association for Computational Linguistics.

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Lawrence Erlbaum Associates, Inc.

Givón, T. (1983). Topic continuity in discourse: A quantitative cross-language study. chapter Topic continuity in discourse: An introduction. (pp. 1–42). Amsterdam: John Benjamins Publishing.

Givón, T. (1988). The pragmatics of word-order: Predictability, importance and attention. *Studies in Syntactic Typology*, (pp. 243–284).

Givón, T. (1989). *Mind, code and context: Essays in pragmatics.*.

Goodman, N. D., & Frank, M. C. (2016). Pragmatic Language Interpretation as Probabilistic Inference.

Goodman, N. D., & Stuhlmüller, A. (2013). Knowledge and Implicature: Modeling Language Understanding as Social Cognition. *Topics in Cognitive Science*, *5*, 173–184.

Grice, H. P. (1975). Logic and conversation. In P. Cole, & J. L. Morgan (Eds.), *Syntax and Semantics: Vol. 3: Speech Acts* (pp. 41–58). New York: Academic Press.

Grober, E. H., Beardsley, W., & Caramazza, A. (1978). Parallel function strategy in pronoun assignment. *Cognition*, *6*, 117–133.

Grodner, D., & Sedivy, J. C. (2011). The processing and acquisition of reference. chapter The effects of speaker-specific information on pragmatic inferences. (pp. 239–272). MIT Press: Cambridge, MA.

Grodner, D. J., Klein, N. M., Carbary, K. M., & Tanenhaus, M. K. (2010). "some," and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition*, *116*, 42–55.

Grosz, B. J., Joshi, A. K., & Weinstein, S. (1995). Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, *21*, 203–225.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, *69*, 274–307.

Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. *Association for Computational Linguistics*, (pp. 1–8).

Halliday, M. A. K. (1967). *Intonation and grammar in British English*. De Gruyter Mouton.

Horn, L. (1984). Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In D. Schiffrin (Ed.), *Meaning, form and use in context* (pp. 11–42). Washington: Georgetown University Press.

Huang, Y. T., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: Insight into the semantics-pragmatics interface. *Cognitive Psychology*, *58*, 376–415.

Jaeger, T. F. (2010). Redundancy and Reduction: Speakers Manage Syntactic Information Density. *Cognitive Psychology*, *61*, 23–62.

Jaeger, T. F. (2011). Corpus-based research on language production: Information density and reducible subject relatives. In E. M. Bender, & J. E. Arnold (Eds.), *Language From a Cognitive Perspective: Grammar, Usage, and Processing* (pp. 161–197). CSLI Publishers.

Jaeger, T. F., & Buz, E. (2017). The handbook of psycholinguistics. In E. M. Fernández, & H. S. Cairns (Eds.), *Handbook of Psycholinguistics* chapter Signal reduction and linguistic encoding. (pp. 38–81). Wiley-Blackwell. To appear.

Jäger, G. (2012). Semantics: An international handbook of natural language meaning. In C. Maienborn, K. von Heusinger, & P. Portner (Eds.), *Volume 3* chapter Game theory in semantics and pragmatics. (pp. 2487–2516). De Gruyter Mouton. URL: `https://doi.org/10.1515/9783110253382.2487`. doi:doi: doi: 10.1515/9783110253382.2487.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee, & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 229–254). Amsterdam: John Benjamins.

Kao, J., Bergen, L., & Goodman, N. (2014a). Formalizing the pragmatics of metaphor understanding. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *36*.

Kao, J. T., & Goodman, N. D. (2015). Let's talk (ironically) about the weather: Modeling verbal irony. In *Proceedings of the 36th Conference of the Cognitive Science Society* (pp. 1051–1056).

Kao, J. T., Wu, J. Y., Bergen, L., & Goodman, N. D. (2014b). Nonliteral understanding of number words. *Proceedings of the National Academy of Sciences*, *111*, 12002–12007.

Karttunen, L. (1974). Presupposition and linguistic context. *Theoretical Linguistics*, *1*, 181–194.

Kasher, A. (1976). Language in focus: Foundations, methods and systems. essays in memory of yehoshua bar-hillel. chapter Conversational Maxims and Rationality. (pp. 197–216).

Kehler, A., Kertz, L., Rohde, H., & Elman, J. L. (2008). Coherence and coreference revisited. *Journal of Semantics*, *25*, 1–44.

Kravtchenko, E. (2014). Predictability and syntactic production: Evidence from subject omission in Russian. In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society* (pp. 785–790).

Kravtchenko, E., & Demberg, V. (2015). Semantically underinformative utterances trigger pragmatic inferences. In *Proceedings of the 37th Annual Meeting of the Cognitive Science Society (CogSci2015)* (pp. 1207–1212).

Kuno, S. (1972). Functional sentence perspective. *Linguistic Inquiry*, *3*, 269–320.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*, 32–59.

Kurumada, C., Brown, M., & Tanenhaus, M. K. (2012). Pragmatic interpretation of contrastive prosody: It looks like speech adaptation. In *Proceedings of the 34th Annual Meeting of the Cognitive Science Society (CogSci2012)* (pp. 647–652).

Kurumada, C., & Jaeger, T. F. (2015). Communicative efficiency in language production : Optional case-marking in Japanese. *Journal of Memory and Language*, *83*, 152–178.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*, 1–26.

Lassiter, D., & Goodman, N. D. (2013). Context, scale structure, and statistics in the interpretation of positive form adjectives. *Proceedings of SALT*, *23*, 587–610.

Lassiter, D., & Goodman, N. D. (2017). Adjectival vagueness in a bayesian model of interpretation. *Synthese*, *194*, 3801–3836.

Levin, B. (1993). *English verb classes and alternations : a preliminary investigation*. University of Chicago Press.

Levinson, S. C. (2000). *Presumptive meanings - the theory of generalized conversational implicature*. The MIT Press.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*, 1126–1177.

Levy, R., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In B. Schölkopf, J. Platt, & T. Hoffman (Eds.), *Advances in Neural Information Processing Systems 19*. Cambridge, MA: MIT Press.

Levy, R. P. (2018). Communicative efficiency, uniform information density, and the rational speech act theory. In *Proceedings of the 40th Annual Meeting of the Cognitive Science Society* (pp. 684–689).

Lockridge, C. B., & Brennan, S. E. (2002). Addressees' needs influence speakers' early syntactic choices. *Psychonomic Bulletin & Review*, *9*, 550–557.

Mahowald, K., Fedorenko, E., Piantadosi, S. T., & Gibson, E. (2013). Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition*, *126*, 313–318.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Henry Holt and Co., Inc.

McDonald, J. L., & Macwhinney, B. (1995). The time course of anaphor resolution: Effects of implicit verb causality and gender. *Journal of memory and language*, *34*, 543–566.

Minsky, M. (1975). A framework for representing knowledge. In P. H. Winston (Ed.), *The psychology of computer vision*. New York: McGraw-Hill.

Modi, A., Titov, I., Demberg, V., Sayeed, A., & Pinkal, M. (2017). Modelling semantic expectation: Using script knowledge for referent prediction. *Transactions of the Association for Computational Linguistics*, *5*, 31–44.

Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science*, *13*, 329–336.

Norcliffe, E., & Jaeger, T. F. (2014). Predicting head-marking variability in Yucatec Maya relative clause production. *Language and Cognition*, (pp. 1–39).

Norcliffe, E. J. (2009). *Head-marking in usage and grammar: A study of variation and change in Yucatec Maya*. Doctoral Dissertation, Stanford University.

Noveck, I. A., & Posada, A. (2003). Characterizing the time course of an implicature: An evoked potentials study. *Brain and Language*, *85*, 203–210.

Orita, N., Vornov, E., Feldman, N., & Daumé III, H. (2015). Why discourse affects speakers' choice of referring expressions. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)* (pp. 1639–1649).

Parikh, P. (1991). Communication and strategic inference. *Linguistics and Philosophy*, *14*, 473–514.

Piantadosi, S. T., Tily, H. J., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, *108*, 3526.

Pogue, A., Kurumada, C., & Tanenhaus, M. K. (2016). Talker-specific generalization of pragmatic inferences based on under- and over-informative prenominal adjective use. *Frontiers in Psychology*, *6*.

Potts, C., Lassiter, D., Levy, R., & Frank, M. C. (2015). Embedded Implicatures as Pragmatic Inferences under Compositional Lexical Uncertainty. *Journal of Semantics*, *33*, 755–802.

Prince, E. (1981). On the reference of indefinite-*this* NPs. In A. K. Joshi, B. L. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 231–250). Cambridge: Cambridge University Press.

Prince, E. (1992). The ZPG letter: Subjects, definiteness, and information-status. In S. Thompson, & W. Mann (Eds.), *Discourse Description: Diverse Analyses of a Fundraising Text* (pp. 295–325). John Benjamins.

Qing, C., & Franke, M. (2014). Meaning and use of gradable adjectives: Formal modeling meets empirical data. In *Proceedings of the 36th annual meeting of the Cognitive Science Society (CogSci-2014)* (pp. 1204–1209).

Qing, C., & Franke, M. (2015). Variations on a bayesian theme: Comparing bayesian models of referential reasoning. In H. Zeevat, & H.-C. Schmitz (Eds.), *Bayesian Natural Language Semantics and Pragmatics* (pp. 201–220). Springer International Publishing volume 2 of *Language, Cognition, and Mind*.

Regneri, M., Koller, A., & Pinkal, M. (2010). Learning script knowledge with web experiments. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics* (pp. 979–988).

Resnik, P. (1996). Selectional constraints: An information-theoretic model and its computational realization. *Cognition*, *61*, 127–159.

Rett, J. (2011). Exclamatives, degrees and speech acts. *Linguistics and Philosophy*, *34*, 411–442.

Rohde, H. (2008). *Coherence-Driven Effects in Sentence and Discourse Processing*. Doctoral Dissertation, University of California, San Diego.

Rohde, H., & Kehler, A. (2014). Grammatical and information-structural influences on pronoun production. *Language, Cognition, and Neuroscience*, *29*, 912–927.

Rosa, E. C., & Arnold, J. E. (2017). Predictability affects production: Thematic roles can affect reference form selection. *Journal of Memory and Language*, *94*, 43–60.

Rubio-Fernández, P. (2016). How redundant are redundant color adjectives? An efficiency-based analysis of color overspecification. *Frontiers in Psychology*, *7*, 153.

Sanford, A. J. S., Sanford, A. J., Molle, J., & Emmott, C. (2006). Shallow processing and attention capture in written and spoken discourse. *Discourse Processes*, *42*, 109–130.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, Plans, Goals and Understanding*. Hillsdale, NJ: Lawrence Erlbaum.

Schöller, A., & Franke, M. (2017). Semantic values as latent parameters: Testing a fixed threshold hypothesis for cardinal readings of few & many. *Linguistics Vanguard*, *3*.

Scontras, G., & Goodman, N. D. (2017). Resolving uncertainty in plural predication. *Cognition*, *168*, 294–311.

Sedivy, J. C. (2003). Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psycholinguistic Research*, *32*, 3–23.

Sedivy, J. C. (2007). Implicature during real time conversation: A view from language processing research. *Philosophy Compass*, *2*, 475–496.

Seyfarth, S. (2014). Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition*, *133*, 140–155.

Sgall, P., Hajičová, E., & Panevová, J. (1986). *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. Dordrecht: Reidel.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 623–656.

Sheldon, A. (1974). The role of parallel function in the acquisition of relative clauses in english. *Journal of Verbal Learning and Verbal Behavior*, *13*, 272–281.

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*, 302–319.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and Cognition*. Harvard University Press.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition (2nd edition with postface)*. Oxford: Blackwell.

Springston, J. A. (1975). *Some cognitive aspects of presupposed coreferential anaphora. PhD dissertation, Stanford University.*.

Stalnaker, R. (1973). Presuppositions. *Journal of Philosophical Logic*, *2*, 447–457.

Stalnaker, R. (1974). Pragmatic presuppositions. In R. Stalnaker (Ed.), *Context and Content* (pp. 47–62). Oxford University Press.

Stevens, S. S. (1971). Issues in psychophysical measurement. *Psychological Review*, *78*, 426–450.

Stevenson, R., Crawley, R., & Kleinman, D. (1994). Thematic roles, focus and the representation of events. *Language and Cognitive Processes*, *9(4)*, 519–548.

Stiller, A., Goodman, N., & Frank, M. (2011). Ad-hoc scalar implicature in adults and children. *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society*, .

Tily, H. J., & Piantadosi, S. T. (2009). Refer efficiently: Use less informative expressions for more predictable meanings. In *Proceedings of the workshop on the production of referring expressions: Bridging the gap between computational and empirical approaches to reference.*.

Vogel, A., Potts, C., & Jurafsky, D. (2013). Implicatures and nested beliefs in approximate decentralized-POMDPs. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (pp. 74–80).

Walker, M. A. (1993). *Informational redundancy and resource bounds in dialogue*. Doctoral Dissertation, University of Pennsylvania, Philadelphia, PA.

Ward, G., & Hirschberg, J. (1985). Implicating uncertainty: The pragmatics of fall-rise intonation. *Language*, *61*, 747–776.

Wasow, T., Jaeger, T. F., & Orr, D. M. (2011). Lexical variation in relativizer frequency. In H. Simon, & H. Wiese (Eds.), *Workshop on Expecting the unexpected: Exceptions in Grammar at the 27th Annual Meeting of the German Linguistic Association (DGfS)* (pp. 205–211). De Gruyter Mouton.

Weischedel, R., Pradhan, S., Ramshaw, L., Palmer, M., Xue, N., Marcus, M., Taylor, A., Greenberg, C., Hovy, E., Belvin, R., & Houston, A. (2008). *Ontonotes release 2.0*. Philadelphia: Linguistic Data Consortium. Philadelphia: Linguistic Data ConsortiumPhiladelphia: Linguistic Data Consortium.

Wilson, D., & Sperber, D. (2004). Relevance Theory. In L. R. Horn, & G. Ward (Eds.), *The Handbook of Pragmatics* (pp. 606–632). Oxford, UK: Blackwell Publishing volume 1.

Yamamoto, M. (1999). *Animacy and Reference: A Cognitive Approach to Corpus Linguistics*. John Benjamins Publishing.

Zipf, G. K. (1949). *Human Behaviour and the Principle of Least Effort*. Cambridge, MA: Addison-Wesley.

Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 386–397.