Informationally Redundant Utterances Alter Prior Beliefs about Event Typicality

Ekaterina Kravtchenko
Saarland University
Vera Demberg
Saarland University

Abstract Most theories of pragmatics and language processing predict that speakers avoid excessive informational redundancy in order to conserve unnecessary cognitive or articulatory effort. Informationally redundant utterances are, however, quite common in natural dialog. From a comprehension standpoint, it remains unclear how listeners interpret these utterances, and whether they make attempts to reconcile the ‘dips’ in informational utility with expectations of ‘appropriate’ or ‘rational’ speaker informativity. We show that informationally redundant (overinformative) utterances can trigger compensatory pragmatic inferences, which increase utterance utility in line with listener expectations. In a series of 3 studies, we look at utterances which refer to stereotyped event sequences describing common activities (scripts). When comprehenders encounter utterances describing events that can be easily inferred from prior context, they interpret them as signifying that the event conveys new, unstated information (i.e. an event otherwise assumed to be typical, such as paying the cashier when shopping, is reinterpreted as atypical in context). Further, we show that the degree to which such inferences are triggered depends on the framing of the utterance. In the absence of prosodic or discourse markers indicating the speaker’s specific intent to communicate the given information, or an utterance’s relevance to the discourse, such inferences are far less likely to arise. Overall, the results demonstrate that excessive conceptual redundancy leads to listeners revising the conversational common ground, in an effort to accommodate unexpected dips in informational utility.

Keywords: Psycholinguistics; redundancy; context-dependent implicatures; accommodation; overinformativeness; prosody.

Word count: 17,622

1 Introduction

Theories of pragmatics or language processing typically include constraints against utterances that add no new information to the discourse, or are informationally redundant (Aylett & Turk 2004, Cohen 1978, Grice 1975, Jaeger 2010). Redundancy
in linguistic form may include overt mention of, or increased articulatory effort towards producing material, at any linguistic level, that is easily predictable or recoverable in context. Examples include expending excessive articulatory effort on a predictable word (e.g. via vowel lengthening, Aylett & Turk 2004), using a longer (meaning-equivalent) linguistic variant in place of a shorter one (Mahowald et al. 2013), or including an optional complementizer when it does not help to predict upcoming material (Jaeger 2010).

The majority of studies investigate predictability at the form level, focusing on the probability that a specific (string of) linguistic units will follow a preceding string. At a conceptual level, predictability refers to the probability of a certain (discourse-relevant) world state, given preceding linguistic material and existing world knowledge. Utterances that are informationally redundant may overtly state information that’s already in the common ground, echo information previously stated, or provide information either entailed or very strongly implied by the preceding discourse. For example, given the context of My best friend is Mary, the world state of Mary is a girl is highly likely, given conventional understanding of gender-name correspondence. Further, overtly mentioning that Mary is a girl, in this context, may be perceived as unnecessarily redundant at a conceptual level, but likely unpredictable at the form level. Whether and how comprehenders react to and resolve conceptual redundancy is to date unresolved.

Most mainstream pragmatic theories are based on the idea that listeners generally expect speakers to behave rationally (for example, not to expend unnecessary effort on what does not need to be said), and predict that listeners should be sensitive to abnormal or unexpected dips in informational utility (Levinson 2000). Moreover, as Grice pointed out (while doubting that there was a general constraint against overinformativity), listeners may be driven to rationalize unexpected dips in utility by infusing them with additional pragmatic meaning; i.e., making compensatory inferences about the world state or the speaker’s intended meaning, that would render the utterance less redundant and more relevant to communicative goals. On this basis, we argue that speakers should avoid overtly stating conceptually predictable information, except when they intend it to communicate a more informative, non-literal meaning. Correspondingly, rational listeners ought to interpret informationally (conceptually) redundant utterances as either conveying additional (non-literal) meaning, that is more informative or relevant to communicative goals than would appear at face value, or as reflecting a background world state where the information conveyed cannot be taken for granted (i.e., is not redundant from the speaker or listener’s perspective). This latter question about comprehender interpretation is what we set out to test in the studies reported here.

While formal pragmatic theories do address cases where the information contained in an utterance (rather than the string itself) may be redundant, they largely
Informationally Redundant Utterances Alter Prior Beliefs

Informationally redundant utterances alter prior beliefs, leaving open or not exploring in any detail, the question of how speakers and listeners treat such utterances. For example, while informational redundancy may be suboptimal from the point of view of the speaker, who expends effort on uttering something that can be automatically inferred by the listener, it is not clear whether informationally redundant utterances pose any problem for the listener (Grice 1975). They do not clearly interfere with comprehension, as under informativeness or underspecification does, and may aid comprehension in some cases (Nadig & Sedivy 2002; e.g., object identification). The question of whether listeners attempt to accommodate redundancy, by making compensatory inferences regarding the speaker’s intended meaning, or the current situation model, is currently unclear. Grice (1975) himself questioned whether there is in fact a pragmatic constraint against informational redundancy, and evidence from use of overinformative (contextually redundant) nominal modification has been equivocal.

Most experimental work on speaker production, and listener comprehension of informationally redundant material has focused on nominal modification in referent identification tasks (Engelhardt et al. 2006, Nadig & Sedivy 2002, Davies & Katsos 2010, 2013). This work has showed that conceptual redundancy may be helpful to the listener, though there is evidence that use of some redundant adjectives is penalized by listeners; however, listeners do not necessarily accommodate the violation, by making (e.g., contrastive) inferences which increase the utility of redundant descriptions. It is, however, unlikely that findings based on referent identification tasks can be easily generalized to make predictions about how speakers and listeners should regard non-conventionalized forms of ‘larger-scale’ redundancy, at the level of an entire clause or sentence, or when detecting and resolving redundancy requires extensive reasoning about general world knowledge. The present studies therefore address an important gap in the literature, by studying the effect of utterances that are redundant at the clausal or sentential level.

Further, while research on conventionalized or lexicalized implicatures has been invaluable to linguistic theory, due to their role in disambiguating pragmatic and semantic contributions to utterance meaning, context-dependent implicatures and inferences (being far more numerous and frequent) are arguably more important to the study of more general human communication. To date there has been virtually no experimental work teasing apart which properties of utterances trigger extra pragmatic inferences, what sufficient or necessary criteria for triggering such inferences are, and which contextual cues speakers and listeners exploit in getting their intended meaning across and interpreting intended meaning, respectively. We therefore argue that experimental work on this topic is important both to processing-based understanding of language and to formal computational models of meaning/processing. Here we present a first attempt at experimentally testing 1) whether a specific inference, tentatively predicted by rational speaker models such as (Grice 1975), is drawn
by comprehenders; and 2) whether and to what degree altering various contextual factors and cues affects the comprehenders’ willingness to draw this inference.

To summarize, Grice (1975) suggested that the constraint against overinformativeness (in the sense of excessive time and effort spent on communicating a piece of information) may barely be a constraint at all, with the only consequence of violations being a ‘waste of time.’ He did, however, note that listeners may be confused by overinformativeness, and presume that there must be a ‘point’ to the ‘excess’ information. How listeners do, in fact, react to overinformativeness (and, by extension, the extent to which they monitor and expect speaker rationality/conciseness), has not to date been investigated in the theoretical or experimental literature, outside the relatively narrow scope of nominal modification. Ultimately, these questions are relevant to a more general theory of human communication, and must be answered to determine the extent to which a) listeners and/or speakers are Gricean; b) under which conditions they behave in a Gricean manner (some violations of rational communicative behavior, which do not directly affect understanding – such as informational redundancy – may be better tolerated by both); c) how and whether listeners resolve apparent violations, and the assumptions about implied meaning or common ground that they are willing to make in order to do so.

In order to investigate this question experimentally, we created a series of short narratives, which contain a sentence describing an event that is inferable from common sense knowledge about event sequences (these are also referred to as scripts Schank & Abelson (1977); see also section 2.2). Scripts describe events that are part of typical or stereotyped activity sequences such as eating when going to a restaurant, or buying a ticket when using the subway. For instance, participants may read passages stating overtly that somebody paid the cashier when they went shopping. Paying the cashier is highly expected in the shopping scenario and is hence highly predictable at the event level, but conceptually redundant at the utterance level.

We present three experiments which test whether such informationally redundant event descriptions trigger compensatory inferences, by leading comprehenders to alter their initial beliefs about how predictable the event described is (on the premise that less predictable events are more likely to be mentioned). We predict that informationally redundant descriptions should generate so-called ‘atypicality’ implicatures, where comprehenders resolve the apparent dip in informational utility by assuming that the apparently inferable event is in fact relatively atypical, therefore justifying its mention by the speaker. For example, in our scenario involving shopping and paying the cashier, a possible ‘compensatory’ inference would be that the conversation concerns someone who does not pay predictably (e.g., a habitual shoplifter). We first look at the comprehension of such utterances in discourse contexts which implicitly support the inference in question, through use
of semantically vacuous prosody or discourse markers\textsuperscript{1}. The first experiment uses implicit exclamatory intonation (!) to mark the utterance message as an intentionally conveyed, important, relevant piece of information. The second experiment uses the discourse marker Oh yeah... to do the same, while avoiding the implied surprise suggested by an exclamation mark. Last, we test whether informational redundancy itself is sufficient to trigger ‘atypicality’ inferences, in the absence of prosodic or discourse cues which support the inference, in order to obtain a better understanding of the circumstances under which ‘atypicality’ implicatures are triggered.

2 Theoretical Framing and Related Work

Although informationally redundant utterances are typically regarded as infelicitous in the linguistics literature, they have been observed to be surprisingly common in natural dialog (Walker 1993). Baker et al. (2008) observed that such utterances are frequently used in response to signs (or the possibility of) listener non-comprehension, when responding to listener questions, or when speaking to strangers. Walker (1993) further concludes that informationally redundant utterances are specifically used to address cognitive resource limitation (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 salient or relevant fact. This literature has addressed in some detail the contexts in which speakers might choose to be informationally redundant, and has the advantage of focusing on conceptual or ‘informational’ redundancy which requires substantial speaker, and is therefore more likely be be noticed or flagged by listeners as a meaningful deviation from conversational norms. However, it has not looked at redundancy from the listener’s perspective.

Most formal or experimental research on pragmatic inferences to date has focused on use and interpretation of scalar implicatures (Horn 1984), such as the use of some to implicate not all, or warm to implicate not hot. In other words, the focus has been on conventionalized inferences based on specific lexical items (or classes of lexical items), with minimal reliance on context or general reasoning about speaker intentions (but see Degen et al. 2015). Accordingly, most formal and experimental work on over- or underinformativeness (or, informational redundancy vs. insufficient information) has focused on the use of scalar terms and nominal modifiers (the cup, the red cup, the small red cup, the cup on the napkin, etc.; Davies & Katsos 2010, 2013, Engelhardt et al. 2006).

\textsuperscript{1} Speakers frequently use multiple cues consistent with their intended meaning, in order to make an inference maximally easy to calculate for the listener (consider, for example, the difficulty of interpreting sarcasm without supportive prosody). We therefore consider supportive prosody to be the ‘default’ case.
Inferences on the basis of scalar terms tend to be conventionalized (minimally contextually sensitive), and do not require particularly effortful reasoning about world knowledge or speaker intentions – nor do they require complex or high-level monitoring of a speaker’s (or one’s own) adherence to or apparent deviation from communicative norms. Work on generalized implicatures in itself is therefore insufficient in determining the degree to which speakers and listeners do in fact stick to Gricean norms, in communicating information or interpreting utterances, or the degree to which they normally engage in or assume rational communicative behavior. Utterances which deviate from communicative norms only in context, and on the basis of reasoning about world knowledge, however, have not been studied systematically by pragmaticists, experimentally or otherwise\(^2\). A more general theory of human communication, as Grice (1975) originally intended, must account for the drawing of inferences both on the basis of relatively fixed, general-currency expressions (Generalized Conversational Implicatures; GCIs), and on the basis of more complex reasoning about context and world knowledge (Particularized Conversational Implicatures; PCIs).

A further note we wish to make is that the ‘atypicality’ inferences readers might draw from overinformative utterances, such as those investigated in our study, are not typical examples of conversational implicatures. Take, for example, the following utterance:

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

It’s unclear whether speakers would use such an utterance to intentionally communicate that John is a habitual non-payer (unless they are flouting other conversational norms). Speakers are typically expected to use less ambiguous, and easier-to-decipher ways of communicating new or unexpected information. A more likely scenario may be that an overhearer of such an exchange between two speakers might conclude that the implied information is already mutually known to the speaker and listener. A critical component, in this case, of what makes something a conversational implicature is speaker intent (Grice 1957). In this paper, however, we are primarily concerned with comprehenders’ reconciliation of informativity expectations with apparent violations thereof, as well as their sensitivity to such violations. The theoretical status of these utterances (or what new information the speaker may be intending to communicate to their targeted listener) is however not critical for our case. Further, it’s not entirely obvious whether it’s implausible for speakers to communicate new information in such a manner, particularly when there is support

\(^2\) But see Degen et al. (2015), for evidence that scalar implicature comprehension may also involve reasoning about world knowledge; beliefs about the world may be altered to resolve ‘odd’ scalar term use.
Informationally Redundant Utterances Alter Prior Beliefs

for the intended interpretation in the form of prosody, body language, etc. Consider for instance the well-known example found in Grice (1975):

(2) A and B are talking about a mutual friend, C, who is now working in a bank. A asks B how C is getting on in his job, and B replies: “Oh quite well, I think; he likes his colleagues, and he hasn’t been to prison yet.” (Grice 1975: p. 43; emphasis added)

The case is related to our study materials, in that the listener hears a highly predictable event described explicitly, and hence may assume that the event is in fact unpredictable in the particular world described by the utterance. Grice argues in this case that the addressee may assume, upon hearing this, that “C” has an unusually high likelihood of ending up in prison – presumably given that people do not typically end up in prison under such circumstances (making it unnecessary to mention, absent reasonable suspicion that C is likely to deviate from this norm). If there was no such suspicion, then mentioning that C has not been to prison yet would be informationally redundant, as the listener would have already automatically assumed a world state in which this would be the case. The addressee would make this inference, in turn, based on an overriding assumption that the speaker is rational and cooperative, and would not intentionally produce utterances with low informational utility. Whether addresses in fact are willing to make such inferences, when there is minimal contextual support for the inference (as above), is another issue we explore in this paper.

Most experimental work on speaker production, and listener comprehension of informationally redundant material has focused on nominal modification (the [red] apple, the [tall] boot). The evidence has been somewhat mixed: in interactive, spontaneous speech, speakers frequently modify nouns with adjectives that are not necessary for referent identification (e.g. referring to a cup as the red cup, in a context where only a single cup is present) (Engelhardt et al. 2006, Nadig & Sedivy 2002: 30% and 50% of nominal descriptions were overspecified in spontaneous speech, respectively). Further, listeners do not appear to penalize such utterances (when restricted to types likely to be produced spontaneously): Engelhardt et al. (2006) showed that comprehenders judge overinformative descriptions as signif-

3 Note that in the redundant utterance in example (2), new or unexpected information is communicated in a relatively obscure manner, requiring complex reasoning about typical world states. It is by no means a given that typical addressees would be able to compute the intended meaning. See also, for example, the famous reference letter example (Grice 1975: p. 52): A is writing a testimonial about a pupil who is a candidate for a philosophy job, and his letter reads as follows: ‘Dear Sir, Mr. X’s command of English is excellent, and his attendance at tutorials has been regular. Yours, etc.’, which likewise requires multiple quite specific types of background world knowledge in order to interpret correctly.
icantly more acceptable than underinformative descriptions, and that they do not draw any additional inferences from overinformative descriptions. They conclude from these findings that listeners, at least when comprehension is not impaired by the violation of conversational norms, are “only moderately Gricean.” Davies & Katsos (2010) similarly find that overinformative expressions are more likely to be both produced and tolerated than underinformative expressions, by both children and adults, although both are recognized as suboptimal. Pogue et al. (2016) further find that listeners are less likely to make generalizations about a speaker’s respective tendencies towards over- or underspecification from their use of overinformative descriptions, in stark contrast to underinformative descriptions, suggesting that listeners are less sensitive to violations of (idealized) conversational norms that do not interfere with comprehension, or recognize them as commonplace and inconsequential.

These studies, however, address only what happens when speakers expend minimal time and speech effort, beyond what might be considered ‘optimal,’ and often in a context where the additional information may be perceptually useful to listeners (and therefore communicatively useful). Settling the question over how problematic overinformativeness is (for either speakers or listeners) based solely on adjective use is hence likely to be problematic – including an extra (non-contrastive) adjective in a description does not substantially increase speaker effort, nor is it obviously superfluous from a listener’s perspective. Additional perceptual information about an object, even when not strictly necessary, may speed up or otherwise aid identification (Davies & Katsos 2013). We therefore argue that in order to test whether overinformativeness is perceived as abnormal, or whether listeners attempt to resolve informational redundancy by assuming more informative non-literal utterance meanings, it would be useful to test materials which, as Example (2) above, are conceptually redundant on a level that requires significant effort for speakers to produce, and that are more likely to catch listeners’ attention.

2.1 Notes on Terminology

In experimental literature, informational redundancy has typically 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 not provide more information than is necessary to get their message across. However, overinformativeness has at times puzzlingly referred to both informational, or conceptual redundancy (Engelhardt et al. 2006, Grice 1975), as well as overinformativeness related to the informativeness of terms in an implicational

---

4 Davies & Katsos (2010) do show higher rejection rates for overinformative descriptions than (Engelhardt et al. 2006), but they also purposefully used adjective types relatively unlikely to be produced spontaneously.
Informationally Redundant Utterances Alter Prior Beliefs

scale (e.g. the use of some when all is sufficient) (Levinson 2000). The latter variety of overinformativeness, associated with the Quantity maxim, is more a problem of vagueness where a more precise description is available; informational redundancy in contrast is a problem of excessive wordiness or precision. A closer analogy (likewise typically associated with the Quantity maxim) is with overinformative nominal modification (such as using the big red cup or the cup on the towel to identify the only cup in a given context), where speakers might choose to describe objects in more detail than strictly necessary (Engelhardt et al. 2006). However, overinformative nominal modification (in an object-identification task) is often perceptually helpful on the one hand (Arts et al. 2010, Mangold & Pobel 1988), and involves minimal additional effort on the speaker’s part on the other, leading to questions over how ‘redundant’ these expressions in fact are. It is similar to cases of larger-scale informational redundancy, in that participants tend to evoke a contrastive interpretation when the additional information is not perceptually useful (Davies & Katsos 2013), suggesting that listeners do attempt to make any ‘extra’ information communicatively useful.

2.2 Script Knowledge as Proxy for World Knowledge

Script knowledge refers to people’s implicit awareness of the typical event structure of various stereotyped activities, such as going shopping or going to a restaurant. The former, for example, normally involves events such as going to a grocery store, selecting food items, and paying the cashier; while the latter typically involves going to a restaurant, selecting something from the menu, and eating food. As script knowledge is generally a systematic, implicitly shared type of world knowledge (few would dispute that people typically eat when they go to restaurants, for example), it makes for an ideal starting point for exploring how and whether comprehenders accommodate utterances which state the generally obvious. Literature on processing of event sequences shows that people anticipate upcoming events or future states once a script is ‘invoked,’ and have faster reading times when information is consistent with previous script knowledge (Zwaan et al. 1995). Likewise, when recalling stories based on a variety of scripts, people have significant difficulty distinguishing between actions that were implied by the script, but not mentioned overtly, and those that were actually mentioned in the story (Bower et al. 1979). These findings suggest that activities or sub-events which are strongly associated with a script are part of its conventional meaning (unless explicitly indicated otherwise), whether they are overtly mentioned or not – and that mentioning them overtly is generally redundant.

Certain types of information may be more or less central to a script - for example, going grocery shopping almost always entails paying for your groceries, but does not typically entail buying apples (although the latter is entirely consistent with the
script). The informationally redundant utterances that we concern ourselves with refer to central events which are easily inferred in the context of the event sequence – for example, *entering a restaurant*, in the context of a restaurant visit:

(3) a. Yesterday, Marie went out to eat. She **entered the restaurant**. She ordered pasta, and read a book.
   b. Yesterday, Marie went out to eat. She ordered pasta, and read a book.

In both variants, comprehenders should automatically infer, after the *going out to eat* script is invoked, that *Marie* would be entering a restaurant, although it is not mentioned explicitly in (b). The first variant, therefore, may be perceived as somewhat anomalous. 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 typically not mentioned overtly (Bower et al. 1979, Regneri et al. 2010). A question that emerges is whether such ‘redundancies’ are perceived as simply odd, or whether listeners resolve the apparent oddness, in order to maintain the assumption that speakers are rational and efficient (Grice 1975), by assigning to the apparently redundant information some kind of narrative function or other informational utility.

3 How Comprehenders Resolve (Apparent) Informational Redundancy

Dialogue literature on informational redundancy (Baker et al. 2008, Walker 1993) is primarily concerned with determining the contexts in which speakers tend to make use of it. Informationally redundant utterances are frequently used for clarification (when presenting new or important information), to make explicit inferences that are relatively difficult to make, to draw attention to a salient point, or to address apparent (or predicted) listener non-comprehension. Apart from indirect evidence about speaker intent, based on annotator or author judgements, there is however no existing literature on how comprehenders resolve apparent informational redundancy when it is encountered, or whether (and when) they perceive it as flouting or a violation of conversational norms.

We distinguish between three theoretical possibilities for how comprehenders might react to informational redundancy. Firstly, if listeners do monitor speakers’ output for an expected level of informational utility, and have an expectation that

---

5 Although Brown & Dell (1987) and Dell & Brown (1991) do not explore what inferences (if any) comprehenders make on the basis of explicit mention of typical (inferable) instruments, they do find that instruments more predictable in the context of a scripted event are less likely to be mentioned by speakers in an storytelling task.
speakers will be ‘concisely informative,’ then they should attempt to resolve any apparent violations, by drawing context-sensitive and utility-maximizing pragmatic inferences from utterances that are (at face value) informationally redundant. Doing so allows them to preserve the belief that the speaker is being cooperative, as assigning an ‘informative’ meaning to an apparently informationally redundant utterance in effect removes the informational redundancy.

Alternatively, listeners may not pay much attention to informational redundancy, provided it does not interfere with comprehension – as has been found in some studies of apparently overinformative nominal descriptions – or at most, find them odd or suboptimal. It is likely that listeners do not expect speakers to be fully rational in all cases and at all times, and that less effort might be expended on monitoring speech for minor deviations from rationality, that have no effect on the listener’s ability to understand the scenario as described. Similarly, listeners may note speech abnormalities, but not draw any inferences regarding intended meaning as a result, making sense of it by ascribing the abnormality to some sort of speaker error (for example, that the speaker is stalling for something else to say, is having production difficulty, or is simply being temporarily non-optimal).

There is a third possibility, suggested by the observation that implicatures must be calculable, or sufficiently supported in context that a listener would not encounter undue difficulty in inferring the intended meaning (Levinson 2000). Ability or willingness to compensate for apparent violations of conversational norms, with respect to informational utility, may hence be a function of the amount of effort listeners must expend on doing so (Sperber & Wilson 1995), as well as the amount of evidence that listeners have of the speaker having intended to say something ‘appropriately’ informative in the first place. It may also be affected by how large of a violation the listener is otherwise left to ‘explain away’ – with smaller violations, such as slightly overspecified nominal descriptions, perhaps routinely expected in the course of conversation.

Let’s now take a moment to go through the types of stimuli we use, to discuss what behaviour we would expect in the first case. Our target utterances introduce a well-known script or event sequence, and our redundant utterance (marked in bold below) then consists of mentioning a highly predictable event from such an event sequence, as in Example (4):

(4) S1: “John went grocery shopping. **He paid the cashier!**”

In this example, the speaker is apparently violating conversational norms, by stating something which can be normally inferred from preceding information. The event described by the second sentence is already conventionally implied to have occurred, by the invocation of the grocery shopping script in the first sentence. The underly-
ing assumption is that people overwhelmingly pay cashiers when they go grocery shopping. Mentioning it explicitly, therefore, is informationally redundant, and if the utterance is to conform with listener expectations about speaker informativity, then the listener is likely to attempt to infer what other information the speaker may have been attempting to convey.

A natural inference is therefore that the listener could not have otherwise assumed, given what had already been stated, that the behavior described (cashier-paying) had taken place; i.e. that this behavior is somehow atypical for John. A straightforward interpretation of such an utterance would then be that John is not typically in the habit of paying during his grocery trips. In other words, listeners should be inclined to revise the common ground in order to make it compatible with a more ‘informative’ interpretation of the utterance. Assuming that listeners detect the informational redundancy, the only context in which one would expect for the inference to not be drawn, is if it is already mutually known by the speaker and listener that this is an atypical event for John. For example, both might share the knowledge that John is a habitual shoplifter. In this case, overt mention of the cashier-paying event is informative at face value, and would not conflict with assumed conversational norms.

4 Experiment 1: Implicit Prosodic Emphasis

We first test whether informationally redundant event descriptions trigger compensatory atypicality inferences, when there is minor contextual support for the implicature, in the form of an exclamation mark at the end of the utterance. Exclamatory intonation, intuitively, is a natural way of framing a description of something ‘atypical.’ When implicatures are context-dependent (and even if they are not; see Degen & Tanenhaus 2015), speakers generally provide multiple signals of their intended meaning, in order to make the implicature easier to calculate for the listener. One would expect this to particularly be the case when the intended meaning of the utterance substantially violates previously held beliefs, as opposed to simply providing new but marginally expected (or at least unsurprising) information. Unexpected or surprising information is frequently accompanied by exclamatory intonation, which in text can be simply represented with an exclamation mark (Rett 2011), without otherwise altering the semantic content of the utterance.

To test this, we 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 normal behavior (where relevant). Some of the narratives include brief dialogue between two discourse participants at the end (which may include informationally redundant or informative event descriptions). After reading the narratives, participants rate how typical they
believe certain behaviors in the story to be. We expect participants who read informationally redundant event descriptions to infer that the utterance in fact signals that the event in question is relatively unexpected, or atypical (as only relatively unexpected events warrant explicit mention). In contrast, those participants who read concisely informative event descriptions should draw no such inferences.

4.1 Methods

Participants

400 participants, located in the US, were recruited on Amazon Mechanical Turk.

Design

The primary question of interest is whether informationally redundant utterances (in this case, descriptions of highly predictable events) are perceived as violating conversational norms, and whether they consequently trigger pragmatic inferences. These inferences should revise common-ground beliefs about the (a)typicality of said events (and so ‘repair’ the violation, or dip in informational utility):

(5) “John went grocery shopping. He paid the cashier!”

Here, we hypothesize that readers will infer that John does not typically pay the cashier, as such a scenario would justify overt mention of John’s cashier-paying. The informational redundancy arises due to the high predictability of paying the cashier, and is resolved if one assumes that the event is not as predictable (in John’s case) as initially assumed.

We also wanted to check whether the inference (that an event is less typical than would normally be expected) could be cancelled by manipulating the common ground, by making the event in question an atypical-in-context activity. When the event described is atypical in context, the description ceases to be informationally redundant (i.e., it becomes informative), and the inference should therefore not arise. This control condition, by keeping the description itself constant and manipulating only the context, ensures that any effect we measure is in fact due to the presence of informational redundancy, and verifies that comprehenders are sensitive to discourse context:

(6) Context: John habitually doesn’t pay.

“John went grocery shopping. He paid the cashier!”

Finally, we wanted to provide a baseline for ‘typical’ interpretation of ‘appropriately’ or ‘concisely’ informative event descriptions. Additionally, we wanted to confirm
that similarly structured descriptions of *optional* events do not provoke similar inferences (which would suggest a problem with the stimulus design or response measure). In (6), the utterance is not informationally redundant, and is not expected to generate any specific inferences. Further, we wanted to confirm that the atypical common ground inserted into the previous example does not affect the interpretation of *optional* event mentions (which would suggest that there is an unexpected effect of context manipulation on stimulus interpretation, in general):

(7) **CONTEXT:** *Typical* or John habitually doesn’t pay.  

“John went grocery shopping. **He got some apples!**”

As in (6), participants should draw no typicality-based inferences here, as the event described is not overly predictable. These conditions therefore provide a secondary control measure.

**Measures**

To measure comprehender beliefs, each story we presented was followed by 4 questions presented in random order, regarding activities mentioned in the story (including both *predictable* and *optional* 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:

![Figure 1: A sliding scale as used in the experiments.](image)

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, or ‘accuracy’ checks, in the experiment itself. All points on the response scale were associated with a number ranging from 0 (*Never*) to 100 (*Always*).

Prior beliefs, or baseline beliefs regarding activity typicality, were estimated from responses to stimuli presented without the potentially redundant event description, or where the event description/utterance was not visible (see the next section for a
more detailed explanation). The responses, aside from setting baseline measures (prior beliefs) of activity typicality, also provide a 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 highly inferable, within a given class, can be compared against one another.

Updated beliefs regarding activity typicality were estimated from responses to stimuli which included the redundant or informative utterance (event description), or where the event description/utterance was visible.

Belief change due to reading the event description was determined by measuring the magnitude and direction of difference between prior beliefs and updated beliefs.

Materials

24 stimuli were constructed as brief stories/narratives, based on distinct stereotyped scripts or activities. Each story had one of 2 context types (typical vs. atypical common ground, relative to the conventionally predictable or inferable script event). In all stories, declarative utterances, spoken by one of the discourse participants, described one of 2 types of script events (conventionally predictable vs. optional), making a total of 4 conditions.

Conventionally predictable events (example (8)) can be inferred simply from the ‘speaker’ having invoked the script, while optional events (example (9)) can not be inferred automatically, as they may only occasionally occur as part of the script activity. To clarify, we are using the term conventionally predictable to specify that the event can typically be inferred from, or almost invariably occurs as part of the event script(under normal conditions). Initial discourse context was either typical ([1a], in (8)) with respect to the script, or atypical, in that it implied the conventionally or script-predictable event was in fact unusual for the event participant ([1b] below):

(8) CONVENTIONALLY PREDICTABLE EVENT

[1a/b] John often {goes to his local supermarket, as it’s close bya-typical | doesn’t pay at the supermarket, as he’s typically brokeb-atypical}.

[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.

The context/common ground manipulation in [1b] was used in order to render the conventionally predictable event unusual, or at least optional. Optional events could not be automatically inferred from the script having been invoked:

(9) **Optional Event**

[1a/b] John often \{goes to his local supermarket, as it’s close by\textsubscript{a-typical} doesn’t pay at the supermarket, as he’s typically broke\textsubscript{b-atypical}\}.

[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.

[5] I just saw him in the living room.”

Participants saw either only the common ground context [1] and discourse setup [2] (without numbering or special formatting), which enabled us to collect estimates of how typical activities are believed to be, based on the context alone (prior beliefs); or the entire text, which enabled us to collect estimates of how typical activities are believed to be, based on both the context and the event description [4] (updated beliefs).

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

i. How often do you think John usually pays the cashier, when at the grocery store?

ii. How often do you think John usually gets apples, when at the grocery store?

iii. How often do you think John usually goes to the grocery store?

iv. How often do you think John usually talks to Susan about himself?

Each question could be responded to on a continuous sliding scale of ‘Never’ to ‘Always’ (see Figure 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 went grocery shopping. I paid the cashier! I just got back from the store.”.

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; there needed to be predictable as well as optional subactivities (paying the cashier, getting apples); and the conventionally predictable activity had to be such that the global activity sequence could still conceivably occur without it. 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 all participants could be expected to agree on what a script entailed, which activities were or weren’t obligatory to the script sequence, etc.. Many of the scenarios, as well as estimates of the ‘centrality’ or predictability of component activities in the script sequence, were taken from unpublished pre-test data provided by the authors of Raisig et al. (2009), who looked at retrieval of script knowledge from memory. Additional scenarios were created where participant norming suggested that some scripts did not meet criteria, were insufficiently familiar to most people, or excessively culturally variable.

**Procedure**

Participants were asked to read 6 stories randomly selected out of the total of 24. Each condition was therefore only presented once, as follows (given the very small number of items, there were no fillers). 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 allowed us to collect measures of prior beliefs regarding activity typicality, and the 4 full stories gave us measures of updated beliefs conditioned on the event description. Comparisons between prior and updated beliefs could therefore only be made between subjects:
John often goes to his local supermarket, as it’s close by.

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.

She commented: “John went shopping. He paid the cashier! I just saw him in the living room.”

# How often do you think John usually pays the cashier, when at the grocery store?

The experiment thus employed a between-subject design, where prior and updated belief estimates for any given item were provided by different participants, to eliminate the possibility of participants conditioning their updated estimates not only on inferences made from the text, but also on their own prior estimates.

### 4.2 Results

For the purposes of determining whether participants made any inferences regarding activity typicality, we modeled belief change, i.e. the difference between prior and updated beliefs, or activity typicality estimates made with and without seeing the redundant or informative event description. Conventionally predictable and optional activities were modeled separately, as the optional 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 Predictable Activities

As expected, prior belief ratings (obtained from participants who did not see the event descriptions) showed that typical context events are perceived as highly typical (84.43 on a 0-100 scale), and atypical context events as relatively atypical (37.79).
Further, as predicted, updated belief ratings (obtained from participants who saw the conventionally predictable, or redundant event descriptions) show lower typicality for the typical context events (71.98) than prior belief ratings; see Figure 2. There was little change in participants’ ratings of the atypical context event (i.e. in the condition where the conventionally predictable event was made unpredictable by the common ground context), from prior beliefs (37.79) to updated beliefs (38.62).

A linear mixed effects regression analysis, fitted using \( \chi^2 \)-test model comparison (backward selection), showed an interaction between context and utterance/event description visibility, which is driven by lowered activity typicality ratings when the readers see the utterance in a typical context (\( \beta = -14.61, p < .001 \)). We used the maximal model, with by-subject random intercepts and slopes for common ground context (typical/atypical) and belief measure (prior/updated), as well as by-item random intercepts and slopes for both factors and their interaction (Barr et al. 2013). A plot illustrating the interaction can be seen in Figure 2.

<table>
<thead>
<tr>
<th></th>
<th>Coef ( \beta )</th>
<th>SE(( \beta ))</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>58.94</td>
<td>2.18</td>
<td>27.04</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Context: Typical</td>
<td>38.40</td>
<td>3.54</td>
<td>10.85</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Utterance: Visible</td>
<td>−3.85</td>
<td>1.92</td>
<td>−2.01</td>
<td>0.06</td>
</tr>
<tr>
<td>Context * Utterance</td>
<td>−14.61</td>
<td>2.96</td>
<td>−4.94</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Table 2  
Experiment 1: conventionally predictable activities 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.

These results show that, as predicted, when a conventionally predictable activity is seen in a typical world context (i.e. a context in which the activity can be automatically inferred), readers infer that the conventionally predictable activity must in fact be optional, i.e., atypical for the individual who is the subject of the story, and therefore worth mentioning explicitly.

Optional Activities

There was little change in participants’ ratings of optional events from prior beliefs (typical: 38.43; atypical: 34.87) to updated beliefs (typical: 38.46; atypical: 38.82), see Figure 3.

A linear mixed effects regression analysis, fitted using \( \chi^2 \)-test model comparison, showed that estimates of activity typicality do not vary with variations in the common
Figure 2  Experiment 1: conventionally predictable activities analysis. This plot shows changes in activity typicality estimates depending on whether the utterance is present, as well as whether the context causes the utterance to be perceived as atypical. Violin plots, overlaid with box plots, show the distribution of estimates. Circles represent mean values. Arrows show significant differences between before/prior and after/updated ratings.

ground context (or world typicality), nor are conditioned on the utterance describing the optional activity (see table 3). This is also consistent with our predictions, and indicates both that the context alteration does not inherently cause a change in activity typicality estimates (regardless of how script-central the activity is), and that optional activities, given our typical context, are not interpreted as atypical when mentioned.

<table>
<thead>
<tr>
<th></th>
<th>Coef β</th>
<th>SE(β)</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>37.44</td>
<td>2.51</td>
<td>14.93</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Context: Typical</td>
<td>1.52</td>
<td>2.46</td>
<td>0.62</td>
<td>0.54</td>
</tr>
<tr>
<td>Utterance: Visible</td>
<td>2.28</td>
<td>2.28</td>
<td>1.00</td>
<td>0.33</td>
</tr>
<tr>
<td>Context * Utterance</td>
<td>−1.29</td>
<td>3.66</td>
<td>−0.35</td>
<td>0.73</td>
</tr>
</tbody>
</table>

Table 3  Experiment 1: optional activities analysis.
Informationally Redundant Utterances Alter Prior Beliefs

Figure 3  Experiment 1: *optional* activities analysis.

### 4.3 Discussion

The results of the first experiment indicate that comprehenders do in fact perceive informational redundancy, in the form of descriptions of overly predictable events, as a violation of conversational norms, and that they attempt to resolve this violation by reinterpreting the events described as atypical. These results provide evidence that informational redundancy constitutes a maxim, or conversational norm violation, and that comprehenders react to it as they typically do to maxim violations – by assuming an implied non-literal meaning that resolves the violation. This runs in some contradiction to Grice’s initial ambivalence about the existence of such a constraint, and equivocal evidence from studies of informationally redundant nominal modification.

These results raise two questions that we address in the following experiments, regarding the importance of prosody, and that of contextual support for the inference. First, an exclamation point may serve multiple purposes: it may signal surprise as to the course of described events, a speaker’s intentionality in communicating a piece of information⁶, the importance and relevance of the information conveyed to the general discourse and listener’s interests, and that the information preceding the exclamation point constitutes an “encapsulated” message in its own right (rather than serving as a temporal or causal anchor⁷). Although it could be argued that the exclamation point (often a signal of surprise, Rett 2011) forces the ‘atypical event’ interpretation independent of utterance informativity, this is not a tenable

---

⁶ I.e., the speakers displays clear and conscious intent to draw to the listener’s attention 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!*.
Ekaterina Kravtchenko and Vera Demberg

explanation, as no signs of a similar effect are present in any of the ‘informative’ event descriptions.

Therefore, the first question is: how generalisable 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 remarkably little work on the question of which contextual cues specifically people employ in computing context-dependent inferences, as well as how these cues influence final pragmatic interpretation. To test this, in Experiment 2 we use a discourse marker (Oh yeah, and...) which does not explicitly signal surprise, in order to frame the event description as intentionally conveyed, important/relevant to the topic at hand, and as an “encapsulated message.”

The second question raised is whether informational redundancy itself is sufficient to trigger such an inference. As mentioned previously, we start from the premise that rational speakers mention only that which cannot be automatically inferred by the listener, to lower the time and effort needed to get a message across. A charitable listener may be expected to expend considerable effort on rescuing the assumption of a cooperative or rational speaker (Davidson 1974). If only events under a certain threshold of predictability deserve mention (assuming a rational speaker), then listeners should conclude that the event mentioned is relatively atypical.

On the other hand, implicatures must be calculable (Levinson 2000) – that is, particularly for non-generalized (context-sensitive) implicatures, the context must offer sufficient support that the reader can guess the speaker’s intent with reasonable certainty. Likewise, while rational speakers may only mention atypical events, forcing a listener to expend significant effort on recovering an utterance’s intended meaning is arguably non-rational behavior itself. In addition, the degree of “intentionality” on the part of the speaker (also signalled in our 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. To test whether informational redundancy itself is sufficient for triggering the inference, or whether some amount of discourse or prosodic contextual support is necessary for its generation, in Experiment 3 we present readers with the same task and stimuli, but strip the event description of prosodic or discourse cues supporting the inference.

5 Experiment 2: Implicit Discourse Support

The second experiment tests whether the effect, of informationally redundant event descriptions being interpreted by readers as signalling event atypicality, is generalizable. To do so, we replace the exclamation point with a non-prosodic discourse marker that signals speaker intentionality and utterance relevance (but crucially, not
surprise). In this experiment, we frame the informationally redundant event description as an apparent recalling of information specifically intended to be mentioned to the listener, 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 event having occurred, nor explicitly support the intended inference otherwise – and in contrast to the exclamation mark in Exp. 1, is a non-prosodic manipulation of the event description. We therefore consider it a good test of whether the effect generalizes beyond the specific context used in the first experiment.

5.1 Methods

Participants

400 participants, located in the US, were recruited on Amazon Mechanical Turk.

Design

The design of this experiment was motivated by the same factors as Experiment 1 - with the exception of how the event description was framed. In Experiment 1, contextual support for the inference was provided in the form of an exclamation point:

(10) “John went grocery shopping. **He paid the cashier!**

In Experiment 2, in contrast, we framed the same utterance as a piece of information the speaker had just recalled, apparently having previously intended to mention it to the listener:

(11) “John went grocery shopping. **Oh yeah, and he paid the cashier.**

The *oh yeah...* discourse marker does not conventionally signal surprise, and therefore does not specifically signal the inference that we are testing for. It does, however, imply substantial speaker intent behind conveying precisely this message, the importance and relevance of the message to the current discourse and listener, as well as that the message is able to stand alone, and is not intended to simply serve as causal or temporal scaffolding for a further message/event.

As in Experiment 1, we used the same discourse marker in the remaining conditions, where we test whether altering the common ground causes the inference to fall through, and whether similar descriptions of optional events fail to trigger any similar inferences (as we expect).
Materials

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

(12)  \[1a/b\] John often \{goes to his local supermarket, as it’s close by_{a\text{-typical}} | doesn’t pay at the supermarket, as he’s typically broke_{b\text{-atypical}}\}.

[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] She commented: “John went shopping. [4a/b] Oh yeah, and he \{paid the cashier_{a\text{-conventionally predictable}} | got some apples_{b\text{-optional}}\}.

[5] I just saw him in the living room.”

Procedure

The procedure was identical to that of Exp. 1.

Measures

The same response measures as in Exp. 1 were used to estimate prior beliefs, updated beliefs, and belief change.

5.2 Results

As in Experiment 1, to determine whether participants made inferences regarding activity typicality, we modeled belief change - the difference between prior and updated beliefs. Conventionally predictable and optional activities were again modelled separately. All factors were effect/sum coded.

*Conventionally Predictable Activities*

As we predicted, prior belief ratings for typical context events showed that these events are judged to be highly typical (83.65), and atypical context events are judged to be moderately atypical (39.63). As in Experiment 1, updated beliefs about the typicality of typical context events were significantly lower (73.85). In this experiment, we also found a small change in ratings between prior and updated beliefs for the atypical context condition (from 39.63 to 44.18), see also Figure 4. However, this change is in the opposite direction as the belief change for the
typical context condition. That is, in the atypical context (the agent is introduced as e.g. not usually paying), the belief in the agent usually paying increases when one such occurrence is explicitly discussed in the narrative. This change is consistent with a trend in the same direction which we observed in Experiment 1, in the same condition, as well as with trends of belief change in the optional activity conditions.

A linear mixed effects regression analysis, fitted using $\chi^2$-test model comparison, showed an interaction between context and utterance visibility that is driven primarily by lowered activity typicality ratings, when the readers see the utterance in a typical context ($\beta=-15.44$, p<.001). We used the maximal model, with by-subject random intercepts and slopes for context and utterance presence in the story, as well as by-item random intercepts and slopes for both factors and their interaction. A plot illustrating the interaction can be seen in Figure 4, which shows a pattern of results that is remarkably quantitatively and qualitatively similar to that of Exp. 1.

<table>
<thead>
<tr>
<th>Coef $\beta$</th>
<th>SE($\beta$)</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>60.06</td>
<td>2.19</td>
<td>27.40</td>
</tr>
<tr>
<td>Context: Typical</td>
<td>37.32</td>
<td>3.98</td>
<td>9.37</td>
</tr>
<tr>
<td>Utterance: Visible</td>
<td>-2.05</td>
<td>1.72</td>
<td>-1.19</td>
</tr>
<tr>
<td>Context * Utterance</td>
<td>-15.44</td>
<td>2.75</td>
<td>-5.61</td>
</tr>
</tbody>
</table>

Table 4  Experiment 2: conventionally predictable activities 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.

These results support our prediction that, as long as the ‘atypicality’ inference has contextual support, in the form of implied speaker intentionality, message relevance, or message “encapsulation,” readers perceive the informationally redundant utterances as abnormal and make compensatory inferences (of event atypicality), regardless of whether prosody or other markers typically associated with surprisal are present.

Optional Activities

As in Experiment 1, there was little change in ratings of optional events, from prior beliefs (typical: 38.77; atypical: 34.30) to updated beliefs (typical: 42.29; atypical: 42.33).

A linear mixed effects regression analysis, fitted using $\chi^2$-test model comparison, showed that estimates of activity typicality increase slightly when the utterance
Figure 4  Experiment 2: conventionally predictable activities analysis. This plot shows changes in activity typicality estimates depending on whether the utterance is present, as well as whether the context causes the utterance to be perceived as atypical. Violin plots, overlaid with box plots, show the distribution of estimates. Circles represent mean values. Arrows show significant differences between before/prior and after/updated ratings.

describing the optional activity (see Table 5) is visible, with the difference significant in the case of the atypical context ($\beta = -8.53, p < .01$).

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 we made prior to running the experiments: simply mentioning a non-inferable, or concisely informative event may increase the perception of its typicality, by providing some evidence that, e.g., John is at least an occasional apple purchaser. While this result also shows some indication that the context alteration may in itself cause a change in activity typicality estimates, at least given how the utterance is framed in this experiment, and is unexpected, it does not change our interpretation of the primary effect, given that there is no rating decrease (significant or numerical) in the typical condition. We therefore leave it aside for future exploration.

5.3 Discussion

Together with Experiment 1, these results show that readers find conceptual informational redundancy abnormal, and will make compensatory inferences to reconcile apparent informational redundancy with their expectations of utterance utility. This
Table 5  Experiment 2: optional activities analysis.

<table>
<thead>
<tr>
<th></th>
<th>Coef $\beta$</th>
<th>SE($\beta$)</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>39.66</td>
<td>2.41</td>
<td>16.48</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Context: Typical</td>
<td>2.87</td>
<td>2.18</td>
<td>1.32</td>
<td>0.20</td>
</tr>
<tr>
<td>Utterance: Visible</td>
<td>5.95</td>
<td>2.38</td>
<td>2.49</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Context * Utterance</td>
<td>-5.34</td>
<td>2.54</td>
<td>-2.10</td>
<td>&lt;.05</td>
</tr>
</tbody>
</table>

indicates that the effect is generalisable, and not dependent on conventional indicators of event atypicality, such as exclamatory intonation.

This experiment still leaves open the question of whether informationally redundant utterances trigger compensatory inferences only in the presence of cues that support the inference, or if the effect is independent of contextual cues. If it is dependent on discourse context that supports the inference, this would suggest one of the following. Comprehenders may simply be unwilling to expend substantial effort on decoding an inference in the absence of supportive cues. Alternately, they may stop short in their efforts, on the assumption that it is more likely that speakers would violate expectations of informational utility, than that they would provide insufficient contextual support for a comprehender to decode a non-literal meaning. Finally, they may be tolerant of informational redundancy, unless context suggests that the redundancy has a ‘point.’ Experiment 3 attempts to answer this question by presenting the same task and materials to participants, but removing the prosody or discourse markers that support the inference.
6 Experiment 3: Removing Contextual Support

The results of Experiments 1 and 2 leave open the question of whether the atypicality effect is dependent on context that supports the inference, or applies independently of discourse context. While some degree of contextual support is necessary for a (non-generalized) conversational implicature to be calculated (Levinson 2000), informational redundancy should be problematic in itself, and the implicit support offered by the implicit prosody and discourse marker in Experiments 1 and 2 is relatively minor. Further, outside the effects of contrastive prosody (Kurumada et al. 2012), there is little evidence of semantically ‘vacuous’ prosodic changes to utterances having a substantial effect on implicature generation.

The third experiment is intended to clarify the range of discourse contexts in which the atypicality effect applies. The event descriptions in Exp. 1 ended in exclamation points, which implicitly signal the speaker’s intentionality in uttering the specific sentence, as well as the relevance of the sentence to the discourse (assuming the speaker is at all cooperative). Further, exclamation points may signal surprise, which could increase the magnitude of the ‘surprisal’ effect in the informationally redundant condition – although the effect cannot in full be attributed to this, as in that case one would expect the exclamation point to have a similar effect in other conditions, or to produce a larger effect than that found in the second experiment, where there was no clear signal of surprise.

Similarly, the utterances in Exp. 2 were framed as a piece of (presumably sufficiently important and relevant) information that the speaker suddenly recalled, with prior intent to communicate to the listener. To investigate whether implicit contextual support for the implicature has an influence on the strength of the atypicality effect, we designed a third experiment which differed only in the absence of contextual support for the implicature, or indeed for the importance/informativity of the event description (in the form of prosody or discourse marker). Our prediction is that while the effect may be attenuated, listeners should nevertheless make a measurable attempt to compensate for a violation in expected informational utility.

6.1 Methods

Participants

400 participants, located in the US, were recruited on Amazon Mechanical Turk.

Design

The design was motivated by the same factors as Experiments 1 and 2, but all implicit contextual support was removed from the event description:
Informationally Redundant Utterances Alter Prior Beliefs

(13) “John went grocery shopping. He paid the cashier.”

In this case, there is no specific support for the inference, or any signal of 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, e.g.:

(14) “John went grocery shopping. 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:

(15) TYPICAL CONTEXT

[1a/b] John often {goes to his local supermarket, as it’s close by a-typical | doesn’t pay at the supermarket, as he’s typically broke b-atypical}.

[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.


Procedure

The procedure was identical to that of previous experiments.

Measures

The same response measures as in the previous experiments were used to estimate prior beliefs, updated beliefs, and belief change.
6.2 Results

As in previous experiments, we modeled the difference between prior and updated beliefs. Conventionally predictable and optional activities were modeled separately. All factors were effect/sum coded.

**Conventionally Predictable Activities**

As in the previous experiments, prior belief ratings showed typical context events to be highly typical (84.21), and atypical context events to be less typical (43.02).

In contrast to the previous two experiments, and our predictions, prior and updated beliefs did not significantly differ in this experiment, for the typical context condition (84.21 to 84.27). In the atypical context condition, we again observed a significant increase in belief change, see Figure 6 (43.02 to 47.96).

A linear mixed effects regression analysis, fitted using $\chi^2$-test model comparison, showed an interaction between context and utterance visibility, but driven by an unexpected contrast in the atypical context condition ($\beta=-8.324$, $p<.01$). We used the maximal model, with by-subject random intercepts and slopes for context and utterance presence in the story, as well as by-item random intercepts and slopes for both factors and their interaction. A plot illustrating the interaction can be see in Figure 6.

<table>
<thead>
<tr>
<th>Coef $\beta$</th>
<th>SE($\beta$)</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>64.28</td>
<td>2.22</td>
<td>28.92</td>
</tr>
<tr>
<td>Context: Typical</td>
<td>40.03</td>
<td>3.89</td>
<td>10.30</td>
</tr>
<tr>
<td>Utterance: Visible</td>
<td>4.09</td>
<td>1.67</td>
<td>2.45</td>
</tr>
<tr>
<td>Context * Utterance</td>
<td>$-8.32$</td>
<td>2.73</td>
<td>$-3.05$</td>
</tr>
</tbody>
</table>

**Table 6** Experiment 3: conventionally predictable activities 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.

These results indicate that, entirely contrary to our predictions and the results of Exp. 1 and 2, when an easily inferable activity is overtly mentioned in a typical world context, without implicit prosody or discourse markers signalling, at minimum, the clear intentionality and presumed discourse relevance of the utterance, listeners do not appear to infer anything regarding the typicality of the activity.
Experiment 3: conventionally predictable activities analysis. This plot shows changes in activity typicality estimates depending on whether the utterance is present, as well as whether the context causes the utterance to be perceived as atypical. Violin plots, overlaid with box plots, show the distribution of estimates. Circles represent mean values. Arrows show significant differences between before/prior and after/updated ratings.

In addition, it appears that participants are making pragmatic inferences only in the atypical context, and that these inferences reflect participants’ general tendency to assume that less inferable activities are more typical if they are mentioned overtly.

Optional Activities

There was some change in ratings of optional events, from prior beliefs (typical: 40.242; atypical: 32.676) to updated beliefs (typical: 45.400; atypical: 46.004).

A linear mixed effects regression analysis, fitted using $\chi^2$-test model comparison, again showed, contrary to predictions and the results of Exp. 1, both a main effect of utterance visibility, which indicates that participants are drawing inferences of greater activity typicality from mentions of non-inferable events, and an interaction that shows a greater effect of overt event mention in the atypical context condition (as in the previous experiment, this effect is somewhat difficult to interpret, and given no present theoretical motivation or impact, we leave it aside for now).
Ekaterina Kravtchenko and Vera Demberg

<table>
<thead>
<tr>
<th></th>
<th>Coef $\beta$</th>
<th>SE($\beta$)</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>41.25</td>
<td>2.94</td>
<td>14.03</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Context: Typical</td>
<td>3.72</td>
<td>2.16</td>
<td>1.72</td>
<td>0.10</td>
</tr>
<tr>
<td>Utterance: Visible</td>
<td>9.88</td>
<td>1.92</td>
<td>5.15</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Context * Utterance</td>
<td>−6.68</td>
<td>2.46</td>
<td>−2.71</td>
<td>&lt;.05</td>
</tr>
</tbody>
</table>

Table 8  Experiment 3: *optional* activities analysis.

6.3 Discussion

In contrast to the results of the first two experiments, these results show that, when informationally redundant utterances are presented without a discourse context which implicitly supports the inference (via prosody or a discourse marker), comprehenders are unlikely to draw *atypical* inferences. This is consistent with the premise that, while rational speakers may typically avoid making utterances that have no literal or implied informational utility, and while such utterances may prompt ‘compensatory’ inferences on the part of listeners (which increase the informational utility of such utterances), such inferences must be sufficiently plausible in context and relatively easy for listeners to calculate (Levinson 2000).

In short, in the absence of contextual support for the inference (a signal of clear speaker intentionality, the event description being an ‘encapsulated’ message, some indication of relevance to the listener/discourse), it appears that comprehenders are either unable to successfully resolve the violation of conversational norms, resolve it in a manner that is not reflected in our response measures (possibly assuming that
the speaker is simply not acting rationally), or do not detect the violation 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 rational conversational choices (Davidson 1974). However, it is also possible that listeners are less ‘charitable’ when presented with oddly phrased psycholinguistic stimuli in an artificial setting – as well as less motivated on expending 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.

Similarly, when the atypical event interpretation is not available or salient, comprehenders may make other inferences on the basis of informationally redundant utterances. For example, they may simply assume that the speaker is strange or ‘irrational,’ or perhaps that they are having production difficulties, or are/were in the process of planning a more informative utterance (where, for example, the informationally redundant event description might serve as a temporal/causal anchor; see Example (14)). Determining which strategies comprehenders do in fact resort to, and in which contexts, is left to future research. Finally, there is the possibility that, given the relatively artificial and non-interactive setting, comprehenders are processing the utterances at a relatively shallow level, and absent some (prosodic, discourse) indication that they should pay attention to an utterance, they simply gloss over it (Sanford et al. 2006). Similarly, it has frequently been observed that comprehenders often do not make expected generalized inferences in behavioral studies, for reasons that are not yet fully known (e.g., Noveck & Posada 2003). Determining whether this plays a role in our studies is left to future work, as is the question of whether similar effects can be observed in less artificial, and/or more interactive settings.

7 Is the Effect of Typicality on Pragmatic Inferences Gradient?

Figure 8 plots the average typicality estimate, before vs. after the target utterance, for each item in each condition, for all three experiments. The diagonal dashed line demonstrates what the null-hypothesis would predict: i.e., no effect of the utterance on belief change (‘before’ ratings corresponding to ‘after’ 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 an atypicality 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), we see that for typical-world, conventionally-predictable activities (e.g., paying the cashier in a typical world), most datapoints
Figure 8 The plots show by-item belief change for all conditions of our three experiments. The dotted diagonal line represents the null hypothesis; i.e., what we would expect the data to look like if the critical utterance had no effect on beliefs. The blue line is a regression line with 95% CIs across all conditions.
fall below the line, indicating an atypicality inference. Interestingly, we also see a gradual ‘trend’ towards atypicality in the items from the typical-world optional-activity condition (e.g., buying apples in a typical world): typical optional items that are similar to typical predictable items, in terms of prior typicality estimates, tend to result in atypicality inferences. In contrast, typical optional items with low prior typicality estimates show the opposite effect: i.e., if it’s mentioned that an individual engaged in a particularly unusual 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), we again see a gradual effect of prior beliefs in activity typicality on the likelihood of an atypicality inference, but this time the slope of the regression line has shifted upwards (Exp. 1: $\beta=0.64$; Exp. 2: $\beta=0.61$; Exp. 3: $\beta=0.73$). We still see that there is a gradient difference between highly expected vs. optional events. For those items that we classified as conventionally predictable, but which our participants regarded as close to optional, we can for instance see that people did not draw atypicality inferences, but rather either inferred that the activity is relatively typical for the individual, or did not alter their beliefs in response to seeing the utterance. For activities rated mostly highly typical, there is still a small tendency for participants to make atypicality inferences. This suggests that even in the full stop (period) condition, atypicality inferences are in principle possible.

Taken together, we 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 atypicality inferences will arise for typical-world predictable-activity event mentions. Furthermore, we can see that the effect of predictability on atypicality inferences is clearly gradient rather than binary: more predictable activities generally elicit larger atypicality effects. The figures also show that the conventionally “predictable” items initially rated by participants to be least predictable, and the “optional” items initially rated most predictable, fall into the same general range with respect to their initial typicality ratings, and likewise show the same degree and direction of belief change. This clearly demonstrates that the binary experimental distinction of predictable and optional activities in fact represents a wide range of schema-relevant activities at all levels of typicality.

8 General Discussion

Our results have several implications for pragmatic theory and experimental study of pragmatic inferences – specifically in the area concerning comprehenders’ (context-dependent) sensitivity to utterance informativity, and the degree to which comprehenders attempt to reconcile, or draw inferences on the basis of informativity violations. Previous observations in theoretical pragmatic literature, as well as experimental
studies of ‘overinformative’ nominal modification, have suggested that comprehenders are relatively unlikely to penalize informational redundancy, and often do not appear to draw additional inferences from use of redundant expressions (Davies & Katsos 2009, Engelhardt et al. 2006). This is contrary to Gricean (or rational speaker) norms, which suggest that informational redundancy is suboptimal, at least from the speaker’s perspective – although Grice (1975: p. 46) himself doubted the existence of a general constraint against overinformativeness. Our results confirm that comprehenders attempt to resolve violations of expected utterance informativity – including in cases that do not directly impair comprehension, as in the case of informational redundancy.

Further, we illustrate a case where 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. This is a strategy that, in contrast to shifting assumptions about intended utterance meaning, has not received much attention to date (but see Degen et al. 2015). Finally, we show that semantically ‘vacuous’ utterance features (those that do not make any conventional contribution to meaning), in the form of implicit prosody or discourse markers, significantly influence the extent to which comprehenders are willing to make an inference predicted by rational speaker norms. Aside from the case of contrastive prosody (Kurumada et al. 2012, Ward & Hirschberg 1985), this has not to date been systematically observed or investigated in formal or experimental literature. On the experimental side, we provide a novel method of measuring comprehender beliefs, as well as belief change in response to experimental stimuli. This should allow for better understanding of pragmatic inferences which are based primarily on reasoning about context and world knowledge (rather than simply the specific proposition expressed).

The possibility of informationally redundant utterances giving rise to ‘compensatory’ pragmatic inferences was tentatively discussed by Grice (1975), but not clearly addressed by later theories. Most of these theories, as well as some theories of online language production/processing (Aylett & Turk 2004, Jaeger 2010, Levy & Jaeger 2007), predict that speakers should avoid excessive redundancy in order to conserve (articulatory, cognitive) effort, and in order to communicate their message in a maximally short amount of time (provided comprehension is not impaired). However, there has remained an open question of whether comprehenders, acting on the expectation that (typical) speakers will be neither under- nor overinformative (Levinson 2000), will either penalize or draw pragmatic inferences from utterances that are apparently informationally redundant. One may expect that they would do so, in order to reconcile any apparent deviation from their assumption that a speaker will behave rationally, and that any apparent deviations from rationality must be purposeful (and further, meaningful). However, as redundancy does not interfere with comprehension, it has repeatedly been questioned whether listeners penalize it
(there is limited evidence that they do), and whether they draw ‘compensatory’ inferences from informationally redundant utterances (existing experimental evidence has suggested that, outside of inherently contrastive or scalar terms, they do not).

As Walker (1993) points out, informationally redundant utterances may not be so anomalous, or rare, as might be predicted by theories of rational communicative behavior (Grice 1975, Levinson 2000). Most theories predicting their scarcity or marked anomaly assume a) unlimited working memory on the part of listeners, and b) ability to easily or quickly generate all inferences (intended to be) conveyed by the discourse. Redundant utterances can serve a narrative function, and are often used to draw attention to salient aspects of a discourse, or to showcase evidence for certain beliefs. And, as shown in this study, they can be reinterpreted by listeners to reflect new or useful information, even when such an interpretation is not explicitly supported. This indicates that the phenomena that we are dealing with are not edge cases, and should be accounted for by pragmatic theories and formal language models. Further, it should neither be assumed that such utterances are a) absent in normal speech, and therefore do not need to be accounted for; or b) while present in speech, do not have any effect on comprehension or processing, and therefore have no implications for formal modeling.

The results firmly point to listeners penalizing, and/or attempting to compensate for informational redundancy, although it does not interfere with comprehension of the events described, and strictly is suboptimal only from the speaker’s perspective (unlike other strains of pragmatic violation, particularly underinformativity or underspecification). This suggests that listeners do monitor utterances for (speaker-centric) optimality, and are sensitive to deviations from such. Further, although redundancy would not appear to negatively impact comprehension, listeners may have an interest in speakers being relatively concise, as it allows for them to receive more information in a shorter span of time, and allocate their attention appropriately. Excessive redundancy may make it more difficult to follow the point of a conversation, or to reliably distinguish important information. Optimally, all utterances should therefore have some degree of informational utility.

The results of the 3rd experiment, intriguingly, suggest however that informational redundancy is in itself not necessarily sufficient to trigger a compensatory pragmatic inference. Semantically vacuous features such as exclamation points, or some discourse markers, significantly increase the likelihood that comprehenders will make informativity-based inferences – an observation which (outside the limited scope of contrastive prosody, or discourse markers which signal a specific discourse relation) has not received attention in theoretical or experimental pragmatic literature, and which may well extend to other pragmatic phenomena. It is not altogether clear whether this effect is due to listeners interpreting the prosody and discourse markers as signals of surprise, or simply as increased focus on the event description
as an important, intentionally conveyed, and discourse-relevant piece of information. It should be noted that any markers which may explicitly convey surprise to comprehenders would support, but could not fully explain, the inferences drawn (given that the ‘surprisal’ effect does not extend to the ‘informative’ conditions – i.e., it still remains clear that the informationally redundant condition is anomalous).

A question that is raised for future research is whether encountering informationally redundant utterances results in processing difficulty on the part of comprehenders. We see several different potential sources of processing difficulty associated with such utterances, resulting from the processing of the surface form on the one hand, and from the process of making any compensatory pragmatic inferences on the other. Firstly, there could be processing difficulty associated with the predictability of the surface form of the utterance. *John paid the cashier!* is an unpredictable utterance in a typical context, as paying the cashier is normal, and mentioning it overtly is redundant – reading such an unpredictable utterance should cause some difficulty. This is a type of processing difficulty that current notions of surprisal can capture: (Smith & Levy 2013) predict that difficulty should be the same for all strings, or surface forms, which are matched in predictability. We argue that this may be not be enough to predict processing cost: consider, for example, two strings hypothetically matched in predictability: *John paid the cashier!* and *John ran over the cashier’s foot with his cart!*.

Considering only surface-level predictability doesn’t take into account that the utterances are unpredictable for entirely different reasons (dispreference for redundancy vs. event unpredictability) – and further, that the first utterance may contribute additional processing cost due to encountering pragmatic abnormality, or to the cost of making a pragmatic inference to ‘compensate’ for the redundancy. In this case, despite identical surface-form predictability, conceptually redundant utterances would be associated with greater processing difficulty.

Further, it might also be the case that conceptually redundant utterances are relatively easy to process, due to the ease of semantic integration, regardless of their apparent anomaly, and surface-level unpredictability (*surprisal*). In this case, we may see lower processing difficulty for conceptually predictable vs. unpredictable utterances, even when they are matched in surface-level predictability. This would likewise require that we revise the linking hypothesis between *surprisal* and processing difficulty. To note, however: while listeners presumably must engage in some form of additional processing to draw non-generalized inferences, which should be associated with a processing signature, it’s not clear whether a) this would be a clear signature detectible by typical (relatively noisy) methods; b) enough participants would be drawing such an inference, or detecting the pragmatic abnormality at all, for the effect not to wash out at the level of processing difficulty. We next briefly discuss whether a detectable processing signature should be associated with making a context-dependent inference of the type considered in this paper.
Informationally Redundant Utterances Alter Prior Beliefs

Context-dependent implicatures particularly are presumed to incur processing cost (Sedivy 2007), although it has recently been argued that they may be processed relatively effortlessly, provided the context supports the inference (Degen & Tanenhaus 2015, Grodner et al. 2010). It may therefore be fruitful to investigate the processing of these utterances using online measures such as eye-tracking or self-paced reading. Further, as typical speaker behavior is to avoid utterances composed of very conceptually predictable, or easily inferable information, the predictability of informationally redundant utterances at a word level should be comparatively low. Online measures such as reading times have been argued by some to reflect the predictability of linguistic events, rather than the conceptual predictability of the scenarios those events describe (and the corresponding ease of integrating such information into a mental situation model; Smith & Levy 2013). There is evidence, however, that comprehenders predict at multiple levels: for example, the event or schema level, as well as at a more fine-grained lexical level that may be informed by higher-level predictions (Kuperberg & Jaeger 2016). A question remains regarding how high-level pragmatic expectations interface with event- or form-level expectations – e.g., if an event and associated lexical items are predicted given the schema or situation model, but also predicted to go unmentioned in isolation, given conversational norms with respect to informativeness.

In the case of surprisal theory (Hale 2001, Levy 2008), this may have interesting implications, given that the empirical link between processing times and surprisal values is based on measures that are largely obtained using formal language models. If, in fact, informationally redundant utterances result in longer reading times, it’s unclear how formal models could accurately generate the (presumably relatively high) surprisal values one would expect for those utterances, and that one would require to test the link between lexical surprisal and processing times. For example, in the case of the conventionally predictable event utterance, in the typical vs. atypical world context, the event description (John paid the cashier) consists of exactly the same string of words, and the preceding context is identical stretching over multiple preceding sentences. The utterance is informationally redundant in the typical context, and concisely informative in the atypical context. Simple 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.

It should 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).
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. However, the general prediction such models would make is that the more congruent an event is with an invoked 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 events into different grades of predictability, such that utterances describing moderately predictable events are easier to process than those describing not-so-predictable events, yet those describing very predictable events incur difficulty. In light of this, we suggest that to predict any processing difference between informationally redundant and concisely informative utterances, formal models of language comprehension would need to incorporate some form of pragmatic reasoning.

Although attempts to build formal or computational language models may appear to have limited relevance to how humans process language – which intuitively appears to be 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. 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 (i.e., whether a speaker is behaving rationally, or 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 processing speed; or build formal language models which make specific, testable predictions regarding the influence of these cues on processing and comprehension.

There are several limitations to these studies - first, the range of ‘compensatory’ inferences that comprehenders might draw from informationally redundant descriptions may extend well beyond what we tested in this series of experiments. For instance, in the absence of a possible ‘informative’ interpretation, as that suggested by our response measure, comprehenders may simply assume that a speaker is being uncooperative, having some production difficulty, or has unconventional speaking patterns. There is also the possibility that informationally redundant event descriptions, especially as seen in Experiment 3, are initially interpreted as likely (possibly aborted) temporal or causal anchors for more ‘interesting’ information relevant to
Informationally Redundant Utterances Alter Prior Beliefs

the situation under discussion. 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 that overall contributes previously unknown, or not easily inferable information.

The work presented here can be extended in the future in several ways. Although readers fairly consistently came to the interpretation that explicit mention of an overly predictable event implies its atypicality (in one way or another), it’s possible that some readers reached other interpretations as to what the speaker intended (such as narrative framing for another event, as speculated above). They may also have made the conclusion that the speaker is somehow ‘abnormal,’ or were unable and possibly unmotivated to resolve the apparent anomaly.

Overall, this series of experiments shows that comprehenders react to informationally redundant utterances by assigning them a more ‘informative’ non-literal meaning, thus increasing their informational utility. This suggests that listeners actively keep track of expected speaker/utterance informativity, and adjust their interpretations of utterances when such expectations are violated. This is consistent with pragmatic accounts of what constitutes ‘cooperative’ speaker behavior, as well as how listeners attempt to resolve behavior that appears non-rational or non-cooperative at face value – although it was long speculated that informational redundancy may have no negative impact on, and consequently no effect on comprehension (Grice 1975). However, as the last experiment shows, the effect is heavily dependent on how the utterance is framed in the discourse, including semantically vacuous manipulations of prosody and discourse markers, which comprehenders are shown to systematically utilize as cues regarding the speaker’s intent in uttering the phrase.

We thus provide robust evidence that informational redundancy is perceived as anomalous, and that when needed, comprehenders alter their situation models in order to accommodate it. In the case of informationally redundant event descriptions, comprehenders infer that the events described are more atypical than initially assumed, which renders them informative enough to warrant mention. However, we also find that comprehenders make such inferences only in the presence of (semantically vacuous) contextual support for the intended interpretation, in the form of implicit prosody or discourse markers that indicate surprise, intentionality, and/or relevance.

Although our results show that informational redundancy is perceived as abnormal, they provide some support for the observation that, unlike underinformativity, it is not necessarily problematic for listeners (Grice 1975), despite at times being treated similarly in formal literature. Listeners are able to justify the use of exces-
sive information by altering their prior beliefs about individuals’ behavior, or more broadly, the common ground. The results also complement work in the dialogue literature (Walker 1993), which shows that informationally redundant utterances are frequently used to convey ‘informative’ non-literal content. Finally, they raise questions regarding which discourse cues are systematically tracked by comprehenders, as well how those cues are integrated during pragmatic interpretation, including of utterances that have to date been treated as either too complex and idiosyncratic to study systematically.

References


A Appendix: Task Validation

A control experiment, using a blunt signal of activity *atypicality*, was run to confirm that participants behave as expected using our paradigm, producing *prior* and *updated* estimates in the expected range. When an event description is presented with a very strong conventional marker of activity atypicality (*You’d be surprised, but...*), participants should strongly decrease their typicality estimates, regardless of condition. This would confirm that, informational redundancy aside, the *informationally redundant* condition is not the only one vulnerable to such a manipulation – which all conditions in our experiments should in principle be subject to.

We predict that participants’ activity typicality estimates should be significantly lowered in all conditions, as the discourse context of each event description indicates that the activity is, unambiguously, quite atypical compared to the addressee’s prior expectations (whether those expectations of typicality were very high, or moderate). If participants do in fact lower their estimates in each condition, it would indicate, in concert with the other experiments presented in this paper, that the method we use accurately gauges participants’ relative typicality estimates, and that all conditions are in principle subject to the *atypicality* effect.

A.1 Methods

Participants

400 participants, located in the US, were recruited on Amazon Mechanical Turk.

Materials

The same 24 stimuli were used as in the previous experiments. Each utterance was prepended by an expression strongly and conventionally indicating activity atypicality, such as *You’d never believe it, but...*:

(16) Typical context

[1a/b] John often \{goes to his local supermarket, as it’s close by\_a-typical | doesn’t pay at the supermarket, as he’s typically broke\_b-atypical\}.

[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] She commented: “John went shopping. [4a/b] *You’d be surprised, but*
he \{ paid the cashier \textit{a-conventionally predictable} | got some apples \textit{b-optional} \}. [5]  
I just saw him in the living room.”

**Procedure**

The procedure was identical to that of the other experiments.

**A.2 Results**

As in the three experiments reported, we modeled the difference between prior and updated beliefs. Conventionally predictable and optional activities were modeled separately. All factors were effect/sum coded.

**Conventionally Predictable Activities**

Participants who did not see the utterance rated typical context events as highly typical (81.692), and atypical context events as moderately atypical (38.705). Participants who saw the informationally redundant event description rated the event as substantially less typical (43.218, vs. 81.692). There was likewise a significant lowering of ratings in the atypical context condition (28.024, vs. 38.705).

A linear mixed effects regression analysis, fitted using $\chi^2$-test model comparison, showed an expected main effect of utterance visibility ($\beta=-24.55, p<.001$), indicating that participants rated all described events as relatively atypical, as expected. It also showed a significant interaction ($\beta=-28.04, p<.001$). However, this cannot be meaningfully interpreted, given a clear floor effect in all conditions. We used the maximal model, with by-subject random intercepts and slopes for context and utterance visibility, as well as by-item random intercepts and slopes for both factors and their interaction. A plot illustrating the interaction can be see in Figure 9.

**Optional Activities**

As for the conventionally predictable activities, there was a substantial lowering of ratings in the typical (24.307 vs. 41.699) and atypical (23.595 vs. 35.200) contexts.

A linear mixed effects regression analysis, fitted using $\chi^2$-test model comparison, showed an expected main effect of utterance visibility ($\beta=-14.05, p<.001$), again indicating that participants rated all described events as relatively atypical, as expected. It also showed a significant interaction ($\beta=-6.46, p<.05$), which is likewise somewhat difficult to interpret given the floor effect, and which does not affect our current conclusions. We used the maximal model, with by-subject random intercepts and slopes for context and utterance presence in the story, as well as by-item random
Informationally Redundant Utterances Alter Prior Beliefs

---

<table>
<thead>
<tr>
<th></th>
<th>Coef $\beta$</th>
<th>SE($\beta$)</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>47.98</td>
<td>1.86</td>
<td>25.78</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Context: Typical</td>
<td>29.40</td>
<td>3.12</td>
<td>9.41</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Utterance: Visible</td>
<td>-24.55</td>
<td>2.05</td>
<td>-12.00</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Context * Utterance</td>
<td>-28.04</td>
<td>3.99</td>
<td>-7.03</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

**Table 9**: Control Experiment: *conventionally predictable* activities 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.

A plot illustrating the interaction can be see in Figure 10.

---

<table>
<thead>
<tr>
<th></th>
<th>Coef $\beta$</th>
<th>SE($\beta$)</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>31.27</td>
<td>1.92</td>
<td>16.32</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Context: Typical</td>
<td>3.63</td>
<td>1.90</td>
<td>1.91</td>
<td>0.07</td>
</tr>
<tr>
<td>Utterance: Visible</td>
<td>-14.05</td>
<td>2.84</td>
<td>-4.94</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Context * Utterance</td>
<td>-6.46</td>
<td>2.56</td>
<td>-2.52</td>
<td>&lt;.05</td>
</tr>
</tbody>
</table>

**Table 10**: Control Experiment: *optional* activities analysis.

---

A.3 Discussion

The results of the control experiment show that when comprehenders are presented with event descriptions that conventionally imply the atypicality of the event, they correspondingly rate the event as highly atypical, using our paradigm. This demonstrates that participants are in principle quite able to accommodate the atypical common ground that the *atypicality* implicatures require. Further, this confirms that all conditions are in principle subject to an ‘atypicality’ effect that is otherwise seen only in the *informationally redundant* condition, and that given an utterance context that forces that interpretation, participants behave exactly as would be predicted using our rating paradigm.
Figure 9  Control Experiment: *conventionally predictable* activities analysis. This plot shows changes in activity typicality estimates depending on whether the utterance is present, as well as whether the context causes the utterance to be perceived as atypical. Violin plots, overlaid with box plots, show the distribution of estimates. Circles represent mean values. Arrows show significant differences between *before/prior* and *after/updated* ratings.

Figure 10  Control Experiment: *optional* activities analysis.