On the nature of implicit causality and consequentiality: the case of psychological verbs

Torgrim Solstad and Oliver Bott

Department of Linguistics, Faculty of Linguistics & Literary Studies, Bielefeld University, Bielefeld, Germany

ABSTRACT

Implicit Causality (I-Caus) and Implicit Consequentiality (I-Cons) biases (Peter annoyed Mary because/and so …) have been argued to be rooted in the argument structure properties of verbs. In particular, the mirror coreference biases displayed by stimulus-experiencer and experiencer-stimulus predicates have been considered strong evidence for this approach. We provide evidence for the Asymmetry Hypothesis, stating that I-Caus and I-Cons are derived from different mechanisms. While we also assume that I-Caus is driven by verb semantics, we contend that I-Cons follows from general discourse-structural principles. Evidence is provided by four production experiments in German investigating the coreference and coherence properties of the aforementioned verb classes in detail. Experiment 1 establishes that the classes mirror each other with respect to coreference biases. Experiments 2 and 3 show, however, that there is no such symmetry with regard to coherence biases. Finally, Experiment 4 provides fine-grained evidence for the underlying strategies for providing contingency specifications.

1. Introduction

Discourse relations are the glue that connects the sentences in a discourse to a coherent whole (Asher & Lascarides, 2003; Kehler, 2002; Mann & Thompson, 1988; Sanders et al., 1992). In particular, causal relations such as explanations and consequences are at the core of what makes discourse coherent and maximally connected (Trabasso, 1985, among many others). Implicit discourse relations can be used to show that oftentimes causal relations are inferred, even in cases such as (1) in which world knowledge does not support a causal explanation (examples from Asher & Lascarides, 2003).

(1) a. [John pushed Bill, s1 [He fell, s2] (consequence(s2, s1))]
   b. [Bill fell, s3 [John pushed him, s4] (explanation(s2, s1))]
   c. [John annoyed Bill, s1 [He ate soup, s2] (explanation(s2, s1))]

Understanding how humans produce and interpret causal discourse is thus central to properly accounting for discourse processing.

Both contingency relations, explanations and consequences, have also received broad interest in the context of anaphora resolution for the well-established phenomena of Implicit Causality (henceforth, I-Caus; cf. Garvey & Caramazza, 1974) and Implicit Consequentiality (henceforth, I-Cons; cf. Au, 1986; Stewart et al., 1998). The following examples indicate that the preferred anaphora resolution strongly depends both on the verb as well as the connective (Ehrlich, 1980; Kehler et al., 2008).

(2) a. John annoyed Bill because he … (he=John)
   b. John adored Bill because he … (he=Bill)
   c. John annoyed Bill so he … (he=Bill)
   d. John adored Bill so he … (he=John)

In the psycholinguistic literature, several proposals have been made to account for I-Caus and I-Cons, often treating them as two sides of the same coin. The present paper proposes an account with very different mechanisms underlying these phenomena and will thus provide a new answer to the often posed question What are Implicit Causality and Consequentiality? (Crinean & Garnham, 2006; Hartshorne et al., 2015; Hartshorne & Snedeker, 2013; Pickering & Majid, 2007). The present proposal is tested in a text production study systematically investigating I-Caus and I-Cons. By doing so, we will look well beyond anaphoric reference into the coherence-driven mechanisms underlying causal discourse. The empirical investigation in this paper focuses on the psychological verb classes prominent from the I-Caus and I-Cons literature as they present
the most extreme test case for current proposals (including our own). The experiments reported below are part of a larger study also including a complementary set of experiments testing non-psychological verbs. These results are reported in a separate paper (Solstad & Bott, in prep.). Importantly, though, the theoretical discussion in this paper addresses the general nature of I-Caus and I-Cons for all relevant verb classes without a limitation to the kinds of verbs tested in the experiments below.

1.1. Implicit causality (I-Caus)

Implicit Causality is a property of a wide range of (mostly transitive) interpersonal verbs for which research has established strong biases for providing an explanation in which the first-mentioned argument is coreferent with either the subject or the object of the verb (Au, 1986; Brown & Fish, 1983; Garvey & Caramazza, 1974), see (3):

(3) a. Peter annoyed Mary because he …
   b. Peter adored Mary because she …

Verbs for which the explanation makes primary reference to the subject, such as annoy, are commonly referred to as NP1-bias verbs, whereas verbs that trigger explanations about the object argument, such as adore, are said to be NP2-biased. Crucially, the I-Caus bias involves a preference only and can thus be violated. Continuations that validate the expected continuation by referring to the biased argument are said to be congruent with the coreference bias, while continuations that violate this expectation are incongruent with the bias, see (4).

(4) a. congruent: 
   Peter annoyed Mary because he was making a lot of noise.
   b. incongruent: 
   Peter annoyed Debbie because she was trying to get some sleep.

I-Caus coreference biases have been assessed to be highly reliable, as shown through the application of a number of psycholinguistic methods (Bott & Solstad, 2014, 2021; Ferstl et al., 2011; Goikoetxea et al., 2008; Rudolph & Försterling, 1997). Research has focused on the biases from four main classes of verbs (Au, 1986; Brown & Fish, 1983). For two classes of psychological verbs (henceforth, psych verbs), the bias has been shown to be towards the stimulus argument, resulting in NP1 bias for stimulus-experiencer verbs like annoy and NP2 bias for experiencer-stimulus verbs like adore. A third class, agent-evocator verbs (including, among others, judgement verbs like criticize, see Au, 1986; Fillmore, 1969; McCawley, 1975), has been shown to display NP2 bias. Finally, for a number of agent-patient verbs, the bias ranges from NP1 to NP2, with no clear pattern emerging (cf. e.g. Ferstl et al., 2011).

There is evidence that I-Caus biases are cross-linguistically stable. Consistent I-Caus biases have been reported for corresponding verbs in a variety of languages including Cantonese, Dutch, English, Finnish, French, German, Italian, Japanese, Korean, Mandarin, Norwegian, Russian, Spanish, and, cross-modally, also American Sign Language (Frederiksen & Mayberry, 2021, and references therein).

Furthermore, verb classes with a strong I-Caus coreference bias have also been shown to display a coherence bias (Bott & Solstad, 2014, 2021; Kehler et al., 2008). Thus, when prompted to continue sequences such as Peter annoyed Mary or Peter admired Mary after a full stop – without a specification of the discourse relation by means of because – participants provided around 60% explanations, which was approximately three times as many as the next-most frequent relation, Result/Consequence. For the category of “non-IC verbs”, on the other hand, Kehler et al. (2008, p. 34) found only 24% of continuations to constitute explanations.

The coreference and coherence biases of I-Caus are strongly modulated by the linguistic context of sentences with I-Caus bias verbs (Bott & Solstad, 2021; Hoek et al., 2021a, 2021b; Kehler & Rohde, 2019; Solstad & Bott, 2013). Studies by Hoek et al. (2021b), Hoek et al. (2021a), and Kehler and Rohde (2019) investigated the production biases as well as the online interpretation of I-Caus sentences with and without relative clauses which can serve as explanations, such as scold the student who was late or fire the guy who embezzled money. Their studies showed that the presence of a relative clause altered the coherence bias with immediate online processing consequences and furthermore resulted in a shift in bias. The studies by Bott and Solstad (2021) and Solstad and Bott (2013) used adverbiaal modification to show a similar effect. Modifying an I-Caus verb like annoy with a causal by phrase as in by making lots of noise altered the coherence preferences as well as the coreference bias. Furthermore, it was shown that the shifts in coreference bias were fully predictable from a more fine-grained analysis of explanation types employed in their Empty Slot Theory of I-Caus (see Section 1.4.3). Beyond the immediate sentence context, van den Hoven and Ferstl (2018b) showed that the larger discourse context also affects the I-Caus coreference bias.
1.2. Implicit consequentiality (I-Cons)

Implicit Consequentiality refers to yet another coreference bias observed for interpersonal verbs after connectives signalling a consequence relation by means of, for instance, and so or therefore. As noted by Commandeur (2010), I-Cons biases were first reported in Au (1986), who carried out a sentence completion experiment establishing strong biases towards the experiencer for psych verbs and a strong object bias for agent-evocator and agent-patient verbs. Subsequently, Stewart et al. (1998) carried out a language production and a reading time experiment establishing comparably strong coreference biases in sentences realising a consequence relation. Crinean and Garnham (2006) analysed the I-Caus and I-Cons biases from Stewart et al. (1998) as a function of verb class. Their study revealed that psych verbs of the stimulus-experiencer and the experiencer-stimulus types displayed I-Cons biases to the experiencer argument, exactly the mirror image of their respective I-Caus biases. Their analysis further revealed that the tested agent-evocator and agent-patient verbs showed class-internally consistent I-Cons biases towards the object argument. They hypothesised that this pattern can be accounted for by the argument structure of the verbs: they proposed that experiencer and patient arguments are associated with the consequence/resultant state of the eventuality expressed by the verb, which explains why they are primarily referred to in discourse continuations describing a consequence. I-Cons biases prove to be similarly strong and robust as I-Caus biases. Hartshorne et al. (2015) collected both I-Caus and I-Cons biases for a comprehensive sample of 502 verbs from seven verb classes from VerbNet (Schuler, 2005). The results revealed large variation of 502 verbs from seven verb classes from VerbNet Caus and I-Cons biases for a comprehensive sample I-Caus biases. Hartshorne et al. (2015) collected both I-Cons biases prove to be similarly strong and robust as to in discourse continuations describing a consequence. Thus, Kehler et al. (2008, p. 34) found IC-bias verbs to trigger consequences/result relations in 22% of the cases for NP1-biased and in 15% of the cases for NP2-biased verbs (compared to 22% for “non-IC bias” verbs). From this, one may argue that coreference and coherence biases should be considered separately from each other.

Finally, a further study worth mentioning for the discourse structuring potential of causality and consequentiality, Simner and Pickering (2005), investigated the planning of causes and consequences in text production. One of their findings was that when the prompt described the cause of an event, continuations to the prompt tended to describe its consequences. This further shows that the consequences of eventualities play a crucial role in language production.

1.3. Online evidence for I-Caus and I-Cons

I-Caus has played an important role in discussions on the time course of discourse biases in pronoun interpretation. A number of online comprehension studies have addressed the question of when I-Caus information is taken into account during online comprehension. Especially earlier studies – employing different methodologies such as probe tasks or phrase-by-phrase self-paced reading – presented mixed results. While some of these earlier studies provided evidence for anticipation (“focusing”) of the I-Caus bias referent (McKoon et al., 1993), others failed to find early effects and assumed I-Caus effects to result from late integration of complete discourse units (Garnham et al., 1996; McDonald & MacWhinney, 1995; Stewart et al., 2000). Despite the failure of the latter studies to find immediate coreference effects of I-Caus at the pronoun, later studies provided I-Caus congruency effects at or briefly after the pronoun in self-paced reading and eyetracking during reading (Featherstone & Sturt, 2010; Koornneef et al., 2016; Koornneef & Sanders, 2013; Koornneef & van Berkum, 2006) and immediate looks to the biased referent in eyetracking in the visual-world paradigm (Cozijn et al., 2011; Garnham et al., 2020; Itzhak & Baum, 2015; Järvi kivi et al., 2017; Pyykkönen & Järvi kivi, 2010; van den Hoven & Ferstl, 2018b). Taken together, there is firm evidence that I-Caus leads to immediate effects right at the pronoun. As Garnham et al. (2020) point out, the evidence as to whether the biased referent is in fact anticipated is not so clear. The study by Pyykkönen and Järvi kivi (2010) reported looks to the biased referent even before the pronoun, but this finding was not replicated in other visual-world eye-tracking studies.
There is also experimental evidence suggesting that the coherence bias is used rather immediately during real-time comprehension. An implicit learning paradigm was used in Rohde and Horton (2014) to show that I-Caus sentences lead comprehenders to expect an explanation, which was different from sentences with transfer of possession verbs, which triggered expectations about an occasion/continuation relation. In line with I-Caus leading to the anticipation of an explanation, the eye-tracking during a reading study by Hoek et al. (2021a) revealed an I-Caus coherence bias effect on reading times of a connective preceding the pronoun. They tested I-Caus NP2 biased sentences modified by an explanatory relative clause (Diane fired the guy who was embezzling money) or a neutral relative clause (Diane fired the guy from London who was here last month) crossed with type of connective, comparing because and and so. The crucial finding was an interaction between relative clause type and connective with a stronger regression-path duration and total time difference between neutral relative clauses than between causal relative clauses. Thus, the coherence bias is immediately affected by a “preemption” of the I-Caus bias due to a causal relative clause. Moreover, the authors report generally shorter regression-path durations and total times for the connective because than for and so. Although the lexical difference between connectives makes this effect hard to interpret, it is consistent with the coherence bias observed for I-Caus verbs by Bott and Solstad (2021) and Kehler et al. (2008).

The only published study on the online effects of I-Cons that we are aware of is a recent visual-world eye-tracking study by Garnham et al. (2020), in which online processing evidence was provided for an early influence of I-Cons, too. The authors compared the referential interpretation of ambiguous pronouns following statements with psych verbs of the stimulus-experiencer or the experiencer-stimulus kinds. In their second experiment, they further manipulated whether the pronoun followed because or and so. The results revealed early effects right at the pronoun for both I-Cons and I-Caus conditions. The authors further discuss whether I-Caus (i.e. because) may have led to an additional very early focusing effect on the associated referent, but since this effect was not statistically reliable, it is unclear to date whether I-Caus and I-Cons differ in their relative time course.

1.4. Theoretical accounts of I-Caus and I-Cons

I-Caus and I-Cons have played unequal roles in psycholinguistic research, with the large majority of offline and, in particular, online studies investigating I-Caus. However, there have been a number of theoretical accounts addressing and contrasting both phenomena. One major difference between these accounts relates to whether the phenomena are thought to result from some linguistic, verb-semantic mechanism that goes well beyond general world knowledge or not. In the existing literature, the question “linguistics or world knowledge” is often posed as an either-or question. However, we contend that this dividing line is somewhat artificial, because at some level, all linguistic accounts must also include a role for world knowledge. For instance, world knowledge seems to be needed to account for the preferred bias-incongruent interpretation in a sentence such as Bill admired John because he is easily impressed, where he is interpreted to be coreferent with the subject experiencer Bill despite admire being an I-Caus NP2-biased verb. Thus, the bias-congruent coreference to NP2 is ruled out by world knowledge without any plausible linguistic mechanism underlying the inferred coreference. This inference (cf. Hobbs, 1979) simply comes about because the property of being impressionable is not among the properties typically causing admiration. In the following, we classify existing accounts of I-Caus and I-Cons as “pure” world knowledge accounts if they assume linguistic factors, and in particular verb semantics, to be no major determinant of I-Caus nor I-Cons. These are set apart from accounts that assume linguistic factors to play a more critical role in determining biases. Among these latter accounts, we differentiate between One-Mechanism Accounts which assume the same factor to underlie I-Caus and I-Cons, and our own Two-Mechanism Account that assumes that a linguistic mechanism rooted in verb semantics is underlying I-Caus, while at the same time a more general pragmatic discourse mechanism is responsible for I-Cons.

1.4.1. World knowledge-based accounts

For I-Caus, it has been argued that the coreference bias must be inferred from general knowledge about the typical causes and effects of the described events without a primary role for verb semantics (Corrigan, 1988, 2001; Kehler & Rohde, 2019; Pickering & Majid, 2007; Semin & Fiedler, 1988, 1991; van den Hoven & Ferstl, 2018a). We will focus on the discussion in Pickering and Majid (2007) because the authors address both I-Caus and I-Cons in their paper. Pickering and Majid explicitly deny that implicit causes and consequences are determined by (verb) meaning, contending that “the implicit causality bias constitutes the likely reasons given for an event” (Pickering & Majid, 2007, p. 784). Their discussion centres around a single verb, question,
which is taken to be representative of I-Caus and I-Cons. According to Pickering and Majid (2007), the lexically
given cause of question in John questioned Mary would be John asked Mary a question, and the lexically given
consequence Mary was asked a question. It may be noted that question doesn’t seem to fit into the category
of causative predicates, but we won’t go further into this issue here. More importantly, Pickering and Majid argue
that continuations to I-Caus and I-Cons prompts, explicit
causes and consequences, never express these tautological propositions (e.g. John questioned Mary because
he asked her a question). Instead, they contend, participants make reference to other inferred causes and con-
sequences – beyond lexical semantics. Therefore, they argue, verb semantics plays no immediate role in I-
Caus and I-Cons, although it may influence the inferences drawn. Instead, both phenomena must be
explained in terms of world knowledge, that is, “people’s grounds for attributing reasons to events and
determining the likely consequences of events” (Pickering & Majid, 2007, p. 786). These reasons and con-
sequences are considered to be inferences drawn from the events as described by the verbs, their arguments
and the constructions they occur in. As pointed out by one reviewer, one may question how useful the distinc-
tion between verbal event semantics and the event descriptions alluded to by Pickering and Majid really is.
However, as we already mentioned, the authors explicitly deny that the reasons provided have any semantic
representation in the verb.

A straightforward prediction of Pickering and Majid’s claim is that neither I-Caus nor I-Cons continuations for
any relevant verb should specify or further elaborate on an eventuality that is part of the verb’s lexical seman-
tics. As far as we can tell, however, the theory does not make any predictions as to differences between I-Caus
and I-Cons coreference and coherence biases within and across verb-classes since they are both inferences from
event descriptions “presumably determined by a similar range of factors” (Pickering & Majid, 2007, p. 786).

1.4.2. One-Mechanism Accounts

Among the approaches anchoring the bias in verb semantics, an alternative view to the just outlined
account is that both phenomena can be rooted in verb semantics. This is what Hartshorne et al. (2015) refer to
as the Semantic Structure Account, which has a long-
standing history in psycholinguistic work on I-Caus
(Au, 1986; Brown & Fish, 1983; Crinean & Garnham,
2006; Garvey & Caramazza, 1974; Hartshorne, 2014; Hart-
shorne et al., 2015; Hartshorne & Snedeker, 2013; Rudolph & Försterling, 1997; Stevenson et al., 2000).
Researchers working in this tradition have mostly tried
to reduce the implicit causes and consequences of
events to aspects of the lexical semantics of verbs.

Crinean and Garnham (2006) propose that I-Caus and
I-Cons can be explained by relating causes and conse-
quences to the thematic roles from a verb’s argument
structure. Take, for instance, fascinate. As a causative
psych verb it assigns two thematic roles, the stimulus
role realised as the subject argument and the experien-
cer as the object. The eventuality expressed by the verb
can be decomposed as follows. The stimulus causes the
experiencer to become fascinated. This causal event
thus has a cause associated with the stimulus and a con-
sequence associated with the experiencer. The lexical
semantics of experiencer-stimulus predicates such as
adore can be exploited in a similar way to predict the
same mirror image of I-Caus and I-Cons: the stimulus,
in this case the object, causes the experiencer, now
the subject, to be in a state of admiration (note,
however, that the two classes differ in their status as ca-
sative predicates, where only stimulus-experiencer verbs
are considered to be causative in a proper sense; see, for
instance, Landau, 2010; Pesetsky, 1995). Agent-patient
verbs denote events in which an agent intentionally
acts in a certain way that affects the patient argument.
Therefore, I-Caus should be associated with the agent,
that is, the subject, whereas I-Cons should be associated
with the object according to Crinean and Garnham
(2006). And this is exactly what could be observed for
the agent-patient verbs in the study by Stewart et al.
(1998). The last category from the revised action-state
taxonomy by Rudolph and Försterling (1997) are
agent-evocator verbs like thank. Here, the patient
argument is not only the affected entity in the event, it is also
involved in some other eventuality providing a cause for
the agent to develop the intention to act (characterised as
a “stimulus” in Crinean & Garnham, 2006). Thus, the
evocator argument, that is, the object plays dual
“roles” as a patient and stimulus at the same time – it is
both associated with the cause and the consequence of
the eventuality expressed by the verb (Crinean &
Garnham, 2006). Empirical support for this approach
comes from comprehensive norming studies reported
by Ferstl et al. (2011) for I-Caus and by Garnham et al.
(2020) for I-Cons – with the exception that for I-Caus,
Ferstl et al. (2011) found agent-patient verbs to display
a balanced bias, contrary to the finding in Stewart
et al. (1998) (more on this below).

To summarise the One-Mechanism Account, the
uniform mechanism underlying both I-Caus and I-Cons
is a causal analysis of the verb classes considered in
I-Caus research. Since a contingency relation involves
both a cause and an effect, it is strongly expected that
these verbs display both I-Caus as well as I-Cons biases.
and that these biases take the form they do – with the important exception of agent-patient predicates. Causes and consequences are just two sides of the same coin, a causal relation, and can thus be expected to affect discourse production and comprehension in similar ways.

Hartshorne et al. (2015) propose another semantic structure account of I-Caus and I-Cons. Instead of employing the revised action-state taxonomy, I-Caus and I-Cons biases of verbs are shown to be related to the semantics of the verb classes in Levin (1993). Hartshorne and colleagues exhaustively included seven VerbNet classes in their study (Kipper et al., 2008; Schuler, 2005). The authors discussed the following approximation to I-Caus and I-Cons biases based on verb semantics. The verb admire of class 31.2, for instance, inherits its semantics EMOTIONAL_STATE(E, EMOTION, EXPERIENCER) and IN_REACTION_TO(E, STIMULUS) from the class. For admire, the EMOTION argument in the predicate EMOTIONAL_STATE is set to the more specific emotional state of ADMIRATION. As in Crinean and Garnham (2006), I-Caus and I-Cons can thus be read off of the verb’s argument structure representation. IN_REACTION_TO relates an effect to its cause, and effects and causes for their part are associated with the stimulus and the expericer argument, respectively. Verbs in VerbNet classes 31.1, 31.2 and 31.3 differ in their syntactic subcategorisation frames, but they all share this underlying semantic representation. For other verb classes in the sample such as VerbNet class 33 (judgement verbs like congratulate or criticise) the link to causality is not that obvious, as these verbs are primarily interpreted as speech act verbs involving the relation DECLARE. As the authors admit, their account is not fully worked out, but their conclusion “that implicit causality and consequentivity biases are a systematic function of Levin verb class” (p. 726) seems to suggest that both phenomena should be driven by the same mechanism. With respect to discourse expectations at the point when a pronoun is encountered in NP1 verb’s NP2 + connective + pronoun configurations, they centre their discussion on I-Caus and do not consider I-Cons at all. Thus, it is hard to tell what Hartshorne et al.’s real stance is on I-Cons. However, we interpret their proposal to imply that for psych verbs, I-Caus bias is driven by causal specifications of the stimulus argument (what properties of the stimulus caused the emotional state?), whereas I-Cons results from specifications of the emotional state encoded in the verb’s lexical entry (what emotional state resulted from the stimulus?).

Summing up, the different One-Mechanism Accounts share the cause-effect duality view of I-Caus and I-Cons: Since causes and consequences are ultimately parts of relational notions, it’s expected that they’re interrelated also in terms of bias. While these accounts capture psych verbs with their common semantic base and mirror biases well, they fare less well with the two other main verb classes, agent-patient and agent-evocator predicates. On the one hand, they aren’t uniformly causative, and on the other, not all causative agent-patient predicates display a mirror bias for I-Caus and I-Cons.

1.4.3. Two-Mechanism Account
The Two-Mechanism Account we present here shares with the One-Mechanism Account the idea that I-Caus is strongly determined by semantic properties of the verbs. We contend that it follows from a strategy to specify lexically determined explanatory slots provided by the verb, following the Empty Slot Theory in Bott and Solstad (2014) and Bott and Solstad (2021). This allows us to account for both coherence and coreference biases, we argue. For I-Cons, however, the Two-Mechanism Account takes a wholly different approach, assuming that discourse structural principles determine most strongly what attributions are made. More specifically, it assumes that participants follow the Contiguity Principle, a strategy that can be coined as “continue from the end-point of the previous eventuality” when providing continuations of consequence type (following general discourse assumptions in Kehler, 2002, 2004; Stevenson et al., 1994).

1.4.3.1 Mechanism 1: Empty slots in I-Caus. The mechanism made responsible for I-Caus in the Empty Slot Theory was first presented in Solstad and Bott (2013), in more detail in Bott and Solstad (2014), and developed further with additional empirical evidence in Bott and Solstad (2021). We present its main features in some detail here since it not only represents an alternative to the One-Mechanism Accounts above. It is also the only lexically based theory to our knowledge that sets out to incorporate both coherence and coherence bias properties for I-Caus as described above.2

The Empty Slot Theory for I-Caus is motivated by the observation that the bias distribution across verb classes assumed for instance by Crinean and Garnham (2006) or Hartshorne and Snedeker (2013) does not explain why the bias comes about in the first place, but merely states a correlation between bias and certain types of verbal arguments. Also, the more fine-grained classification in Hartshorne and Snedeker (2013) does not achieve the correct generalisation for a verb’s semantic influence of the bias with regard to causative predicates, which don’t show an overall NP1 bias. In other words, the Empty Slot Theory sets out to explain why it is that...
stimulus and evocator arguments are biased arguments, whereas others, such as agents and patients, are not. At the same time, it accounts for the coherence bias associated with I-Caus.

Using standard categories from semantic-pragmatic research, the Empty Slot Theory ultimately reduces the coreference and coherence preferences to the fact that the stimulus arguments of psych verbs and the evocator argument of agent-evocator verbs are associated with an underspecified proposition which can be seen as an implicit question evoked by the predicate. The source of I-Caus is thus attributed to lexical semantic properties of I-Caus verbs, accounting for the cross-linguistic robustness of I-Caus effects (see Section 1.1). The coherence and coreference biases then follow from a mechanism that evokes the specification of lexically given underspecified content (in order to avoid accommodation on the part of the hearer; see Altmann & Steedman, 1988). Thus, when participants provide an explanation about the stimulus argument Peter in Peter annoyed Mary because ..., they follow a strategy of filling a slot that is given by annoy. This slot is causally related to the main eventuality, which is why the specification will tend to constitute an explanation even in the absence of a causal connective such as because. A parallel strategy is followed for experiencer-stimulus predicates, for which the stimulus is again a cause for the psychological state in the experiencer (even though the predicate is not a causative predicate in the proper sense, see for instance the analysis of psych verbs in Asher & Lascarides, 2003). For evocator arguments, the underspecified proposition derives from the presupposition that is introduced by agent-evocator predicates such as congratulate, criticise or apologise to (Fillmore, 1969; McCawley, 1975): It is a precondition on the proper use of these verbs that there exists a preceding eventuality upon which the agent acts. Mostly, the presupposition is associated with the object argument, which contributes to NP2 bias, but for some verbs, such as apologise to, the presupposition resides in the agent itself, which is why such predicates are NP1-biased. Consequently, this is a very different take from the World Knowledge-Based Account proposed by, for instance, Pickering and Majid (2007). The empty slots correspond to semantic entities that are part of the verb’s semantics. Hence, a major influencing factor of I-Caus is verb-based. However, we obviously do not want to claim that slots are the only factor driving I-Caus. After all, I-Caus is not deterministic, but a bias phenomenon. What is more, I-Caus bias varies within verb classes. We cannot go into much detail here, but for a number of stimulus-experiencer predicates such as disturb, it may be noted that they also allow for a purposeful, agent-patient reading. Since participants may choose either reading, and the latter does not introduce a slot, the bias should be influenced by the proportion of agent-patient interpretations for a verb. Conversely, less variation and a stronger bias is expected for verbs that are not ambiguous in this respect, such as fascinate. Furthermore, a number of other linguistic factors in addition to world knowledge can also exert an influence on the bias (see, in particular, van den Hoven & Ferstl, 2017). Moreover, speakers can leave underspecified content unspecified or force the hearer to accommodate a presupposition. The crucial aspect about the Empty Slot theory is that it claims that verb semantics is the major determinant of I-Caus.

Ultimately, we consider the slots in the Empty Slot Theory triggers of discourse expectations (Bott & Solstad, 2021; Kehler et al., 2008). It should be noted that these slots do not correspond to one particular linguistic or semantic category, but may have a number of sources. Going beyond I-Caus, expressions associated with Questions under Discussion, for instance (Clifton & Frazier, 2012; Kehler & Rohde, 2017; Onea, 2016; van Kuppevelt, 1995) or concealed questions (Aloni & Roelofsen, 2011) and other types of complement coercion (Asher, 2011; Pustejovsky, 1995) may be analysed as triggers of discourse expectations. In addition, presuppositions that cannot be verified in the preceding context may also lead to “slot-filling strategies” at the discourse level. However, such discourse expectations may not be associated with coreference biases, as these phenomena don’t necessarily involve interpersonal predicates.

Bott and Solstad (2014) and Bott and Solstad (2021) provided evidence for the Empty Slot Theory. First, Bott and Solstad (2014) and Bott and Solstad (2021) found evidence supporting the coherence bias observed by Kehler et al. (2008). Second, Bott and Solstad showed that the preferred explanations follow particular patterns, constituting uniform types of explanations for bias-congruent continuations (see below). Furthermore, Bott and Solstad (2021) showed that the bias is driven by underspecification. Once the preferred explanation is already specified in the prompt, such as Peter annoyed Mary by singing loudly (because) ..., the proportion of explanations after a full stop drops to the level of non-biased agent-patient verbs and the pattern of explanations shifts away from the types preferred for bias-congruent explanations towards the explanations provided in bias-incongruent continuations. The effects for relative clauses discussed above (Hoek et al., 2021a, 2021b; Kehler & Rohde, 2019) are also explained by this approach: The relative clauses leading to shifts in coreference and coherence bias provide explanations that fill the lexically given slot.
What is more, the theory makes predictions even more fine-grained than at the level of discourse relations. The slots, as stated above, are an integral part of the verb’s semantic representation. Thus, the underspecified proposition for psych verbs is identified with the causing entity in the causal relation provided by the verb. This predicts that explanations should constitute such simple causes (Bott & Solstad, 2014, 2021). For agent-evocator verbs, the presupposition causes the intention of the agent to act (e.g. performing an act of congratulating or criticising) and explanations thus constitute what may be characterised as an external reason (it is anchored outside the agent’s mind). Having shown that these are indeed the dominating categories for explanations in Bott and Solstad (2014) and Bott and Solstad (2021) provided evidence that bias-incongruent explanations display a shift in explanatory categories away from simple causes and external reasons, respectively.

1.4.3.2 Mechanism 2: General discourse strategies for I-Cons. As we just saw, I-Caus can be said to follow from a “slot filling” strategy based on the fact that the sequence Name1 verb-ed Name2 because … leaves the particular explanation, or cause, unspecified for the eventuality introduced by the predicate and its arguments. If the Empty Slot Theory were to offer a unified account of I-Caus and I-Cons also involving only one mechanism, one would need to identify slots for consequences that would be specified by continuations after I-Cons prompts (e.g. and so …). For psych verbs, for instance, these would be slots specifying the psychological effect, that is, the psychological state of the experiencer (as specified in the approach by Hartshorne et al., 2015, cf. Section 1.4.2). However, we contend that for the predicates from the classes with a prominent I-Caus and/or I-Cons bias, no relevant slots are available that would be associated with a consequence of the predicates in question. Rather, it would seem that agent-evocator verbs do not involve an end-state in any proper sense and for psych verbs their psychological (end) state is already specified in the verb’s root. Thus, for the I-Cons NP2-biased verb annoy, the verb’s root specifies the state of annoyance that the experiencer is in (and not the actions or properties of the stimulus). Similarly, for the I-Cons NP1-biased verb admire, the root also specifies the psychological state that the experiencer is in (and again not a property of the stimulus; see also the above discussion of the VerbNet analysis of Hartshorne et al., 2015). From these considerations, it follows that the I-Cons bias merely correlates with certain thematic roles, but doesn’t result from a strategy determined by these roles. In this particular respect, our considerations are in accordance with the theoretical approach in Pickering and Majid (2007).

What mechanism could account for the observed I-Cons bias patterns if verb thematic roles are not at the heart of I-Cons bias? We propose the Contiguity Principle, a discourse structural constraint akin to the inference principle assumed for contiguity relations by Kehler (2002) and Kehler (2004), according to which we interpret “the final state of one eventuality as being the initial state of the next” (Kehler, 2004, p. 259). While consequences belong to the class of cause-effect relations in the Hobbs classification of Kehler (2002) and Kehler (2004), they share the property with contiguity relations of an ordering between the eventualities introduced by different discourse segments. Thus, since there is no underspecified end-state slot to be specified by a consequence clause (as introduced by and so), we are left with a principle of “continuing from the end state of the previous eventuality”, i.e. the one introduced in the prompt. As a general principle of discourse structuring, the Contiguity Principle is also at play with biases that are semantically unrelated to I-Caus and I-Cons such as the goal biases observed by Stevenson et al. (1994), according to which the goal argument in transfer of possession events displays a stronger bias than the source argument in continuations prompted by the consequence connective so.

This second mechanism makes several predictions for I-Cons bias. First, it predicts that in general, I-Cons bias will go in the direction of the object, NP2 argument, as most eventualities involving (at least) two participants describe actions at the end point of which the (direct) object is affected. From this perspective, the I-Cons NP1-bias of experiencer-stimulus verbs constitutes an exception in terms of bias pattern but still fits with the Contiguity Principle, as for these predicates the argument associated with the final (and only) state of the eventuality is the subject argument. The Contiguity Principle in terms of coreference bias is supported by the results from large-scale bias studies as in Ferstl et al. (2011) and Garnham et al. (2020). Thus, for I-Caus, Ferstl et al. (2011) found the bias to be evenly distributed across verb classes and an evenly distributed I-Caus bias for the “slot-less” agent-patient verbs. Contrary to this, for I-Cons bias, Garnham et al. (2020) observe an overall object bias for the verb classes under investigation (which were almost identical to the ones investigated in the 2011 paper). Garnham et al. (2020) only find an I-Cons NP1-bias for experiencer-stimulus predicates, whereas agent-patient verbs (as in all other classes) display a pronounced I-Cons NP2 bias.

Second, the contiguity approach to consequences makes predictions also for the types of consequences
that occur. Thus, it predicts for psych verbs that consequences do not describe the (end) state of the experiencer, but rather describe some consequence subsequent to the eventuality introduced by the predicate as a whole. Put differently, consequences describe eventualities with no temporal overlap with the eventuality in the prompt. Thus, if the consequence describes a mental state of the experiencer, this state should be subsequent to the experiencer’s state as specified by the predicate. Experiment 4 will show in detail how this works for stimulus-experiencer and experiencer-stimulus predicates.

Figure 1 illustrates the Two-Mechanism Account for psych verbs (here shown for the stimulus-experiencer predicate *annoy*). On the one hand, the explanation targets the causal slot included in the predicate (“Peter does something”) and thus also temporally overlaps with it. On the other hand, the consequence derives from the Contiguity Principle and is thus subsequent and temporally disjoint from the eventuality introduced by the predicate as a whole.

In combination, the two mechanisms also make predictions for the distribution of explanations and consequences in “full stop” conditions, that is, where no connective is offered. It may be assumed that the slot filling strategy, which is tied to specific properties of the preceding context, that is, the prompt, takes precedence over a more general discourse-structural mechanism such as the Contiguity Principle (see the inference mechanisms discussed by Asher & Lascarides, 2003). Therefore, we expect all “slot predicates” to trigger explanations above consequences. This is exactly what Kehler et al. (2008) found for NP1- and NP2-biased verbs and also what Bott and Solstad (2014) and Bott and Solstad (2021) found for the three individual classes of stimulus-experiencer, experiencer-stimulus and agent-evocator verbs. For predicates without a slot, such as agent-patient verbs, however, explanations and consequences should be more evenly distributed, as there is no “disadvantage” for consequences evoked by explanatory slots (explanations may still be more frequent due to other principles, such as the causality by default principle, cf. Sanders, 2005). This is also what Kehler et al. (2008) found for bias-neutral (“non-IC-bias” predicates in their terms) verbs. Experiments 2 and 3 will provide additional evidence in this regard for stimulus-experiencer and experiencer-stimulus predicates.

1.4.4. The present investigation: asymmetry or symmetry?

The just outlined I-Caus and I-Cons theories make rather different predictions for discourse production. These differences become most obvious when looking beyond coreference production. In our experiments, we will, therefore, not only consider coreference biases but also distributions of coherence relations and even a more fine-grained causal, or contingency, typology distinguishing between verb-inherent and verb-external causes and consequences.

The symmetric relation between I-Caus and I-Cons coreference biases is best supported in the case of psych verbs, that is, stimulus-experiencer and experiencer-stimulus verbs. The present series of experiments thus focuses on these verb classes to see up to which point the symmetry between the two phenomena extends. While the One-Mechanism Accounts predict symmetrical behaviour of I-Caus and I-Cons with respect to discourse coherence, the World Knowledge-Based Accounts do not allow us to derive an a priori prediction of asymmetry for I-Caus and I-Cons in this respect. From the Two-Mechanism Account, however, which says that the biases observed for I-Caus and I-Cons have very different origins, we can derive the Asymmetry Hypothesis:

**Asymmetry Hypothesis:** Even for psych verbs of the stimulus-experiencer and the experiencer-stimulus kind that display mirror-like coreference biases for I-Caus and I-Cons, discourse coherence is strongly biased towards (particular) explanation relations. These explanations specify a causally relevant property of the stimulus argument and constitute the default expectation/strategy for discourse continuations in language production.

According to the Asymmetry Hypothesis, the biases observed for I-Caus and I-Cons have very different origins. I-Caus is triggered by underspecified semantics associated with a verb’s arguments, and I-Cons follows overall discourse-structural principles. The hypothesis was tested in four discourse continuation experiments. Experiment 1 investigates the coreference biases for I-Caus (*because*) and I-Cons (*and so*) for stimulus-experiencer and experiencer-stimulus verbs, establishing the supposed mirror-like coreference biases. Based on these results, Experiment 2 presents a first comprehensive and systematic investigation of the coherence bias for these verbs, that is, the preference for particular
discourse relations, showing that explanations are indeed the default relation. Experiment 3 further corroborates this finding by showing that even under experimentally controlled circumstances where participants are forced to produce a continuation about the I-Cons consistent referent, explanations are at least as likely as consequences. Finally, Experiment 4 provides direct evidence that the explanations and consequences for I-Caus and I-Cons are of the particular types predicted by the Two-Mechanism Account (causal “empty slot” specifications and subsequent “contiguity” eventualities, respectively).

2. Experiment 1: coreference biases

The first experiment established I-Caus as well as I-Cons coreference biases for the psych verbs in the present study. The selected verbs were clear instances of stimulus-experiencer (henceforth, STIM-EXP) and experiencer-stimulus (henceforth, EXP-STIM) verbs, most of them taken from verb-class classifications in previous research (Bott & Solstad, 2014). This by itself sets the present research apart from related investigations by Kehler, Rohde and colleagues whose I-Caus verbs consisted of a mix of various verb-classes. For both I-Caus and I-Cons, all three types of accounts mentioned above assume clear biases for STIM-EXP and EXP-STIM verbs with I-Caus and I-Cons biases forming mirror images. Beyond the direction of the bias, we were also interested in the strength of the verbs’ I-Caus and I-Cons biases.

The present experiment is the first comparative investigation of I-Caus and I-Cons biases within the same experiment using an ordinary sentence completion task. Hartshorne et al. (2015) also compared I-Caus and I-Cons within the same experiment but used a referent choice task with pseudo-verbs in the continuation, e.g. Sally verbed Lisa because she daxed. However, comparing the I-Caus biases with those observed in Hartshorne and Snedeker (2013), they observed a shift toward the object. They hypothesised that this shift could have resulted from the slightly different aspeantial properties of the stimuli. Hartshorne and Snedeker (2013) used pseudo-nouns in a predicative stative sentence frame (because she is a dax), which may induce a bias of its own related to event interpretation (cf. Dowty, 1979, on eventive vs. stative predicates). For the specification of the empty slot introduced by the stimulus argument of a psychological I-Caus verb, it makes quite a difference whether the continuation makes reference to an abstract property (is a dax) or an abstract event (daxed), see the semantic representations in Bott and Solstad (2014) and Bott and Solstad (2021). A similar point could be made for I-Cons as will become clear when scrutinising on the types of consequences in Experiment 4. As a consequence, it is not clear whether the method used in Hartshorne et al. (2015) by itself induced a bias complicating the comparison between I-Caus and I-Cons coreference biases. We, therefore, consider it crucial to compare their results to I-Caus and I-Cons biases elicited from free production in an ordinary story continuation paradigm.

2.1. Design

The sentence continuation experiment employed a $2 \times 2( \times 2)$ within-participants and within-items design manipulating the factors VERB CLASS (German STIM-EXP vs. EXP-STIM verbs; corresponding to VerbNet class 31.1 and VerbNet class 31.2, respectively), CONNECTIVE (well “because” explanation vs. sodass “and so” consequence prompts), and GENDER ORDER (NP1fem – NP2masc vs. NP1masc – NP2fem). The latter factor was included in the design as a counterbalancing factor. The dependent variable was subject vs. object coreference of the first anaphor in the elicited sentence continuations.

2.2. Methods

2.2.1. Participants

52 students from the University of Tübingen (39 female, 13 male; mean age 23.9 years, range 18–41 years) gave their informed consent to take part in the experiment for monetary compensation (5 Euro for half an hour). All participants reported German to be their native language.

2.2.2. Materials

The items were constructed according to a name verb-ed name, connective scheme according to the above-mentioned design using 20 German STIM-EXP verbs and 20 EXP-STIM verbs (see Appendix) in combination with 40 unambiguously female and 40 unambiguously male German forenames (the complete set of materials for all experiments in this study are provided in the Open Science Framework repository for this paper). Verbs were paired in items by matching them semantically as closely as possible. The verb pairs ranged from minimal or closely related pairs such as fear and frighten or admire and fascinate to pairs that only shared the same emotional valency like despise and shock. A Latin Square was used to distribute the resulting 20 items in eight prompt conditions on four lists. This was done in such a way that each item appeared twice in each list, once with a STIM-EXP verb and one of the connectives, and a second time with the respective EXP-STIM verb and the other connective, yielding a total of 40
experimental trials in each list. Here is a sample item translated to English with a sample mapping to lists in parentheses:

(5) a. Johanna shocked Lars because … (list 1)
    b. Lars shocked Johanna because … (list 2)
    c. Johanna despised Lars because … (list 3)
    d. Lars despised Johanna because … (list 4)
    e. Johanna shocked Lars, and so … (list 4)
    f. Lars shocked Johanna, and so … (list 3)
    g. Johanna despised Lars, and so … (list 2)
    h. Lars despised Johanna, and so … (list 1)

The 40 trials of Experiment 2 with full stop continuations and 40 trials with connective plus pronoun prompts served as filler trials. All pronouns in the fillers unambiguously referred to antecedents disambiguated by gender. Half of these pronouns referred back to the subject and the other half to the object of the filler sentences. The fillers also contained transitive interpersonal verbs with two proper name arguments.

2.2.3. Procedure

The experiment was conducted via the internet employing the free Onexp software (version 1.3.1, see http://onexp.textstrukturen.uni-goettingen.de/). After reading written instructions, participants proceeded to a short practice of three trials, upon which they received the experiment with 120 individually randomised trials in a single block. In each trial a text field with the prompt appeared at the top of the browser with a sentence fragment ending in “…”. Participants were asked to type the first continuation that came to their mind in another text field just below the prompt. There was no time limit for providing an answer. Only participants that completed the experiment were included in the analysis. Each list was randomly assigned to 13 participants. A typical experimental session took about 30 minutes (median time spent on task 31 minutes, range 18–53 minutes).

2.2.4. Data annotation

The resulting data set of 2080 continuations was annotated according to the following annotation scheme. First of all, it was coded whether the continuation was complete and sensible (excluding 46 cases (= 1.2%) from the further analysis). It was then coded whether the continuation contained at least one anaphor to NP1 or NP2 (excluding another 117 cases (= 5.6%) of OVS cases). Thus, a total of 1849 continuations were included in the annotation. These were coded with respect to the following five categories:

(1) the coreference of the first anaphor to the subject or the object, respectively;
(2) anaphoric form (personal pronoun, repeated name, demonstrative pronoun, other form);
(3) its position in the continuation for surprisal analysis (anaphor immediately following the connective or not);
(4) whether there was another anaphor coreferent with the other referent;
(5) whether the continuation was itself a complex sentence including the subordination of another sentence.

The annotations were performed by two student assistants trained on a random subset of 200 continuations. Another random subset of 300 continuations were then annotated by both of them to make sure that inter-annotator reliability was high enough (Cohen’s $\kappa = 0.82$). As for all experiments reported in this paper, an agreement was determined before the exclusion of any data.

In the course of the annotation of the data, it became clear that the presumed STIM-EXP verb gefallen (“appeal to”) was exceptional in that it rather often received an OVS reading of the prompt (14 out of 50 sensible continuations). After annotation, we, therefore, excluded this verb from the statistical analysis in this and all other experiments reported in this study.

2.2.5. Statistical analysis

All remaining 1813 continuations were statistically analysed by fitting mixed-effects binomial logistic regression models with the lme4 package (Bates et al., 2015) in R (version 3.6.3). The dependent variable was congruency with the presumed bias. This allowed us to directly compare bias strength across connectives and subject vs. object coreference. The presumed I-Caus bias for STIM-EXP verbs is coreference with the subject, whereas for EXP-STIM verbs it is coreference with the object. Exactly the reverse I-Cons coreference biases are expected for prompts with a consequence connective. All factors in the models were centred. Our initial models always included the maximal random effects structure with random intercepts and random slopes by participants and items (Barr et al., 2013). However, due to failed convergence of any models including random slopes, the models reported below only include random intercepts for both participants and
items. We started with the model including all centred main effects and interactions in the fixed-effects structure, and this model was then further simplified in line with our research questions by taking out fixed effects. We tested for the significance of fixed effects by performing likelihood ratio tests of the full model with the effect in question against the model without this effect. We estimated 95% confidence intervals for the I-Caus and I-Cons biases of individual verbs applying non-parametric bootstrapping (Efron & Tibshirani, 1986) with the bootstrapping function from R’s bootstrap package. All analyses and data are publicly available in an OSF repository (see note 3).

### 2.3. Results and discussion

The model comparison between the global analysis including the centred factors of CONNECTIVE, VERB TYPE, and GENDER ORDER as well as their interactions and a model without any fixed effects of GENDER ORDER showed that GENDER ORDER did not affect the data significantly neither in a global analysis of coreference to the subject vs. object (χ²(4) = 2.75; p = .60) nor in the analysis of coreference with the bias-congruent vs. bias-incongruent referent (χ²(4) = 7.74; p = .10). In the following, we will, therefore, present the data aggregated over both gender orders.

Table 1 summarises the descriptive statistics analysing the likelihood to refer back to the subject or the object and Figure 2 presents the coreference biases of the individual verbs in the current experiment. A table of all verbs with their individual I-Caus and I-Cons biases can be found in Appendix. The verbs formed two clear clusters: While STIM-EXP verbs exhibit a uniform I-Caus bias towards the subject and an I-Cons bias towards the object, EXP-STIM verbs were biased exactly in the opposite direction. A correlation analysis for the average I-Caus and I-Cons biases for all verbs in the analysis revealed an almost perfect negative correlation between the two biases (r = −0.94). This shows that as far as coreference is concerned, I-Caus and I-Cons biases for the two verb classes in the study indeed seem to mirror each other. However, a closer look at bias strength and congruency effects reveals slight differences between the conditions (see the inferential statistics below).

Inferential statistical analyses were computed in order to test the strength of congruency effects. We started the logit mixed-effects model analyses with the global model including the fixed effects of PRESUMED BIAS (subject vs. object bias) and CONNECTIVE and their interaction as predictors as well as random intercepts for participants and items. There was no reliable interaction PRESUMED BIAS × CONNECTIVE (χ²(1) = 1.70; p = .19), and we ended with the final model including only the two main effects, which both turned out to be significant. First, there was a significant effect of CONNECTIVE (β = 0.34, SE = 0.18, χ²(1) = 8.82, p < .01), which was due to the fact that the I-Cons bias was weaker overall than the I-Caus bias. This is a first indication that there is no perfect symmetry between I-Caus and I-Cons even for the two classes of psych verbs tested here. Second, across connectives and verb types, the object bias conditions led to a stronger coreference bias than did the subject bias conditions, as was reflected by a highly significant main effect of PRESUMED BIAS (β = −1.42, SE = 0.18, χ²(1) = 74.71, p < .001). A plausible explanation of the latter effect is a general preference for recency in anaphoric dependencies (Gernsbacher & Hargreaves, 1988).

In addition to the just reported coreference analysis, we looked into the specific anaphoric forms that were used to refer back to the subject and object referents in the four conditions, respectively. The results of this analysis are summarised in Table 2. Across both because and and so conditions, subject coreference was almost exclusively established with personal pronouns (I-Caus: 99.3%; I-Cons: 99.2%). Other forms only surfaced when referring back to the object. There was a slight difference between explanations and consequences with more forms other than personal pronouns for I-Cons (10.1%) than I-Caus (3.6%). Although both connectives are syntactically subordinating, consequences and explanations can be taken to represent different discourse structural configurations, with explanations being subordinating and consequences being coordinating discourse relations (for discussion see, in particular Asher & Vieu, 2005). The two relations thus differ with regard to the Right-Frontier Constraint (Asher & Lascarides, 2003; Polanyi, 1988), which influences anaphora resolution and hence possibly also the choice of anaphoric forms.

We conducted an exploratory data analysis analysing coreference with a personal pronoun vs. some other form in a logit mixed-effects regression analysis including...
the centred fixed effects of CONNECTIVE and of GRAMMATICAL FUNCTION (subject vs. object coreference) and their interaction. Furthermore, the models included the random intercepts of PARTICIPANTS and ITEMS. The analysis revealed no significant interaction \( \chi^2(1) = 2.38; p = .12 \) but two reliable main effects: the significant main effect of GRAMMATICAL FUNCTION \( \beta = 2.36, SE = 0.45, \chi^2(1) = 46.64, p < .001 \) was due to the fact that other forms than personal pronouns emerged almost exclusively in coreference with object referents. The effect of CONNECTIVE \( \beta = 1.26, SE = 0.29, \chi^2(1) = 21.73, p < .001 \) was also significant, showing that coreference after the connective sodass “and so” did in fact lead to more complex forms than after the connective weil “because”.

3. Experiment 2: coherence relations

Experiment 2 investigated the distribution of discourse relations and anaphoric dependencies after a full stop (cf. Bott & Solstad, 2014, 2021; Kehler et al., 2008, among others, for a similar approach). In particular, we were interested in the distribution of explanation and consequence relations, corresponding to the two connectives weil "because" and sodass "and so". If there is a slot in psych verbs that triggers particular types of explanations, this slot should be present even in the absence of a causal connective. With no corresponding slot for consequences, the Asymmetry Hypothesis predicts significantly more explanations than consequences with full stop prompts.

The studies by Bott and Solstad (2014) and Kehler et al. (2008) suggested that explanations are indeed strongly preferred over consequences for I-Caus verbs.
In both studies, verbs with a pronounced I-Caus bias were followed by explanations in \( \approx 60\% \) of all cases. Furthermore, in the study by Kehler et al. (2008) explanations were three times more frequent than consequences overall.

Investigating distributions of discourse relations associated with STIM-EXP and EXP-STIM predicates is desirable for a number of reasons. First, Kehler et al. (2008) (with three exceptions) used I-Caus verbs from the study by McKoon et al. (1993), which were only controlled for I-Caus coreference bias (NP1 vs. NP2), but neither for verb class nor for I-Cons bias. In particular, the set of I-Caus NP2-biased verbs included EXP-STIM as well as agent-evocator verbs known to differ in terms of I-Cons (Garnham et al., 2020). This makes them less suited to compare I-Caus and I-Cons biases and discourse relation behaviour.

Second, given that the two verb classes are parallel in terms of semantic roles and – as shown in Experiment 1 – in terms of their coreference bias, it is important to also compare the verb classes with respect to the number of explanations and consequences they give rise to. The Two-Mechanism Account predicts comparably high frequencies of explanations for the two verb classes, but differences may be expected with respect to the likelihood to continue with a consequence relation. This derives from the difference in the end-point properties of the two verb classes. STIM-EXP predicates are causative in a proper sense and thus involve a (telic) end-point proper. However, EXP-STIM predicates are stative, and therefore do not involve an end point (cf. Moens & Steedman, 1988).

Last but not least, Experiment 2 provides a link between the explicit explanations in the previous experiment and implicit explanation relations. Following Kehler et al. (2008) we hypothesised that, conditioned on explanations and consequences, respectively, I-Caus and I-Cons biases for the two verb classes should be comparable with those found for prompts involving weil “because” and sodass “and so”.

### 3.1. Design

The sentence continuation experiment employed a \( 2(\times)2 \) within-participants and within-items design manipulating the factors VERB CLASS (STIM-EXP vs. EXP-STIM; corresponding to VerbNet classes 31.1 and 31.2, respectively), and GENDER ORDER (NP1\(_{fem} \)–NP2\(_{masc} \) vs. NP1\(_{masc} \)–NP2\(_{fem} \)). The latter factor was included in the design as a counterbalancing factor. The continuations were annotated with respect to discourse relations and anaphoric dependencies.

### 3.2. Methods

#### 3.2.1. Participants

The experiment was run together with Experiment 1 testing the same 52 participants.

#### 3.2.2. Materials

The 20 STIM-EXP and 20 EXP-STIM items from Experiment 2 were modified by taking out the connectives and inserting a full stop. A Latin Square was used to distribute the resulting 20 items in four prompt conditions on two lists such that each item appeared twice in both lists but each verb appeared only once. The 40 trials of Experiment 1 and 40 trials comparable to those in Experiment 4 served as fillers for the current experiment.\(^6\)

#### 3.2.3. Procedure

The procedure was the same as in Experiment 1.

#### 3.2.4. Data annotation

The resulting data set of 2080 continuations was annotated according to the following annotation scheme. First, it was coded whether the continuation was complete and sensible (excluding 99 cases (= 4.4\%) from the further analysis). For the discourse relation analysis, the remaining 1981 continuations were categorised into the following discourse relations: in a first step, we annotated EXPLANATION (using a weil “because” insertion test) and CONSEQUENCE (using a sodass “and so” insertion test) relations. If the continuation did not pass either of these tests, we applied insertion tests for nachher “afterwards”, und zwar “that is”, and aber “but” to code OCCASION, ELABORATION and CONTRAST/VIOLATED EXPECTATION relations, respectively. Furthermore, the remaining un categorised continuations were tested for a PARALLEL relation (e.g. Mary loves John. (And) John loves Mary.). These six categories accounted for 95.5\% of the data. The remaining continuations including questions and ambiguous cases were merged into a category OTHER. The data were coded by two student assistants. At the beginning of their annotations, their inter-annotator agreement was checked on a randomly drawn subset of 300 continuations (Cohen’s \( \kappa = 0.80 \) for the discourse annotation). Before data analysis, the complete set of annotations was again checked by the second author.

For the coreference analysis, the same coding schema was applied as in the previous experiment. It was coded whether the continuation contained at least one anaphor to NP1 or NP2 (excluding another 239 cases (= 11.5\%)). Again, only subject-verb-object cases were included in the analysis (excluding 78 (= 3.8\%) object-verb-subject cases). The resulting 1608 continuations were included in the coreference analysis. These were
The verb *gefallen* ("appeal to") was again exceptional in that it often received an object-verb-subject reading of the prompt (27 out of 50 sensible continuations). This verb was, therefore, excluded from the statistical analysis.

### 3.2.5. Statistical analysis

For the analysis of discourse relations, GLMER analyses were computed. Two logit mixed-effects regression models were fit to the data. The first analysed explanations vs. other discourse relations and the second analysed consequences vs. other discourse relations. Models included the fixed effects of VERB CLASS, GENDER ORDER, and their interaction as well as random intercepts of participants and items. Statistical significance of fixed effects was determined by model comparisons of models only varying in the effect of interest.

The analysis of coreference was performed in the same way as in the previous experiment. The only difference was that coreference was conditioned on the discourse relation realised by the participants. A total of 1325 continuations were contingency-related (1056 explanations and 269 consequences), and only these were included in the GLMER models analysing coreference. As for the previous experiment, all analyses are publicly available (see note 3).

### 3.3. Results and discussion

The distribution of discourse relations is shown in Figure 3. Participants produced far more EXPLANATIONS than any other discourse relation – 60.2% for EXP-STIM verbs and 58.2% for STIM-EXP verbs. The second most common discourse relations were CONSEQUENCES for STIM-EXP verbs in 21.1% of the cases (vs. only 9.9% for EXP-STIM) and CONTRASTS/VIOLATED EXPECTATIONS for EXP-STIM verbs in 14.8% of all cases (STIM-EXP 5.8%). ELABORATIONS were quite rare (7.7%) and OCCASION and PARALLEL hardly occurred with less than 2% continuations overall. The Appendix contains an overview of the distribution of all these discourse relations for the individual verbs investigated in Experiment 2.

The GLMER analyses of explanations showed that there were no differences between the two GENDER ORDERS ($\chi^2(1) = 1.41, p = .49$). Follow-up logit mixed-effects model analyses were thus conducted including only the fixed effect of VERB CLASS. The analysis of the binary categorical response variable EXPLANATIONS vs. all other discourse relations revealed that the inclusion of VERB CLASS did not enhance model fit (model comparison: $\chi^2(1) = 0.22, p = .64$). In addition, the analysis revealed that the intercept significantly differed from zero with a positive parameter estimate ($\beta = .52, SE = .23, \chi^2(1) = 5.02, p < .05$). This intercept shows that explanations were provided more frequently than all other discourse relations together. Thus, explanations did in fact constitute the default discourse relation and this tendency proved to be equally strong for both classes of psych verbs.

This was different in the analysis of CONSEQUENCES vs. other relations showing that the two verb classes in fact differed in how likely a consequence relation was chosen (model comparison: $\beta = 1.01, SE = 0.24, \chi^2(1) = 13.54, p < .01$). STIM-EXP verb conditions gave rise to consequence continuations 21.4% of the time, whereas EXP-STIM verb conditions received consequence continuations only 9.9% of the time. In the latter conditions, contrast relations (14.8%) appeared even more frequently than consequence relations. In addition to the VERB CLASS effect, the intercept significantly differed from zero with a negative parameter estimate ($\beta = -2.01, SE = .15, \chi^2(1) = 58.84, p < .01$). Thus, asymmetry between relations was observed in two respects. First, explanations were far more frequent than consequences. Second, while explanations were equally likely for both verb classes, the likelihood to continue with a consequence differed between STIM-EXP and EXP-STIM verb types.

Could there be reasons other than verb semantics for explanations to be more frequent than consequences? Some researchers have suggested that explanation constitutes a default category in discourse (Sanders, 2005). So, is it possible that a One-Mechanism Account augmented by causality as default can account for the observed differences between causes and consequences? Such an account cannot fully accommodate the present findings because of the observed differences between verb classes with respect to the likelihood of consequence relations. If causality by default were the only principle responsible for a different distribution between explanations and consequences, one would expect these verb classes to display a comparable distribution of contingency relations (and other discourse relations), which is clearly not the case. What is more, the fact that consequence relations were more likely for STIM-EXP than for EXP-STIM verbs is fully expected under the discourse-structural mechanism for consequences. After all, the two types of psych verbs differ semantically with respect to natural end points. While STIM-EXP verbs involve an end-point, EXP-STIM verbs are stative. Thus, the former constitute a better trigger for the Contiguity Principle, as discussed in Section 1.
For coreference, the statistical analysis again yielded strong I-Caus and I-Cons bias effects with roughly the same pattern as in the previous experiment. Table 3 presents the results of the coreference annotation within the major three discourse relations. Continuations that were explanations showed a strong bias towards the stimulus argument, whereas consequences were associated with the experiencer argument. The discourse relation of contrast showed yet another pattern, that is a general bias to the object argument, irrespective of the verb type. Logit mixed-effects regression analyses were performed to compare the strength of congruency effects in the I-Caus and I-Cons conditions. The models included the centered effects presumed bias (NP1 vs. NP2), discourse relation (explanation vs. consequence), gender order (NP1fem–NP2masc vs. NP1masc–NP2fem), and their interactions as well as the random intercept of participants. This was the maximal model that converged. A first model comparison revealed that gender order had no effects and was, therefore, removed from all subsequent analyses ($\chi^2(4) = 2.36, p = .62$). There was also no significant interaction between bias and relation ($\chi^2(1) = 0.73, p = .39$) and the model was further simplified to a model only including the intercept and the main effects of bias and relation. All three estimates contributed significantly to model fit. The reliable intercept ($\beta = 2.32, SE=0.16, \chi^2(1) = 90.12, p<.01$) reflects the fact that all conditions were at least biased about 80% to the congruent referent. As in the previous experiment, the bias turned out to be stronger for coreference to the immediately preceding object as indicated by the significant main effect of presumed bias ($\beta = 1.42, SE=0.19, \chi^2(1) = 57.93, p<.01$). And finally, consequentiality biases were on average stronger than explanation biases as indicated by a significant main effect of discourse relation ($\beta = 0.70, SE=0.29, \chi^2(1) = 5.59, p<.05$). Note that the latter effect points in the opposite direction as in the previous experiment, where I-Caus

Table 3. Relative (%) and absolute (n) frequencies of subject vs. object coreference for STIM-EXP and EXP-STIM verbs in Exp. 2 for the three most common discourse relations explanation, consequence, and contrast.

|                | subject |     |     | object |     |     |
|----------------|---------|-----|-----|--------|-----|-----|
|                | %       | n   | %   | n      |     |     |
| Explanation    |         |     |     |        |     |     |
| STIM-EXP       | 79.2    | 401 | 20.8| 105    |     |     |
| EXP-STIM       | 7.3     | 41  | 92.7| 520    |     |     |
| Consequence    |         |     |     |        |     |     |
| STIM-EXP       | 3.2     | 6   | 96.8| 184    |     |     |
| EXP-STIM       | 84.5    | 71  | 15.5| 13     |     |     |
| Contrast       |         |     |     |        |     |     |
| STIM-EXP       | 25.7    | 9   | 74.3| 26     |     |     |
| EXP-STIM       | 19.2    | 23  | 80.8| 97     |     |     |

Figure 3. Distribution of discourse relations in Exp. 2 for STIM-EXP and EXP-STIM verbs.
turned out to be stronger than I-Cons: In the present experiment, consequences were associated with a slightly stronger bias than explanations. We would like to point out that the present analysis is based on much fewer observations than our previous analysis, especially for the consequence continuations. Further research is required to investigate whether the I-Caus and I-Cons biases of the two classes of psych verbs do in fact differ. Except for this difference, the coreference in implicit explanations and consequences closely resembled the distributions observed for explicit relations in the previous experiment. This further corroborates the findings reported in Bott and Solstad (2014) and Kehler et al. (2008).

Given the strong across-the-board differences between discourse relations, we will refrain from any inferential statistical analysis of anaphoric forms chosen for explanations, consequences and other relations. We would just like to mention that personal pronouns were the most common means to refer back to the antecedent in both explanations (STIM-EXP 95.8%, and EXP-STIM 90.7% personal pronouns) and consequences (STIM-EXP 81.6% relative to EXP-STIM 95.2% personal pronouns). This pattern resembles the distribution observed in Experiment 1. Interestingly, other relations such as contrast gave rise to greater numbers of other forms, mainly repeated names with less than 60% personal pronouns in contrast relations overall.

4. Experiment 3: forced coreference and coherence

While Experiment 2 found a clear asymmetry between explanations and consequences after a full stop, one might argue that explanations are preferred over consequences for independent reasons. For example, Sanders (2005) proposed that providing explanations constitutes a default strategy in discourse. We argued above that this leaves unexplained the distributional differences between consequences and other discourse relations for STIM-EXP and EXP-STIM verbs observed in Experiment 2. Experiment 3 was designed to add further evidence to the explanation-consequence asymmetry.

To this end, the present experiment also investigated discourse continuations after a full stop. In this case, however, a forced coreference paradigm was employed (see, e.g. Fukumura & van Gompel, 2010), where participants had to provide continuations about the subject or the object. When crossing forced coreference with the two psych verb classes, conditions are achieved in which participants are forced to write about the I-Caus-biased referent or the I-Cons-biased referent, respectively. We reasoned that enforcing a continuation about the stimulus argument should trigger explanations, while being forced to provide a continuation about the experiencer should create optimal conditions for producing a consequence. The main question Experiment 3 was designed to answer is thus: Is the I-Caus explanation bias observed in Experiment 2 still present in continuations incongruent with the I-Caus coreference bias? If so, this would provide strong evidence in favour of the Asymmetry Hypothesis and the Two-Mechanism Account.

Rohde (2008, Exp. V, Ch. 4) conducted a related experiment, forcing particular coreference patterns by introducing a pronoun in the prompt (John infuriated Mary. He/She …). However, as the pronoun prompt comes with interpretation biases of its own, as argued by Rohde and colleagues (cf. e.g. Kehler Rohde, 2013), we believe that the forced coreference paradigm is better suited to investigate this issue. Also, since the verbs in Rohde’s items were chosen based on I-Caus bias patterns, and not verb class, the results of Rohde’s (2008) Experiment V isn’t directly informative for the balanced STIM-EXP and EXP-STIM verb classes.7

4.1. Design

The sentence continuation experiment employed a 2 × 2( × 2) within-participants and within-items design manipulating the factors VERB CLASS (STIM-EXP vs. EXP-STIM verbs), FORCED REFERENCE (forced coreference about NP1 or NP2), and GENDER ORDER (NP1 fem – NP2 masc vs. NP1 masc – NP2 fem). The latter factor was included in the design as a counterbalancing factor. As in the previous experiment, the continuations were annotated with respect to discourse relations and coreference.

4.2. Methods

4.2.1. Participants

64 students from the University of Tübingen (50 female, 14 male; mean age 23.5 years, range 18–32 years), all reporting to be native speakers of German, participated in the experiment for monetary compensation (5 Euro for half an hour). Participants were randomly assigned to the four lists with 16 participants in each list.

4.2.2. Materials

The 20 STIM-EXP and 20 EXP-STIM prompt items from the previous experiments were tested with a full stop. GENDER ORDER was again manipulated within items. In addition, FORCED REFERENCE was manipulated by
drawing a box around either the first or the second referent in the prompt. A Latin Square was used to distribute the resulting 20 items in four prompt conditions to four lists such that each item appeared twice in each list and each verb appeared once, that is, each STIM-EXP and each EXP-STIM verb was tested within each participant. 40 sentences with two referents of the same gender served as fillers and were added to each list. The lists were individually randomised for each participant.

4.2.3. Procedure
The procedure was the same as in the previous experiments with one crucial difference. In the present study, participants had to write a continuation about the referent in the box. They received written instructions with examples of continuations in accord with the forced coreference. They subsequently familiarised themselves with the task in a short practice of three trials. Experimental sessions took between 30 and 45 minutes.

4.2.4. Data annotation
The elicited data set of 2560 continuations was annotated for discourse relations and coreference using the same annotation scheme as in the previous experiment. In addition, it was annotated whether the continuation matched the forced reference. 145 (= 5.2%) continuations that were incomplete or nonsensical were excluded from the analysis; 74 (2.8%) continuations did not mention the focused referent as the first referent and were, therefore, also excluded from the analysis. The remaining 2282 continuations were analysed in a logistic mixed-effects regression analysis. The data were annotated by the first author and two student assistants. The inter-annotator agreement was determined on a subset of 200 continuations and each pairwise comparison had Cohen’s κ scores of at least 0.81 for the coreference annotation and at least 0.77 for the annotation of discourse relations. Before data analysis, the complete set of annotations was again checked for consistency by the second author. As in the previous analyses, the verb ‘getallen’ (“appeal to”) was excluded from the analysis.

4.3. Statistical analysis
The GLMER analysis of discourse relations was conducted as in the previous experiment. In addition to the fixed effect of VERB CLASS, the analyses included the main effect of FORCED REFERENCE (subject vs. object) as well as its interaction with VERB CLASS. As for the previous experiments, all analyses are publicly available (see note 3).

4.4. Results and discussion
The overall distribution of discourse relations is shown in Figure 4. As in the previous experiment, a clear asymmetry between explanations and consequences was observed. In I-Caus-congruent conditions, participants produced explanations 83.8% of the time after STIM-EXP verb prompts and 76.7% of the time after EXP-STIM prompts (3.7% and 3.0% consequences, respectively). Crucially, even in the I-Caus-inconsistent STIM-EXP NP2 and EXP-STIM NP1 forced reference conditions participants still produced 42.9% and 49.1% explanations, respectively. In the latter two conditions, there were overall even more explanations than consequences (44.2% and 32.4%, respectively), even though the forced reference prompts matched the verbs’ I-Cons bias and were inconsistent with their I-Caus bias.

Because explanation and consequence relations were the predominant relations used in the present experiment, we performed logit mixed-effects regression model analysis only analysing these two relations (coding consequence relations as 0 and explanation relations as 1). Since GENDER ORDER did not lead to any reliable effects ($\chi^2(4) = 5.76; p = .22$), this factor was left out in subsequent regression modelling. The GLMER analysis revealed a strong crossover interaction between VERB CLASS and FORCED REFERENCE in the direction of the respective I-Caus and I-Cons biases ($\beta = -6.80$, $SE = 0.41$, $\chi^2(1) = 513.47$, $p < .01$): the likelihood to provide an explanation was higher in I-Caus bias-congruent conditions than in incongruent ones. In addition, the intercept was significantly above a value of zero ($\beta = 1.97$, $SE = 0.17$, $\chi^2(1) = 60.17$, $p < .01$), showing that explanations were much more likely overall than consequence relations. This provides further evidence against symmetry between the two biases in terms of coherence and further corroborates the Asymmetry Hypothesis. To break down the interaction, subset analyses were computed separately for the two verb types. Both analyses revealed clear FORCED REFERENCE effects (STIM-EXP verbs: $\beta = -3.63$, $SE = 0.28$, $\chi^2(1) = 324.37$, $p < .01$; EXP-STIM verbs: $\beta = 3.41$, $SE = 0.32$, $\chi^2(1) = 211.08$, $p < .01$) as well as highly reliable positive intercepts (STIM-EXP verbs: $\beta = 1.79$, $SE = 0.20$, $\chi^2(1) = 46.92$, $p < .01$; EXP-STIM verbs: $\beta = 2.37$, $SE = 0.26$, $\chi^2(1) = 50.70$, $p < .01$).

As in the previous experiment, the distribution of anaphoric forms was analysed in explanation and consequence continuations ($N = 1821$) in which prompts received a subject-verb-object reading and which contained at least one anaphoric expression. Again, participants overwhelmingly used personal pronouns more than 90% of the time. An exploratory GLMER analysis on
anaphoric form (personal pronouns vs. other forms) taking into account the fixed effects of forced reference and discourse relation as well as their interaction led to a significant main effect of forced reference ($\beta = -1.15$, $SE = 0.25$, $\chi^2(1) = 23.44$, $p < .01$) in addition to a highly significant intercept ($\beta = 5.91$, $SE = 1.08$, $\chi^2(1) = 84.60$, $p < .01$). The fixed effect of forced reference was due to the fact that participants produced more personal pronouns in the subject forced reference conditions ($\bar{x} = 94.0\%$) than they did in the object forced reference conditions ($\bar{x} = 89.2\%$) – as would be expected under a first-mention or a subject-preference account (Brennan, 1995; Gernsbacher & Hargreaves, 1988).

5. Experiment 4: types of explanations and consequences

Experiments 1 through 3 showed that while I-Caus and I-Cons biases are mirror-like for STIM-EXP and EXP-STIM predicates, the status of the corresponding unmarked relations of explanations and consequences are strongly asymmetric. The Two-Mechanism Account, from which the Asymmetry Hypothesis was derived, claims that this stems from two different strategies for providing explanations and consequences for these predicates. Whereas explanations target a slot in the predicate’s semantic structure, consequences cannot target any comparable slot and can thus only specify consequences subsequent to the eventuality as a whole, following the Contiguity Principle (see Section 1.4.3).

Experiment 4 investigated the detailed properties of those explanations and consequences. More precisely, we investigated what characterises bias-congruent and bias-incongruent explanations and consequences for STIM-EXP and EXP-STIM verbs. The Asymmetry Hypothesis predicts that the specification of explanations and consequences should follow two distinct patterns, as they are driven by wholly different mechanisms, with different specification possibilities (see Figure 1 in Section 1.4.3). Whereas the properties of I-Caus explanations were investigated in Bott and Solstad (2014) and Bott and Solstad (2021), we are not aware of any studies that provide an in-depth investigation for I-Cons and compare them with the I-Caus patterns.

5.1. Design

A further manipulation was added to the design of Experiment 1. The congruency of explanations was manipulated by including a personal pronoun following the connective that was either congruent or incongruent

Figure 4. Distribution of discourse relations in Exp. 3 for STIM-EXP and EXP-STIM verbs with forced reference to NP1 or NP2.
with the expected I-Caus (for weil “because”) or I-Cons bias (for sodass “and so”), for instance, John annoyed Mary because/and so he/she. Consequently, the sentence continuation experiment employed a $2 \times 2 \times 2$ design manipulating within-participants and within-items the factors VERB CLASS (STIM-EXP vs. EXP-STIM), CONNECTIVE (weil “because” vs. sodass “and so”), CONGRUENCY (bias-congruent vs. bias-incongruent pronouns in the prompt) and GENDER ORDER (NP$_{1\text{fem}}$–NP$_{2\text{masc}}$ vs. NP$_{1\text{masc}}$–NP$_{2\text{fem}}$). GENDER ORDER was included in the design as a counterbalancing factor.

5.2. Methods

5.2.1. Participants
Fifty-six native German speakers (28 female, 28 male; mean age 24.6 years, range 18–35 years) who were recruited via the platform Prolific (www.prolific.co) participated in the experiment for monetary compensation (reward: 4.5 GBP per half hour). Participants were randomly assigned to eight lists with seven participants in each list. Participants gave their informed consent to the study. All participants reported German as their native language.

5.2.2. Materials
The 20 items (20 STIM-EXP and 20 EXP-STIM verbs) from the previous experiments were tested with prompts as in Johanna shocked/despised Lars because/and so she/he …. A Latin Square was used to distribute the resulting 20 items in 16 prompt conditions (including gender manipulation) to eight lists such that each item appeared twice in each list, but each verb appeared only once. It was furthermore ensured that each list contained five trials in each of the eight experimental conditions (i.e. two plus three trials in the two gender orders). The same set of 80 filler sentences from other verb classes (agent-patient and agent-evocator verbs) either followed by weil “because” or sodass “and so” alone or simply a full stop were added to each list. The lists were individually randomised for each participant.

5.2.3. Procedure
The procedure was identical to the one in Experiments 1 and 2. On average, the experiment took just over 30 minutes to complete.

5.2.4. Data annotation
The elicited data set of 2,240 continuations was annotated for subtypes of explanation and consequence relations. The annotation included some further, general categories that were applied also in the previous experiments (see above); 100 ($= 4.5\%$) continuations that were incomplete or nonsensical were excluded from the analysis. Another 155 ($= 6.9\%$) continuations were excluded in which participants had interpreted the prompt with object-verb-subject order. For reasons of consistency, all continuations for the verb gefallen (“appeal to”) were again excluded. The remaining 1963 continuations were annotated as follows.

The categories of the explanation relations were based on the ones applied in Bott and Solstad (2014) and Bott and Solstad (2021), but adjusted somewhat to allow for better comparison with consequence relations, in particular with reference to the Asymmetry Hypothesis. In the following, we merely list examples along with the categories. More elaborate remarks on the particular examples can be found in the Appendix. Explanations, that is, continuations as prompted by weil “because”, were annotated according to the following categories:

- **Explanatory specifications**: Does the explanation provide the direct, simple cause (Bott & Solstad, 2014, 2021) of the psychological state in the experiencer argument?
  (6) **Explanatory specification**: Aisha scared Albert because she was wearing a horrible mask.
- **Backgrounds**: Is the explanation necessary, but insufficient for the psychological experience?
  - (7) **(Non-mental) background**: Aisha scared Albert because he was sleeping.
  - (8) **Mental background**: Klaus astonished Marie because she had thought that he was stupid.
- **Reasons**: Does the explanation provide a rationale for the intentional action of an agent?
  (9) **Reason**: Aisha scared Albert because he deserved it.

The main category to be expected for congruent continuations is the one of explanatory specifications (6), as they specify the underspecified property of the stimulus argument (Bott & Solstad, 2014, 2021).

Consequences that is, sodass “and so” continuations, were annotated according to the following criteria:

- **Consequence specifications**: Does the consequence specify, or restate the mental state of the experiencer as given in the verb?
  (10) **Consequence specification**: Aisha scared Albert, and so he was very shocked.
• **Subsequent consequences**: Is the reference time of the consequence disjoint (and subsequent) to the end state in the predicate?
  ○ Consequences were additionally annotated for whether they included a mental state of the experiencer that was (i) proximate, or (ii) distant to the state described in the matrix predicate.

(11) **Subsequent (non-mental) consequence**:  
*Albert scared Aisha, and so she dropped her glass.*

(12) **Proximate experiencer-consequence**:  
*Patrick fascinated Anastasia, and so she fell completely in love with him.*

(13) **Distant experiencer-consequence**:  
*Alisa shocked Alexander, and so he needed some time for himself.*

• **Finality**: Does the soddass continuation constitute the goal of an actor? The conjunction soddass has an additional “in order to” reading which must be excluded as it constitutes a reason, that is, a kind of explanation.

For the Asymmetry Hypothesis, the most important category is the consequence specification in (10) because this is the effect provided in the predicate which would be a sub-eventuality mirroring the cause in the psychological relation between stimulus and experiencer.

The additional identification of subsequent experiencer-states that were mental in nature was performed to be able to investigate cases intermediate between consequence specifications and clearly separable subsequent events. In particular, the proximate experiencer-consequences (*Patrick fascinated Anastasia, and so she fell completely in love with him*) relate more closely to the mental state of the experiencer, that is, being fascinated and falling in love. Although we contend that these cases are subsequent in the sense discussed in relation to Figure 1 (Section 1.4.3), including this category in the statistical analysis allows for a better chance of falsifying the Asymmetry Hypothesis.

The annotation was done by the first author. A random sample of 300 continuations was independently annotated by the second author. The inter-annotator agreement proved to be good (Cohen’s \( \kappa = .75 \)), given the complexity of this semantic-pragmatic task. The complete set of annotations was checked for consistency by the first author.

5.2.5. **Statistical analysis**

After excluding 29 (= 1.3%) consequence continuations of finality type (see above), the remaining 1934 continuations were statistically analysed by fitting mixed-effects binomial logistic regression models including the centred fixed effects of **verb class**, **connective**, **congruency**, and their interactions as well as random intercepts for participants and items. Two analyses were conducted. The first analysis contrasted explanatory and consequence specifications (see (6) and (10)) relative to all other explanations and consequences, respectively. A second analysis was computed analysing what we characterise as **broad specifications**. In addition to the explanatory specifications and the consequence specifications (10) in the narrow sense, broad specifications also included subsequent consequences that constituted proximate experiencer-consequences (12). As for the previous experiments, all analyses are publicly available (see note 3).

5.3. **Results and discussion**

The distribution of explanation and consequence relations as detailed above is shown in Figure 5. Two clear patterns were observed. First, specifications in the narrow sense were almost exclusively found for explanations. And second, for those explanations, explanatory specifications were found almost only in congruent conditions, where they made up 93.1% of all explanations across verb classes (STIM-EXP: 87.6%; EXP-STIM: 98.2%). The data thus strongly support the Asymmetry Hypothesis.

Looking at the data in more detail, in the incongruent explanation conditions, explanatory specifications constituted the least frequent category (6.1%) when collapsed over both verb classes (STIM-EXP: 3.0%; EXP-STIM: 8.5%). Instead, incongruent explanations overwhelmingly represented mental backgrounds, see (8), for both verb classes (STIM-EXP: 66.3%; EXP-STIM: 67.1%). Reasons, which presuppose that an agent acted intentionally, were almost exclusively found with STIM-EXP verbs, many of which are ambiguous between a psych verb and an agent-patient reading (Bott & Solstad, 2014, see also the discussion on bias variability in the Introduction). Turning to consequences, the most frequent category was subsequent consequences with 76.9% in congruent (STIM-EXP: 67.4%; EXP-STIM: 86.2%) and 95.0% in incongruent conditions (STIM-EXP: 88.8%; EXP-STIM: 99.6%) across verb classes. Consequence specifications, however, were almost never produced with a total of five occurrences (STIM-EXP congruent: 1.2%, STIM-EXP incongruent: 0.6% and EXP-STIM congruent: 0.4%, EXP-STIM incongruent: 0.0%). Thus, specification of the mental state of the experiencer does not seem to be an available option. The category closest to consequence specifications, that is, proximate experiencer-consequences, see (12), were found mainly with STIM-EXP verbs (16.5% across conditions) and hardly with EXP-STIM verbs (2.2%
Figure 5. Distribution of Explanation types (a) and Consequence types (b) for STIM-EXP and EXP-STIM verbs in the congruent and incongruent conditions of Exp. 4.
overall). This makes sense since these verbs are assumed to be causative in a narrow sense and thus make available an end state which can form the starting point of the consequence of the Contiguity Principle (Section 1.4.3). Given the uniform picture with an overall lack of consequence specifications, no inferential statistics seemed to be warranted analysing explanatory and consequence specifications.

As to the category of broad specifications, a statistical analysis was conducted in which the category of consequence specifications was merged with (subsequent) proximate experiencer-consequences. Table 4 presents the distribution of broad specifications which were further investigated in a logit mixed-effects model regression analysis.

The GLMER analysis starting with a global model including the centred effects VERB CLASS, CONNECTIVE, CONGRUENCY and GENDER ORDER as well as their interactions showed that GENDER ORDER failed to contribute to model fit (χ²(8) = 4.60, p=.60). Consequently, the model reported here is a model only including the effects VERB CLASS, CONNECTIVE, CONGRUENCY and their interactions.

The regression analysis revealed that neither the three-way interaction between VERB CLASS, CONNECTIVE and CONGRUENCY (χ²(1) = .00, p=.98) nor the two-way interaction between CONGRUENCY and VERB CLASS significantly improved model fit (χ²(1) = 3.27; p = .07). The most parsimonious model included only the two two-way interactions between CONNECTIVE and CONGRUENCY (β = 4.32, SE= 0.68, χ²(1) = 86.02, p <.01), and between CONNECTIVE and VERB CLASS (β = −4.14, SE= 0.61, χ²(1) = 89.60, p <.01). The first interaction was due to the fact that CONGRUENCY had a much stronger effect on explanations than it had on consequences. This is exactly the differential effect predicted by the Asymmetry Hypothesis. The second interaction reflects the fact that, across the board, more explanations were of the specifying type for EXP-STIM verbs than for STIM-EXP verbs, whereas the opposite trend could be observed for consequences. On the consequence side, this makes sense because only STIM-EXP verbs are causative with a clear end state. On the explanation side, this pattern is consistent with the results from Experiment 1, where it was observed that the selected EXP-STIM verbs show a generally stronger bias than the STIM-EXP verbs, which should be reflected by a stronger need for specification for the EXP-STIM than for the STIM-EXP verbs. Even though this interaction was not among our initial predictions, this is exactly what the Two-Mechanism Account would lead us to expect.

To summarise, the results from Experiment 4 show that specifications of the contingency relation introduced by the predicate is only a viable option for explanations and that this strategy is almost only available for bias-congruent explanations, as predicted by the Asymmetry Hypothesis (following the Empty Slot Theory). For consequences, a specification strategy is not an option. The results furthermore show that subsequent consequences that are closer to the experiencer’s state of mind (dubbed broad consequence specifications above) are sensitive to congruency manipulations. This follows naturally from the fact that I-Cons bias-incongruent consequences focus on the stimulus argument and must thus involve a separate eventuality subsequent to the one introduced by the psych verb in the prompt.

### 6. General discussion

The study presented in this paper investigated the foundations of Implicit Causality and Implicit Consequentiality for stimulus-experiencer and experiencer-stimulus verbs. For these verb classes, previous research had observed mirror-like coreference biases and it was hypothesised that these biases were caused by one common mechanism related to verb argument structure (Crinean & Garnham, 2006; Hartshorne et al., 2015). Four experiments investigated coreference and coherence biases as well as the finer distribution of congruency types for the two verb classes. We found persistent evidence for the Asymmetry Hypothesis, according to which the biases are driven by two different

### Table 4. Relative (%) and absolute (n) frequencies of explanation and broad consequence specifications for STIM-EXP and EXP-STIM verbs in Exp. 4.

| Explanation        | Consequence        |
|--------------------|--------------------|
|                     | broadly specifying | other |
|                     | %      | n  | %      | n  |
| STIM-EXP           |        |    |        |    |
| congruent          | 87.6   | 227| 12.4   | 32 |
| incongruent        | 3.0    | 6  | 97.0   | 193|
| EXP-STIM           |        |    |        |    |
| congruent          | 98.2   | 271| 1.8    | 5  |
| incongruent        | 8.5    | 22 | 91.5   | 236|

The GLMER analysis starting with a global model including the centred effects VERB CLASS, CONNECTIVE, CONGRUENCY and GENDER ORDER as well as their interactions showed that GENDER ORDER failed to contribute to model fit (χ²(8) = 4.60, p=.60). Consequently, the model reported here is a model only including the effects VERB CLASS, CONNECTIVE, CONGRUENCY and their interactions.
mechanisms, as described in Section 1.4.3: I-Caus, we contend, reflects a slot-filling strategy for underspecified content in the verb’s semantic structure. I-Cons, on the other hand, reflects a general, verb-class independent Contiguity Principle, according to which consequences take the final end-point of previous discourse as their reference point.

Experiment 1 established that the STIM-EXP and EXP-STIM verb classes did indeed display mirror-like coreference biases, with strong I-Caus NP1 biases and I-Cons NP2 biases for STIM-EXP verbs and similarly pronounced I-Caus NP2 and I-Cons NP1 biases for EXP-STIM verbs. These findings corroborate results from previous research (Bott & Solstad, 2014; Crinean & Garnham, 2006; Ferstl et al., 2011; Garnham et al., 2020; Hartshorne et al., 2015; Stewart et al., 1998).

Experiment 2 investigated the coherence bias of the two verb classes, that is, the discourse relation established between the continuation and the prompt after a full stop. On the Asymmetry Hypothesis, explanations are evoked by empty slots anchored in verb semantics, while consequences follow the more general Contiguity Principle. General processing principles of not leaving slots unspecified, that is, providing specifications for underspecified entities, should lead to a general preference to provide slot-filling explanations above contiguity-related consequences. Our findings confirmed the predictions from the Asymmetry Hypothesis: For both verb classes, explanations constituted the most frequent category, occurring more than three times as often as consequences, the second-most frequent category. Also of notice, for experiencer-stimulus verbs, consequences constituted only the third-most frequent category, dominated also by contrast relations. Thus, although the coreference biases are comparable for I-Caus and I-Cons, as shown in Experiment 1, there is a clear asymmetry in the distribution of the discourse relations associated with I-Caus and I-Cons, with a strong preference to produce explanations over consequences. These findings confirm observations in previous research (Kehler et al., 2008), but do so for a more carefully selected set of verbs with regard to coreference biases.

Experiment 3 investigated whether the coherence biases in Experiment 2 reflected a general preference for explanations independent of I-Caus and I-Cons, such as a “causality-by-default” strategy (e.g. Sanders, 2005), according to which explanations could be preferred for independent reasons. To this end, a forced reference task was employed in which participants had to produce continuations to either the subject or the object argument after a full stop. Being forced to provide a continuation about the consequence-associated experiencer argument should produce more consequences than explanations if there is a common mechanism underlying coreference and coherence bias. In any case should the forced reference work against any preference for the Empty Slot strategy over the Contiguity Principle. Although we found that forced reference had an influence on discourse strategies, it was in no way able to fundamentally alter the strategy. Instead, we again found that providing explanations constituted the preferred strategy for continuations, further supporting the Asymmetry Hypothesis.

Finally, Experiment 4 investigated more closely the particular explanation and consequence strategies that participants employ in providing continuations for I-Caus and I-Cons conditions. The Asymmetry Hypothesis leads to the prediction that two rather different strategies are followed: Explanations fill a predicate slot, providing specifications of a cause inherent in the verb, whereas consequences provide a different eventuality subsequent to the eventuality denoted by the verb. The findings confirmed these predictions: Providing explanations congruent with I-Caus is overwhelmingly done by specifying the cause of the mental state in the verb, whereas providing consequences congruent with I-Cons is done by introducing a new eventuality subsequent to the one in the prompt. Correspondingly, being forced to produce I-Caus incongruent continuations leads to a change in strategy, where continuations do not fill the slot, but constitute different types of explanations. For I-Cons, consequence specifications aren’t available in the first place, hence specification in a narrow sense plays no role. Even broadening the category of consequence specifications including more temporally remote consequences displays strong differences between I-Caus and I-Cons strategies and again a difference between verb classes.

As all reported data from Experiments 2 through 4 must be seen as providing strong support for the Asymmetry Hypothesis, it seems fair to conclude that the assumptions behind the Two-Mechanism Account based on the Empty Slot Theory and the Contiguity Principle are strengthened through our study: I-Caus bias is driven by verb semantics, while I-Cons is grounded in more general discourse structuring principles.

This Two-Mechanism Account was contrasted with previous accounts that offer a unified view of I-Caus and I-Cons. These have either assumed that verb semantics is not a central part of the mechanism driving the observed biases for I-Caus and I-Cons (Pickering & Majid, 2007), or, more interestingly from the perspective of the present study, that there is one mechanism common to both biases (Crinean & Garnham, 2006; Hartshorne et al., 2015). According to the One-Mechanism
Accounts, I-Caus and I-Cons are both based on verb semantics. More specifically, explanations and consequences are related to the thematic roles in the verb’s argument structure. Thus, for instance, the most elaborate One-Mechanism Account for I-Caus (Hartshorne et al., 2015) makes use of VerbNet’s semantic structure to predict I-Caus biases, according to which the experiencer’s mental state occurs “in reaction to” an event (uality) associated with the stimulus. Consequently, I-Caus bias prompts with because lead to coreference with the stimulus, since the mental state is contingent on this argument. Conversely, I-Cons bias prompts with and so are associated with the experiencer, as this is the argument that “reacts” to the stimulus.

We believe that the Two-Mechanism Account can better account for the experimental findings of the present study than what we have characterised as One-Mechanism Accounts. First, the propositional, explanatory slots on the explanation side offer a more straightforward connection between the coreference and the coherence bias, as discussed in detail in Bott and Solstad (2014) and Bott and Solstad (2021). While we obviously do not deny that argument structure correlates with I-Cons and I-Caus biases, no semantic or pragmatic mechanisms that predict the coherence bias can be derived directly from, for instance, argument structure (Crinean & Garnham, 2006) or VerbNet formalisations (Hartshorne & Snedeker, 2013). Thus, if a causative semantic component were decisive, all causative predicates should be associated with NP1 bias. However, a number of causative agent-patient verbs (i.e. not psych verbs), such as kill, were found to display a balanced bias in Ferstl et al. (2011). Also, the Two-Mechanism Account is the only currently available account that can predict the explanation and consequence patterns found for continuations congruent and incongruent with I-Caus and I-Cons bias reported in Experiment 4. In Solstad and Bott (in prep.), it is shown in more detail how verb classes that lack a slot according to our theory, such as causative agent-patient verbs, compare to predicates with an explanatory slot, such as agent-evocator verbs. We mention here some results of relevance to the current paper. First, slot-less causative agent-patient predicates display an overall balanced I-Caus bias (45.4% NP1 coreference). Object-biased agent-evocator predicates, on the other hand, which carry an explanatory NP2-associated slot on our theory, display a strong I-Caus NP2 bias (7.9% NP1 coreference). For I-Cons, the Contiguity Principle predicts object continuations for both verb classes and this is indeed what we find (17.1% NP1 coreference for agent-patient and 5.5% NP1 coreference for agent-evocator verbs). With regard to coherence biases, explanation is only the major discourse relation for the agent-evocator verbs, comparable in strength to the biases reported in this paper (58.3% explanations). For slot-less agent-patient verbs, explanations are still frequent (31.9% explanations), as expected from a “causality by default” strategy Sanders (2005). However, for these verbs, consequences constitute the dominant category (41.9% vs. 21.1% consequences for agent-evocator verbs), as would follow from the Contiguity Principle without a competing empty slot.

Our results corroborate both the results of earlier investigations of the I-Caus and I-Cons coreference bias for the relevant verb classes (Abelson & Kanouse, 1966; Crinean & Garnham, 2006; Ferstl et al., 2011; Hartshorne et al., 2015; Hartshorne & Snedeker, 2013; Stewart et al., 1998) as well as those on their coherence bias (Bott & Solstad, 2014, 2021; Kehler et al., 2008; Kehler & Rohde, 2019; Rohde, 2008). By controlling for not only I-Caus bias but also the semantic verb classes, a more elaborate and systematic picture of the coherence relation distribution is provided in the present paper than in Kehler et al. (2008), who first investigated the relation between coreference and coherence. In addition, our study complemented the contingency typology for I-Caus in the studies by Bott and Solstad (2014) and Bott and Solstad (2021) with the first assessment of the types of consequences found with I-Cons.

Given that I-Caus and, more limited, I-Cons have been investigated with respect to their time course during online comprehension, we would like to point out that the Two-Mechanism Account can be applied to existing results but also offers possible avenues for future investigation. Section 1.3 reviewed experimental evidence showing that early effects of I-Caus have been found across different online methodologies, but there is only one study published on the online effects of I-Cons (Garnham et al., 2020), as far as we are aware, and one allowing for the comparison of explicit explanations and consequences (Hoek et al., 2021a). First, although the evidence is somewhat difficult to interpret, as we mentioned in the introduction, our results concerning the primacy of explanations are compatible with the finding by Hoek et al. (2021a) that explanatory connectives are processed faster than consequential ones. Second, in their visual world eyetracking study, Garnham and colleagues found what they refer to as early effects of both I-Caus and I-Cons and observe that for I-Caus even earlier effects may be plausible, although this “very early” effect didn’t turn out to be reliable in their study (for effects prior to the connective, see Pyykönén & Järvikivi, 2010; Rohde & Horton, 2014).

How do these early effects for both I-Caus and I-Cons fit into the present account? Based on our assumptions
of two independent mechanisms, one wouldn’t necessarily expect I-Caus and I-Cons to give rise to similar effects. However, we do think that they are compatible with our approach and that our account can actually offer an interesting perspective on the mechanisms involved in the effects found by Garnham et al. (2020). Assuming incremental semantic and pragmatic interpretation (Altmann & Steedman, 1988; Crain & Steedman, 1985; Millis & Just, 1994, among many others), the connective is immediately integrated into the representation, upon which the Contiguity Principle can be immediately applied.

From this principle, the (end) state argument of psychological predicates would become more easy to integrate, which should lead to early effects. Next, one may ask whether the two mechanisms could lead to any differences between the relative time courses of I-Caus and I-Cons, as discussed by Garnham et al. (2020)? We believe that differences could be observed along two different interpretation aspects of I-Caus and I-Cons. Since the causal connective because also fulfills the coherence bias associated with the involved predicates, there should be a processing advantage for I-Caus over I-Cons. The I-Caus connective because serves the coherence bias, that is, the anticipated coherence relation, and therefore also the coreference bias as argued in the Empty Slot Theory. The I-Cons connective and so, however, works against the coherence bias and should thus lead to some disruption as compared to I-Caus. Although reassignment of probabilities of likely continuations could (and should) be fast, one could expect effects of I-Caus to be even earlier than effects of I-Cons given this “combined advantage”, as contemplated in the distinction between early and very early effects by Garnham et al. (2020). To summarise, the proposed Two-Mechanism Account emphasises the need to compare I-Caus and I-Cons. Although some anticipatory effect based on the Contiguity Principle cannot be excluded, we assume an I-Cons associated connective such as and so to involve a disruption with the coherence bias. Still, after integrating the connective, expectations for a particular referent could be evoked.

In this study, we focused on only two of the four most commonly assumed verb classes that display strong I-Caus and/or I-Cons biases, stimulus-experiencer and experiencer-stimulus verbs. We did so because they present the strongest case for testing the Asymmetry Hypothesis, since the biases relate to the same argument types that are distributed mirror-like on the subject and object arguments. However, to be able to assess the Empty Slot Theory and the influence of the Contiguity Principle more properly, it should be stressed that an investigation of the two other main classes of interest, agent-evocator and ordinary agent-patient verbs, is called for. In Solstad and Bott (in prep.), such a study with agent-evocator verbs and (causative) agent-patient verbs has been conducted. The combined results of the two studies will allow a more comprehensive view on the patterns and interconnections between coreference and coherence biases and strategies for (congruent and incongruent) explanations and consequences.

To conclude, our study has provided the case for an asymmetric view on Implicit Causality and Implicit Consequentiality biases. We made the argument that two different mechanisms are involved. On the one hand, I-Caus is assumed to be governed by a lexically determined specification strategy, whereas I-Cons is assumed to derive from general discourse continuation strategies. If this account is on the right track, it makes the two phenomena potentially more interesting for further investigations of offline and online processing, where one phenomenon, I-Caus, is governed by features implicit in a predicate, whereas the other, I-Cons, occurs only in contexts made explicit by discourse means such as connectives. Hence, it would be more appropriate to refer to the coreference biases as implicit causality and explicit consequentiality, respectively.

---

**Notes**

1. We strongly disagree with this view and consider it highly important to look into representative samples of verbs. It must be emphasised though, that the paper by Pickering and Majid (2007) is a commentary to the proposal by Crinean and Garnham (2006), and may, therefore, necessarily lack empirical breadth.
2. Kehler and Cohen (2016) ground the bias in a “elicitation” implicature – a special kind of inference – mechanism. However, this theory does not assume a role for verb semantics, merely world knowledge in combination with this special inference mechanism.
3. The OSF repository can be found at [https://osf.io/Sewbd/](https://osf.io/Sewbd/).
4. The OVS reading was assessed based on semantic and pragmatic properties of the continuation. Take a sequence like Jennifer gefiel Tobias, weil …er sich gut um seine Schwester kümmerte “Jennifer liked Tobias because …he took good care of his sister”. In this case, it is clear from world knowledge that Tobias must be the stimulus since he is the agent of a deed (taking care of his sister) that is suited to cause someone to like someone else. As the stimulus is realised as the subject for gefallen, the participant must have read the prompt in an OVS order. 
5. An anonymous reviewer raised the concern that the pairing of STIM-EXP and EXP-STIM verbs in our items may not be close enough to really warrant the reported within-items analysis of VERB TYPE. For this reason, the main analyses of all four experiments reported in this
paper were repeated treating VERB TYPE as a between-items factor with 40 items instead of 20. This concerned the coreference analysis of Exp. 1, the discourse analyses of Exp. 2 and 3, and the cause/consequence analysis of Exp. 4. These additional analyses revealed qualitatively the exact same patterns of effects as the analyses reported below.

6. As two reviewers pointed out, one could worry that the ensuing repetition of verbs could have unwanted effects in that a participant’s continuation for a particular verb in one trial could influence the continuation provided by the same participant in a subsequent trial. We looked into the first four items in a sample of 1248 (= 20%) continuations (critical and filler items from Experiment 1 and 2) to check whether this was the case and found that in 18 (= 1.4%) cases, participants had provided a continuation for the prompts in Experiment 1 and 2 that was more or less identical to one of their previous continuations. Given the size of the effects reported in this paper, we consider this effect to be negligible.

7. Rohde (2008) doesn’t state explicitly which verbs she used in Experiment V, only that they were a subset of particularly strongly I-caus NP1- and NP2-biased verbs (in addition to non-biased ones). Also, she doesn’t report on the proportion of Consequences/Results for NP2-biased verbs.

8. Any model including fixed effects of VERB CLASS failed to converge. Thus, this was the maximal model that could be estimated.

Acknowledgments

We would like to thank Emiel van den Hoven as well as two anonymous reviewers for their valuable comments. We are also grateful to our student assistants Elisabeth Grün and Mathias Schrumpf for assisting us with the annotation of the data. Finally, we would like to thank our student assistant Natalie Clarius and the students in an MA seminar in General Linguistics on reference production at the University of Tübingen in the summer term 2018, in which some of the data were collected. The usual disclaimers apply.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was funded by the University of Bielefeld and the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) for projects B2 at the CRC 833 (SFB 833, University of Tübingen) “The construction of meaning” – project ID 75650358 and the project Composition in Context ( CiC) as part of the Priority Program (SPP 1277) XPrag.de: New Pragmatic Theories based on Experimental Evidence.

ORCID

Torgrim Solstad DOI: http://orcid.org/0000-0003-2405-1366
Oliver Bott DOI: http://orcid.org/0000-0003-3213-2123

References

Abelson, R., & Kanouse, D. (1966). Subjective acceptance of verb generalizations. In S. Feldman (Ed.), Cognitive consistency (pp. 173–199). Academic Press.

Aloni, M., & Roelofsen, F. (2011). Interpreting concealed questions. Linguistics and Philosophy, 34(5), 443–478. https://doi.org/10.1007/s10988-011-9102-9

Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. Cognition, 30(3), 191–238. https://doi.org/10.1016/0010-0277(88)90020-0

Asher, N. (2011). Lexical meaning in context. Cambridge University Press.

Asher, N., & Lascarides, A. (2003). Logics of conversation. Cambridge University Press.

Asher, N., & Vieu, L. (2005). Subordinating and coordinating discourse relations. Lingua, 115(4), 591–610. https://doi.org/10.1016/j.lingua.2003.09.017

Au, T. K. (1986). A verb is worth a thousand words: the causes and consequences of interpersonal events implicit in language. Journal of Memory and Language, 21(1), 104–122. https://doi.org/10.1016/0749-596X(86)90024-0

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of Memory and Language, 68(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. Journal of Statistical Software, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bott, O., & Solstad, T. (2014). From verbs to discourse – a novel account of implicit causality. In B. Hemforth, B. Mertins, & C. Fabricius-Hansen (Eds.), Psycholinguistic approaches to meaning and understanding across languages (pp. 213–251). Springer.

Bott, O., & Solstad, T. (2021). Discourse expectations: Explaining the implicit causality biases of verbs. Linguistics, 59(2), 361–416. https://doi.org/10.1515/ling-2021-0007

Brennan, S. E. (1995). Centering attention in discourse. Language and Cognitive Processes, 10(2), 137–167. https://doi.org/10.1080/01690969508407091

Brown, R., & Fish, D. (1983). The psychological causality implicit in language. Cognition, 14(3), 237–273. https://doi.org/10.1016/0010-0277(83)90006-9

Clifton, C. J., & Frazier, L. (2012). Discourse integration guided by the ‘question under discussion’. Cognitive Psychology, 65(2), 352–379. https://doi.org/10.1016/j.cogpsych.2012.04.001

Commandeur, E. (2010). Implicit causality and implicit consequentality in language comprehension. [Doctoral dissertation]. Tilburg University. https://research.tilburguniversity.edu/en/publications/implicit-causality-and-implicit-consequentality-in-language-comp

Corrigan, R. (1988). Who dun it? The influence of actor-patient animacy and type of verb in the making of causal attribution. Journal of Memory and Language, 27(4), 447–465. https://doi.org/10.1016/0749-596X(88)90067-8

Corrigan, R. (2001). Implicit causality in language: Event participants and their interactions. Journal of Language and Social Psychology, 20(3), 285–320. https://doi.org/10.1177/0261927X01020003002

Cozijn, R., Commandeur, E., Vonk, W., & Noordman, L. G. (2011). The time course of the use of implicit causality information
in the processing of pronouns: A visual world paradigm study. Journal of Memory and Language, 64(4), 381–403. https://doi.org/10.1016/j.jml.2011.01.001

Crain, S., & Steedman, M. (1985). On not being let up the garden path: The use of context information by the psychological syntax processor. In D. Dowty, L. Karttunen, & A. Zwicky (Eds.), Natural language parsing: Psychological, computational, and theoretical perspectives (pp. 320–358). Cambridge University Press.

Crinean, M., & Garnham, A. (2006). Implicit causality, implicit consequentiality and semantic roles. Language and Cognitive Processes, 21(5), 636–648. https://doi.org/10.1080/01690960600199763

Dowty, D. (1979). Word meaning and Montague grammar. Reidel.

Efron, B., & Tibshirani, R. (1986). The bootstrap method for standard errors, confidence intervals, and other measures of statistical accuracy. Statistical Science, 1(1), 1–35. https://doi.org/10.1214/ss/1177013815

Ehrlich, K. (1980). Comprehension of pronouns. Quarterly Journal of Experimental Psychology, 32(2), 247–255. https://doi.org/10.1080/14607480808401161

Featherstone, C. R., & Sturt, P. (2010). Because there was a cause for concern: An investigation into a word-specific prediction account of the implicit-causality effect. The Quarterly Journal of Experimental Psychology. 63(1), 3–15. https://doi.org/10.1080/17470210903134344

Ferstl, E. C., Garnham, A., & Manouilidou, C. (2011). Implicit causality bias in English: A corpus of 300 verbs. Behavior Research Methods, 43(1), 124–135. https://doi.org/10.3758/s13428-010-0023-2

Fillmore, C. J. (1969). Verbs of judging: An exercise in semantic description. Paper in Linguistics, 1(1), 91–117. https://doi.org/10.1080/08331816909389109

Frederiksen, A. T., & Mayberry, R. I. (2021). Implicit causality biases and thematic roles in American sign language. Behavior Research Methods, 53(5), 2172–2190. https://doi.org/10.3758/s13428-021-01561-1

Fukumura, K., & van Gompel, R. P. (2010). Choosing anaphoric expressions: Do people take into account likelihood of reference? Journal of Memory and Language, 62(1), 52–66. https://doi.org/10.1016/j.jml.2009.09.001

Garvin, A., Child, S., & Hutton, S. (2020). Anticipating causes and consequences. Journal of Memory and Language, 114, Article 104130. https://doi.org/10.1016/j.jml.2020.104130

Garnham, A., Traxler, M., Oakhill, J., & Gernsbacher, M. A. (1996). The locus of implicit causality effects in comprehension. Journal of Memory and Language, 35(4), 517–543. https://doi.org/10.1006/jmla.1996.0028

Garvin, A., Vorthmann, S., & Kaplanova, K. (2020). Implicit consequentiality bias in English: A corpus of 300+ verbs. Behavior Research Methods, 53(4), 1530–1550. https://doi.org/10.3758/s13428-020-01507-z

Garvey, C., & Caramazza, A. (1974). Implicit causality in verbs. Linguistic Inquiry, 5(3), 459–464. https://www.jstor.org/stable/4177835

Gernsbacher, M. A., & Hargreaves, D. (1988). Accessing sentence participants: The advantage of first mention. Journal of Memory and Language, 27(6), 699–717. https://doi.org/10.1016/0749-596X(88)90016-2

Goikoetxea, E., Pascual, G., & Acha, J. (2008). Normative study of the implicit causality of 100 interpersonal verbs in Spanish. Behavior Research Methods, 40(3), 760–772. https://doi.org/10.3758/BRM.40.3.760

Hartshorne, J. K. (2014). What is implicit causality? Language, Cognition and Neuroscience, 29(7), 804–824. https://doi.org/10.1080/01690965.2013.796396

Hartshorne, J. K., O’Donnell, T. J., & Tenenbaum, J. B. (2015). The causes and consequences explicit in verbs. Language, Cognition and Neuroscience, 30(6), 716–734. https://doi.org/10.1080/23273798.2015.1008524

Hartshorne, J. K., & Snedeker, J. (2013). Verb argument structure predicts implicit causality: The advantages of finer-grained semantics. Language and Cognitive Processes, 28(10), 1474–1508. https://doi.org/10.1080/01690965.2012.689305

Hobbs, J. R. (1979). Coherence and coreference. Cognitive Science, 3(1), 67–90. https://doi.org/10.1207/s15516709cog0301_4

Hoek, J., Rohde, H., Evers-Vermeul, J., & Sanders, T. J. M. (2021a). Expectations from relative clauses: Real-time coherence updates in discourse processing. Cognition, 210, Article 104581. https://doi.org/10.1016/j.cognition.2020.104581

Hoek, J., Rohde, H., Evers-Vermeul, J., & Sanders, T. J. M. (2021b). Scolding the child who threw the scissors: shaping discourse expectations by restricting referents. Language, Cognition and Neuroscience, 36(3), 382–399. https://doi.org/10.1080/23273798.2020.1852292

Itzhak, I., & Baum, S. R. (2015). Misleading bias-driven expectations in referential processing and the facilitative role of contrastive accent. Journal of Psycholinguistic Research, 44(5), 623–650. https://doi.org/10.1007/s10936-014-9306-6

Järvisivu, J., van Gompel, R. P. G., & Hyönä, J. (2017). The interplay of implicit causality, structural heuristics, and anaphor type in ambiguous pronoun resolution. Journal of Psycholinguistic Research, 46(3), 525–550. https://doi.org/10.1007/s10936-016-9451-1

Kehler, A. (2002). Coherence, reference, and the theory of grammar. CSLI Publications.

Kehler, A. (2004). Discourse topics, sentence topics, and coherence. Theoretical Linguistics, 30(2–3), 227–240. https://doi.org/10.1515/thli.2004.30.2-3.227

Kehler, A., & Cohen, J. (2016). On the presuppositional behavior of coherence-driven pragmatic enrichments. In M. Moroney, C. R. Little, J. Collard, & D. Burgdorf (Eds.), Semantics and linguistic theory (SALT) (Vol. 26, pp. 961–979). LSA and CLC Publications.

Kehler, A., Kertz, L., Rohde, H., & J. L. Elman (2008). Coherence and coreference revisited. Journal of Semantics, 25(1), 1–44. https://doi.org/10.1093/jos/ffm018

Kehler, A., & Rohde, H. (2013). A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. Theoretical Linguistics, 39(1–2), 1–37. https://doi.org/10.1515/tl-2013-0001

Kehler, A., & Rohde, H. (2017). Evaluating an expectation-driven question-under-discussion model of discourse interpretation. Discourse Processes, 54(3), 219–238. https://doi.org/10.1080/01638533.2016.1169069

Kehler, A., & Rohde, H. (2019). Prominence and coherence in a Bayesian theory of pronoun interpretation. Journal of Pragmatics, 154, 63–78. https://doi.org/10.1016/j.pragma.2018.04.006

Kipper, K., Korhonen, A., Ryant, N., & Palmer, M. A. (2008). A large-scale classification of English verbs. Language, 84(3), 701–727. https://doi.org/10.1353/lan.0.0199
van den Hoven, E., & Ferstl, E. C. (2018a). Discourse context modulates the effect of implicit causality on rementions. *Language and Cognition, 10*(4), 561–594. https://doi.org/10.1017/langcog.2018.17

van den Hoven, E., & Ferstl, E. C. (2018b). The roles of implicit causality and discourse context in pronoun resolution. *Frontiers in Communication, 3*. https://doi.org/10.3389/fcomm.2018.00053

van Kuppevelt, J. (1995). Discourse structure, topicality and questioning. *Journal of Linguistics, 31*(1), 109–147. https://doi.org/10.1017/S002222670000058X