From Verbs to Discourse: A Novel Account of Implicit Causality

Oliver Bott and Torggrim Solstad

Abstract We present a semantic theory of causal discourse which allows us to derive expectations about upcoming discourse, specifically when to expect various types of explanations. We apply our theory to the phenomenon of *implicit causality* and show how the semantics of implicit causality verbs interacts with discourse structure. In particular, we show why certain verbs trigger explanations per default, what kind of explanations are triggered, and why these explanations are closely related to specific coreference patterns. Predictions derived from the theory were tested in a large-scale crosslinguistic production study comparing discourse continuations in German and Norwegian – two languages which display differing discourse structuring properties. The production study fully confirmed our predictions, with explanation types and implicit causality bias distributing as expected. The study furthermore reveals that our semantic account of implicit causality is cross-linguistically valid.

1 Introduction

A crucial task in processing discourse is the identification of discourse relations pertaining between individual sentences and clauses. Such relations may or may not be marked, cf. (1):

- (1) a. Mary fascinated John because she always knew what to say.
- b. Mary fascinated John. She always knew what to say.

Oliver Bott
SFB 833, University of Tübingen
e-mail: oliver.bott@uni-tuebingen.de

Torggrim Solstad
Department of Language and Literature, Norwegian University of Science and Technology
e-mail: torggrim.solstad@ntnu.no

In (1a), the connective *because* marks the state of affairs expressed in the second clause as the *cause* of John's fascination. Although the relation between the two sentences in (1b) is not marked, it is strongly preferred to interpret the second sentence similarly to the *because* clause in (1a), i.e. as an explanation of the matter introduced in the first clause, i.e. *Mary fascinated John*.

Interestingly, such discourse relations are not only inferred *post hoc*, i.e. after processing the last sentence in sequences such as (1b). It has been argued that in many cases, we may *predict* the discourse relation which will pertain between the current state of discourse and the upcoming sentence to be processed, even based on the information contained in a single sentence (Kehler et al. 2008).

In psycholinguistics and social psychology, such expectations have been discussed with regard to a number of verbs which are characterized as *implicit causality verbs* (henceforth, IC verbs; Garvey & Caramazza 1974; Brown & Fish 1983; Au 1986; Rudolph & Försterling 1997; Pickering & Majid 2007). In general, IC verbs are transitive verbs with two animate arguments characterized by the particular property of triggering explanations focusing systematically on one of the two arguments when followed by a *because* clause. Such preferences are standardly elicited in production studies where participants are prompted to continue the sequence "ARGUMENT-VERB-ARGUMENT *because*" freely (see e.g. Garvey & Caramazza 1974; Ferstl et al. 2011), cf. (2):

- (2) a. *Mary* fascinated *John* because ... *she* always knew what to say.
- b. *Mary* admired *John* because ... *he* always knew what to say.

For *fascinate*, there is a strong preference for providing an explanation referring primarily to the subject argument (2a). For *admire*, on the other hand, continuations preferably refer to the object argument (2b). The proportion of continuations mentioning the subject or object argument is referred to as *IC bias*. With reference to the linear order of the two arguments, *fascinate* is characterized as an *NP1 bias* verb (*Mary*, the subject argument, is the first argument one encounters when processing the sentence from left to right), whereas *admire* is referred to as an *NP2 bias* verb. Bias-compliant continuations are termed *congruent* (3a), whereas non-bias-compliant continuations are *incongruent* (3b):

- (3) a. *Mary* admired *John* because *he* always knew what to say. (=2b)
- b. *Mary* admired *John* because *she* was easily impressed.

While the preference for coreference in (2) is certainly interesting and has proven useful in a number of experimental paradigms (see Sect. 2), two other aspects of IC are of more central importance, as far as we are concerned. First, it has been shown by Kehler et al. (2008) that IC verbs trigger a significantly higher number of explanations when prompted for continuation with a full stop following the "ARGUMENT-VERB-ARGUMENT" sequence. Thus, with e.g. *fascinate*, participants tend to produce continuations of explanation (4a), and not result type (4b):

- (4) a. *Mary* fascinated *John*. *She* always knew what to say. (=Explanation)
- b. *Mary* fascinated *John*. *He* decided to ask her out. (=Result)

A second, highly important aspect concerns the relationship between coreference and explanation. It may be observed that incongruent continuations constitute explanations of a very different kind than bias-congruent continuations. Consider again (3). Whereas the bias-compliant continuation in (3a) explains Mary's admiration for John with reference to a property of John, the non-bias-compliant continuation in (3b) refers to a property of Mary.

There exists to date no comprehensive account of the above observations. Thus, concerning the bias itself, no accounts (including a very recent one by Hartshorne & Snedeker 2012) have gone beyond establishing correlation patterns between the bias and the argument structure properties of IC verbs. However, argument structure does not readily lend itself to explaining a phenomenon which involves *discourse*. Although the paper by Kehler et al. (2008) constitutes a crucial step towards connecting IC and discourse, the authors do not attempt to answer the question of *why* it is that IC verbs tend to trigger explanations in production.

Based on the observations concerning the relation between coreference and explanation, we present a semantic-compositional account of IC. IC verbs, we contend, are underspecified with respect to certain properties of the situation described which are (causally) contingent on one of the two participants. It is this missing information which triggers explanations in full stop continuations and which also triggers primary reference to one of the two participants. Thus, on our analysis the IC bias as a measure of coreference preferences is an epiphenomenon of specific explanatory preferences derived from verb semantics and the particular realization of its arguments. We will substantiate our claims both by way of detailed semantic analysis and by means of a production experiment in German and Norwegian showing that the bias is cross-linguistically stable and that the same mechanisms are operative across languages.

German and Norwegian were chosen because they constitute a particularly interesting pair of languages from a discourse perspective. With reference to the study by Kehler et al. (2008), German and Norwegian differ in substantial ways, making them well-suited for testing universal vs. language-specific aspects of explanatory discourse: While German exhibits more subordinating/hypotactic syntactic structures and tends to make discourse relations explicit by means of connectives, Norwegian is more coordinating/paratactic in nature with discourse relations often left implicit. Translation studies (cf. Solfjeld 1996; Fabricius-Hansen 1999; Fabricius-Hansen et al. 2005; Ramm 2011) provide evidence that there are opposing constraints on discourse segmentation: Norwegian translators of German regularly split up complex German sentences and introduce full stops where there were none in the original. As a consequence, Norwegian explanations are often not explicitly marked since the most commonly used discourse marker, *fordi* ('because'), is subordinating. Showing IC to be the same in explicitly and implicitly marked German and Norwegian explanatory discourse would thus make a strong point for the claim that IC is universally grounded in the semantics of verbs and the discourse relation of explanation. Such a comparison presupposes that verbs may be matched precisely. The close historical relationship between German and Norwegian ensures that we may find a high number of near-synonyms – often sharing the same root – for a given

lexical concept. To our knowledge, our study is the first to present a verb-by-verb comparison of IC bias, where previous studies have compared verb classes.

The paper is structured as follows: In Sect. 2, we present what we view to be the most central observations concerning IC in previous research. In Sect. 3, we present a semantic analysis of IC and derive predictions from this analysis. Section 4 presents the production study, whereas Sect. 5 presents the results of the study. Section 6 concludes the paper with considerations regarding predictions for online studies and directions for future research.

2 Previous Research

IC verbs are special in at least two respects. We already mentioned the finding from the discourse continuation study by Kehler et al. (2008) that IC verbs triggered explanations per default (>60%), whereas no such preference could be observed with “non-IC verbs” (<25% explanations). What is more, the general preference for resolving anaphora to the subject of the preceding sentence (e.g. Grosz et al. 1995) and to the entity mentioned first in the preceding sentence (e.g. Gernsbacher & Hargreaves 1988) is seemingly violated by NP2 bias verbs, cf. (2b). With such verbs there is a strong preference to resolve the pronoun to the (second-mentioned) object (for a comprehensive review of factors affecting pronoun resolution see Garnham 2001). An analysis of IC should bring these two aspects together. The upshot of such an approach would be that the explanation triggering aspects of IC verbs outrank other commonly assumed anaphora resolution constraints. For a discussion of such constraints, see Baumann et al. (this volume) and Colonna et al. (this volume).

Following initial observations in social psychology (Abelson & Kanouse 1966), IC verbs were first studied in psycholinguistics in the early 1970s by Caramazza, Garvey and colleagues (Garvey & Caramazza 1974; Garvey et al. 1974). Caramazza and colleagues were the first to employ the sentence completion task mentioned in Sect. 1. Since then, congruency effects of IC have been assessed to be highly regular, applying a number of psycholinguistic methods across different languages and across adults and children (cf. the review in Rudolph & Försterling 1997).

Garvey and Caramazza attributed this effect to a feature of causal directionality encoded in the lexical semantics of interpersonal verbs. In later work, Brown & Fish (1983) tried to link IC to systematic differences between certain classes of verbs and thematic properties of their arguments. They observed systematic differences between *actions* and *states*, corresponding essentially to the distinction between *agent-patient* (henceforth *a-p*) and two kinds of psychological verbs (henceforth, *psych verbs*), i.e. *stimulus-experiencer* and *experiencer-stimulus* verbs (henceforth *s-e* and *e-s*, respectively). They suggested that *a-p* verbs attribute causes to the *agent* role, while *psych verbs* attribute them to the *stimulus* argument. This line of research has been followed by other researchers and has led to a more fine-grained taxonomy, termed *revised action-state distinction* (e.g. Rudolph & Försterling 1997). Au (1986) noted that *a-p* verbs do not constitute a homogeneous class, but should be

further split into *a-p* verbs proper and a verb class which she termed *agent-evocator* verbs, i.e. verbs that denote interpersonal actions where the patient/theme presumably evoked the intention of the agent to act (like *thank*, *criticize*, *congratulate*). The studies by Rudolph (1997), Goikoetxea et al. (2008) and Ferstl et al. (2011) have provided norms for different languages and a large number of verbs lending support to the four-way classification into *a-p*, *agent-evocator*, *s-e* and *e-s* verbs.

In a framework based on folk psychology, Malle (2002) has criticized this classification partly because it postulates the class of agent-evocator verbs as an ad hoc category merely introduced to satisfy the observed data patterns. He also observed that the two classes of psych verbs do not behave as uniformly as would be expected. Instead, more distinctions are required to describe their bias behavior adequately. More generally, he points out that it has to be explained what constitutes the connection between thematic properties and causal dependencies (see also the discussion in Garnham 2001; Crinean & Garnham 2006; Pickering & Majid 2007). A recent proposal in this direction has been made by Hartshorne & Snedeker (2012) grounding IC in the semantics of VerbNet (Kipper-Schuler 2006).

Complicating matters, albeit undoubtedly very important, verb semantics is only one contributor to IC bias. It has been shown that perceived causality also takes into account the semantic connotations of the nouns denoting the participants (Garvey & Caramazza 1974; Corrigan 2001). Corrigan (2001) provided evidence that some nouns are rated as more agentive than others (e.g. *mugger* vs. *passerby*) and are thus more likely to be perceived as instigators of events. Similarly, Ferstl et al. (2011) demonstrated that even the gender information of proper names (*John fascinates Mary* vs. *Mary fascinates John*) affects IC bias. These observations have led some researchers to the conclusion that IC is not a genuinely linguistic phenomenon but primarily reflects encyclopedic knowledge about specific events in a given social or cultural setting.

On the other hand, it has often been claimed that IC constitutes a cognitive universal, that is the same verb will elicit the same bias across languages (e.g. Rudolph & Försterling 1997; Goikoetxea et al. 2008; Ferstl et al. 2011). Up to date, however, only one study systematically investigated IC cross-linguistically using the same set of verbs and employing the same methods across languages. Hartshorne et al. (2013) elicited IC biases in English, Japanese, Mandarin and Russian emotion verbs and found them to be highly correlated across languages and cultures.

Besides the verbs and their arguments, the connective is a crucial determinant for IC bias, too (Ehrlich 1980; Stevenson et al. 1994, 2000). Testing connectives other than *because* such as *and*, *but*, *so* as well as full stop yielded continuations with different anaphoric preferences. Extending this line of research, Kehler et al. (2008) have provided evidence that IC bias hinges on the discourse relation of explanation. They demonstrated that the phenomenon is stable across different kinds of discourse connectives (explicit explanations with *because* vs. implicit explanations after a full stop), once productions are conditioned on the explanation relation.

The coreferential preferences due to IC have been shown to elicit effects in on-line comprehension tasks. Caramazza et al. (1977) observed longer response times when judging coreference in incongruent continuations than in congruent ones. This

congruency effect has been established in coreference decision tasks, probe recognition tasks and reading time measures (Garnham & Oakhill 1985; Vonk 1985; Garnham et al. 1992; McKoon et al. 1993; Carreiras et al. 1996). An open question is whether IC has an early focusing effect (e.g. Greene & McKoon 1995; McDonald & MacWhinney 1995; Long & Ley 2000; Koornneef & van Berkum 2006; van Berkum et al. 2007; Pyykkönen & Järvisikivi 2010), i.e. whether the congruence effect shows up right at the pronoun, or constitutes a later effect on clausal integration (e.g. Garnham et al. 1996; Stewart et al. 2000; Guerry et al. 2006). Recent studies show that IC verbs yield early congruency effects immediately at the pronoun. This has been demonstrated in eyetracking studies during reading by Koornneef & van Berkum (2006) and Featherstone & Sturt (2010) and in evoked potentials (ERPs) when encountering the pronoun (van Berkum et al. 2007). What is more, the visual world eyetracking studies by Cozijn et al. (2011) and Pyykkönen & Järvisikivi (2010) provide evidence for the activation of IC information even before participants encounter the causal connective. What is clearly needed then is an account that allows us to derive predictions about the anticipation of an explanation that is transparently connected with the lexical semantics of the verb. These predictions have to be calculated without putting too much computational burden on the comprehender (e.g. not relying on inferences in the style of Hobbs 1979) since they seem to arise very fast and automatic.

Obviously, besides explaining early activation of the congruent referent something has to be said about how a continuation that overrides the bias eventually can get integrated. This is essential because end-of-sentence acceptability ratings show that participants are generally able to interpret incongruent continuations. On our view, focusing versus integration is thus not an either–or question and both aspects have to be explained in order to properly understand the nature of the bias.

To conclude, the reviewed studies leave us with the following desiderata a linguistic theory of implicit verb causality will have to account for. First of all, the associations between certain argument roles and preferences for anaphoric links have to be explained. In particular, it should follow from the theory what makes psych verbs and the “class” of ‘agent-evocator’ verbs semantically special and what causes their affinity for explanations. Second, a theory of IC has to explain the link between the lexical meaning of a given verb, the particular realization of its arguments and the semantics/pragmatics of discourse. Third, it should provide us with a semantics of the explanation relation and the constraints it imposes on the entities it relates. Finally, discourse expectations should be derivable in a fast and automatic fashion to be able to account for focusing effects. It is important to stress that all of these points constitute cognitive universals. We would therefore expect to see little or no variation across languages. That this is indeed the case will be shown in the comparative production study in Sect. 4 and 5.

3 A Semantic Account of Implicit Causality

In this section, we present a novel, semantic-compositional account of IC. We will show that IC bias is based on well-established semantic properties of the verbs and the arguments they combine with. What is more, we argue that we can achieve a better understanding of this phenomenon by studying in detail the semantic and pragmatic properties of the discourse relation of explanation.

Briefly recapitulating the verbs involved, we first need to explain why the stimulus argument of *s-e* (5a) and *e-s* (5b) psych verbs, and the ‘evocator’ argument of ‘agent-evocator’ verbs (6) attract the bias.

(5) **Psych verbs:**

- a. *s-e*: Mary fascinated John because *she* danced very well.
- b. *e-s*: Mary admired John because *he* sang beautifully.

(6) **‘Agent-evocator’ verbs:**

John congratulated Mary because *she* won the race.

It should also follow from the analysis why it is that interpersonal verbs belonging to other verb classes (e.g. *a-p* verbs such as *hit* or *kill*) show no consistent bias pattern. Finally, we want to account for the findings of Kehler et al. (2008) that IC verbs trigger an over-proportional amount of explanations. It may be noted that the sets of verbs included in our analysis will not differ radically from those identified in earlier research. Importantly, though, our proposal offers a different take on *why* it is that these verbs share the property of being IC verbs.

Our account may be summarized as follows: What makes IC verbs special is that they *trigger expectations or preferences for specific explanations*. They do so because they are *underspecified* with respect to causal content of a specific type which is crucial to the conceptualization of the situation described by the verb. Put differently, IC verbs carry an empty “slot” for specific explanatory content. Importantly, this slot is intimately *linked to one of the two arguments*. The IC bias follows from a general processing preference for not leaving “missing content” unspecified, i.e. a tendency to avoid accommodation (Altmann & Steedman 1988; van Berkum et al. 1999; van der Sandt 1992; Zeevat 2000). We will show that there is a strong relation between the specific type of explanation that a verb triggers and the ratio of anaphoric reference to the subject or object, i.e. IC bias. We will also see that there is a role to play for world-knowledge or more pragmatic factors. Crucially however, *the bias always reflects the preference for specific explanation patterns*.

Our proposal allows for a more precise definition of what the nature of an IC verb is. We go beyond a mere correlation between argument structure classes and bias patterns and show *why* exactly the verb classes displaying IC do so. We also contend that it is a central aspect of our account that it allows for novel and more precise predictions for discourse processing involving IC verbs.

This section is organized as follows. After a presentation of explanation types relevant to IC (Sect. 3.1), we present a semantic analysis of the crucial factors determining IC bias (Sect. 3.2–3.4) along with the predictions for the study.

3.1 Simple Causes, External and Internal Reasons

A necessary prerequisite for accounting for IC in terms of preferences for specifying “missing explanations” is an elaborate typology of explanations. Based on our own previous work (Solstad 2010), which was developed independently of analyses of IC bias, we distinguish three kinds of explanation relations expressible by means of *because*: (i) simple causes (7a), (ii) externally anchored reasons (7b) and (iii) internally anchored reasons (7c):

- (7) a. **simple cause:**
 John disturbed *Mary* because *he* was making lots of noise.
 b. **externally anchored reason:**
 John disturbed *Mary* because *she* had damaged his bike.
 c. **internally anchored reason:**
 John disturbed *Mary* because *he* was very angry at her.

Simple (direct) causes may be causes of events, states or attitudinal states. Limiting ourselves to interpersonal causal relations for the present purposes, simple causes never involve volition or agentivity on the side of the causing individual. Thus, in (7a), John is not taken to make noise intentionally in order to disturb Mary, the disturbance is rather understood to be a by-product, as it were, of his noise-making. Although *disturb* may be interpreted both as an *s-e* verb, lacking intentionality, as well as an intentional *a-p* verb, simple causes only go with the former variant. A number of verbs display an ambiguity of this kind, with both *a-p* and *s-e* interpretations (e.g. *annoy*, *frighten*, and *scare*).

The two kinds of *reasons* are defined as causes of attitudinal states (involving a *reasoning* agent, cf. Solstad 2010). Thus, in (7b) and (7c) the *because* clauses specify causes of John’s intention to disturb Mary. Since an intention is involved, *disturb* must be interpreted as an *a-p* verb. Externally and internally anchored reasons (henceforth, external and internal reasons, respectively) differ with respect to the nature of the cause for the agent’s intention. External reasons are causes which are external to the attitude-bearer’s, i.e. the agent’s, mind. In (7b), this is the event of Mary stealing John’s bike (which happened prior to John developing his intention to disturb Mary). Internal reasons, on the other hand, designate causes internal to the attitude-bearer’s mind. Internal reasons are mostly attitudes or mental states themselves. Thus, in (7c), what causes John to develop the intention to disturb Mary, is his own attitude of anger towards her. It should be noted that providing an internal reason does not exclude the existence of an external reason. In (7c), John’s anger may result from Mary having stolen his bike. However, in (7b) and (7c), the two *because* clauses focus on two different kinds of causes for intentional action, respectively.

It should be added that the above typology is applicable also outside of the domain of IC. However, in order to avoid unnecessary complexity, it is limited to the categories which are needed to handle the cases of relevance to the current study. For instance, one may additionally assume a criterial or “background” cause, which may

also occur with IC verbs (cf. Sect. 4). Such causes provide background information which make possible, or ‘facilitate’ the situation described by the verb. Compare the two sentences in (8):

- (8) a. Felix frightened Vanessa because he suddenly screamed.
 b. Felix frightened Vanessa because she didn’t hear him coming.

Whereas the *because* clause in (8a) specifies the simple, direct cause of Vanessa’s state of being frightened, the *because* clause in (8b) specifies the background for Vanessa being frightened. In this case, it isn’t specified what it was that Felix did which caused Vanessa to be frightened, but it may be assumed that she would not have been frightened if she had been aware that he was coming, no matter what he actually did. It may also be added that *because* has an ‘epistemic’ reading (see for instance Solstad 2010), which will not be discussed here.

The interdependency between coreference and explanation patterns may be illustrated by the examples in (7b) and (7c). In the case of the internal and external reasons, which involve a sentient attitude bearer, we can see that in the case of the internal reason, primary reference is made to the subject argument (it is mentioned first), whereas the external reason makes primary reference to the object argument. Put differently, internal reasons contribute to the NP1 proportion of the bias in *a-p* verbs, whereas external reasons contribute to the NP2 proportion of the bias. Simple causes may contribute to either NP1 or NP2 bias, but only with psych verbs which do not involve any intentionality. Importantly, there is a strong association between the kind of explanation which is given (simple cause, internal or external reason) and the referent which this explanation is attributed to. In the case of the external reason, for instance, an external event is taken to explain John’s behavior, whereas in the case of the internal reason, an attitude is taken to constitute an explanation of his behavior.

We will return to this typology when discussing the semantics of particular verb classes, but for now it may be noted that not all explanation types are available for all verb classes, as witnessed by the difference between *s-e* and *a-p* interpretations of *disturb*. Thus, simple causes will not be a relevant category for *a-p* verbs, whereas (external) reasons are not relevant for psych verbs.

An additional aspect of importance to our theory is the fact that explanations which are provided in the production studies take the form of *because* clauses or (independent) sentences following a full stop. It may be assumed that *because* and corresponding ‘full stop’ explanations causally relate entities propositional in nature (Solstad 2010). In accordance with this, Solstad (2010) proposes that a *because* clause introduces factive causes. Hence, we are not interested in any kind of under-specified causal content in verbs, but rather such content which is ‘sentential’, or rather propositional, in nature (Solstad 2010). Put differently, the missing content must be specifiable by means of a *because* clause for the bias to show up. This will become clear in the discussion of *s-e* vs. *a-p* verbs in Sect. 3.2.1.

With these notions in mind, we can now turn to the question of why it is that the bias is displayed consistently by *s-e* and *e-s* psych verbs and ‘agent-evocator’ verbs, but not by *a-p* verbs? Are there any semantically well-defined properties of

stimulus and ‘evocator’ arguments which predestine them to attract the bias when explanations are provided by *because* clauses or independent sentences?

3.2 Semantic Verb Properties Determining the Bias

We have identified two semantic factors which determine the bias, both involving semantic features well-known from semantics. One involves the stimulus argument of *s-e* and *e-s* verbs, the other subsumes (but not only) ‘evocator’ arguments of the so-called ‘agent-evocator’ verbs. In what follows, we will show to which extent these arguments may be viewed as (underlyingly) propositional in nature, making them suitable for specification by means of *because* clauses.

3.2.1 Stimulus Arguments of Psych Verbs

It has been observed repeatedly in the literature that stimulus arguments, which occur with psych verbs, are strong bias attractors, whereas experiencer arguments are not. Thus, *s-e* verbs such as *annoy*, *fascinate*, or *disturb* display a strong NP1 bias, whereas *e-s* verbs such as *hate*, *pity*, or *admire* have a strong NP2 bias:

- (9) a. *Peter* annoyed Mary because *he* sang loudly.
 b. Peter admired *Mary* because *she* sang beautifully.

However, as far as we are concerned, there has been no explanation as to *why* stimulus arguments – as opposed to experiencer arguments – should attract the bias. On our account, there is an obvious two-fold semantic connection between the explanation possibilities offered by psych verbs and *because* clauses: First, stimulus arguments are causes of experiences, i.e. there is a causal relation between a stimulus and an experiencer (cf. e.g. Pesetsky 1995). Second, the stimulus argument may be argued to be propositional in nature. To see this, consider (10), where the stimulus argument of *annoy* is realized as a proper name:

- (10) Peter annoyed Mary.

Whenever we make an utterance as in (10) out of context, we leave out an important piece of information, namely what exact action or property of Peter it was that caused an experience of annoyance in Mary. Thus, in (10), the proper noun occurring as the subject, the stimulus argument, is merely a placeholder, as it were, for something which is semantically more complex. We will take this to be an entity which is propositional in nature. It follows from this view that what we do when modifying a matrix clause with a psych verb such as (10) by means of a *because* clause, is to specify exactly which unspecified property of Peter it was that caused annoyance in Mary, cf. (9a).

One piece of evidence in favor of this propositional analysis of stimulus arguments is that for both *s-e* verbs and *e-s* verbs the stimulus arguments alternate with *that* clauses, which are indisputably propositional in nature:

(11) ***s-e* verb:**

- a. *Peter* annoyed Mary (because he sang loudly).
- b. It annoyed Mary *that Peter sang loudly*.

(12) ***e-s* verb:**

- a. Mary hated *Peter* (because he sang loudly).
- b. Mary hated it *that Peter sang loudly*.

The semantic representations in (13) for example (9a) accommodate the above considerations:

The representations in (13) are in the format of Discourse Representation Structures (DRS; Kamp et al. 2011). In the leftmost DRS (A.), we represent semantic aspects of *annoy* relevant to our discussion. This representation may be seen as a simplified lexical entry for *annoy*. In prose, it states that y is in a mental state s_{att} of annoyance and that this state is caused by some fact f which is propositional in nature. Importantly, this fact is not specified if the stimulus argument slot is realized by a noun phrase which denotes an individual, cf. the representation in B. for example (10). This representation reveals nothing more than that the ‘stimulus individual’ is a participant (PART) in some event e (other ontological sorts, such as e.g. states, are possible) which is still to be specified. Finally, the rightmost DRS in C. shows a representation which corresponds to the interpretation for a sentence such as *Peter annoys Mary because he sings loudly* or *It annoys Mary that Peter sings loudly*.

A similar analysis may be applied for *e-s* verbs, although there are some differences between the two classes (Jackendoff 1990; Landau 2010; Pesetsky 1995). In both cases, the stimulus is propositional in nature and causally related to the mental state of the experiencer. It may also be noted that in both cases, the experiencer may only be realized as a noun phrase denoting a sentient being.

Predictions for *s-e* and *e-s* Psych Verbs In terms of the typology of explanations introduced in Sect. 3.1, psych verbs trigger *simple causes* as specifications of the stimulus argument since they encode a propositional causal relation between the causing stimulus and a caused experience. Thus, *s-e* verbs should display not only a strong proportion of NP1-continuations, but these continuations should also be of *simple cause* type. *e-s* verbs, on the other hand, should display a strong proportion of NP2-continuations, but still trigger simple causes as the preferred explanation. In both cases, internal and external reasons should be strongly dispreferred. Variation in the strength of IC bias should be reflected in the proportion of simple causes vs. explanations of other categories. We also predict ambiguous *s-e/a-p* verbs, which allow for external reasons on their *a-p* interpretation, to exhibit no stronger NP1 bias than unambiguous *s-e* verbs.

3.2.2 Presuppositions of Agent-patient Verbs

The second semantic feature which represents a case of content missing in the verb triggering specific explanations, and consequently specific bias patterns, involves the presence of a presupposition suitable to explain the action denoted by an *a-p* verb. In general, a presupposition is a defeasible inference which may be viewed as a prerequisite, an implicit assumption, for (the occurrence of) a situation. Consider the examples involving *punish* in (14):

- (14) a. John punished Mary.
b. John punished Mary because she stole the money.

Intuitively, if someone punishes someone else, there must be some reason for it, at least in the eyes of the punisher. Typically, this is something the patient (object) argument of *punish* did, which occurred prior to the act of punishing. If we take something away from someone, it may only be considered an act of punishment if there occurred an act worthy of punishment prior to this. Otherwise, the act of taking something from someone would only be an act of stealing or cruelty. It is exactly this ‘occasioning state of affairs’, that is the presupposition carried by the verb *punish*. If *punish* occurs only with its two obligatory arguments (14a), this presupposition needs to be accommodated, i.e. crucial information is lacking. Importantly, the presupposition may be made explicit (or justified, cf. Kamp et al. 2011), by means of a *because* clause, given that it constitutes an external reason which temporally precedes the act of punishment, cf. (14b). To see that we are actually dealing with a presupposition, it may be observed that the inference in (14a) survives negation: *John didn’t punish Mary*. In this case, the assumption that Mary did something which John could have considered to be worthy of punishment, may still be made.

Other verbs belonging to this group include e.g. *thank*, *criticize*, *praise* and *reward*. In the IC literature, these verbs have been assumed to belong to the somewhat ad hoc class of ‘agent-evocator’ verbs, a class of verbs not generally adopted in linguistic research on verb semantics. We avoid reference to this class for two reasons: First of all, as opposed to the stimulus argument of psych verbs, the argument itself

is not underspecified in this case. Rather, it is a (underspecified) presupposition *associated with the argument in question* which triggers the causal explanation. Thus, it does not only involve argument structure as such. What is more, resorting to presuppositions allows us to account for the bias of verbs which one has so far not been able to account for. Consider *apologize*. Whereas the classic ‘agent-evocator’ verbs such as *punish* display NP2 bias, *apologize* is an NP1-biased verb:

(15) *John apologized to Mary because he had behaved rudely.*

Apologize is no *s-e* verb and thus seems to belong to the class of ‘other’ *a-p* verbs which do not display a consistent bias pattern overall. However, once we take our analysis of presuppositional *a-p* verbs into account, the bias behavior of *apologize* becomes completely predictable: *Apologize* carries a presupposition (the reader may verify this by applying the negation test, see above), too, but this time, the presupposition is associated with the *subject argument*. There must be an occasion for someone to apologize to someone. In this case, however, what is presupposed is that the agent of *apologize* engaged in an event which, reflexively, as it were, was the external reason for the act of apologizing.

The DRS in (16) shows a simplified semantic representation for *punish*:

$$(16) \quad \left\langle \left\{ \begin{array}{c} \overline{e' \ z} \\ e' <_t e \\ e'=? \\ \text{PART}(e')=z \\ z=y \end{array} \right\}, \begin{array}{|c|} \hline e \ x \ y \\ \hline e: \text{punish}(x,y) \\ \text{AGENT}(e)=x \\ \text{PATIENT}(e)=y \\ \hline \end{array} \right\rangle$$

In prose: There is an event *e* which is an event of *x* punishing *y* with the presupposition that there was an event *e'* which temporally preceded *e* and in which one participant was identical to the patient argument of the punishing act. As in the case of the stimulus argument of psych verbs, we are dealing with unspecified content: We know that the patient of the punishment was involved in an event which occurred prior to the act of punishment, but we do not know which event it was (and consequently not which role the patient of the punishment act played in it).

Finally, it should be added that we do not expect all presuppositions associated with a participant to attract the bias. A presupposed event must also be a *suitable explanation or cause* for the event described by the predicate in the matrix clause. Consider *forgive*. Similarly to *apologize*, it carries a presupposition that the object did something which was perceived of as negative by the subject, which could lead one to expect *forgive* to display NP2 bias. However, this presupposed event does not constitute a reasonable cause for the act of forgiving (# signals pragmatic anomaly):

(17) #Mary forgave John because he cheated on her.

In Sect. 5, we will present results substantiating the above claims concerning *punish* (and other NP2-biased verbs), *apologize*, and *forgive*.

Predictions for *a-p* Verbs with Presuppositions In terms of the causal typology, these verbs should trigger external reasons. For most verbs (*punish*, *congratulate*, *thank*, or *criticize*), they should be linked to NP2. However, they may also be associated with NP1 (e.g. *apologize*). Obviously, internal reasons are not excluded, but they should be clearly dispreferred because they still leave the underspecified content unspecified. Simple causes should not be possible due to ontological restrictions of this explanation relation. Differences in the strength of IC bias should be reflected in differences in the proportion between external and internal reasons.

In this section, we observed two main semantic factors which constitute instances of “missing content” which may be specified by *because* clauses: 1. Underspecified propositions associated with stimulus arguments of psych verbs, and 2. certain presuppositions of *a-p* verbs. Both factors involve underspecification in the sense that if the associated argument is realized by a proper noun, information necessary for the conceptualization of the situation is missing. For Peter to be annoying, there must be a specific property or action associated with him which is annoying, and if Mary is to be punished, there must be an occasion for it, i.e. some fact or action associated with her which occurred prior to the punishing act. Importantly, these factors are semantic constraints encoded in the verb and related to its arguments. We view the influence of both factors as reflecting a general processing constraint against leaving underspecified content unspecified. The specification of this content may thus be seen as an instance of a general principle to avoid accommodation (see above). Consequently, these verbs predominantly trigger explanations, as we will see.

3.3 On the IC Bias of Other Verb Classes

Finally, we address the issue of why other verb classes do not display a consistent bias pattern. We will focus on the class of *a-p* verbs without a presupposition. Why for instance, do continuations of the kind in (18) occur equally often for *hit*?

- (18) a. Mary hit *John* because ... *he* provoked her.
 b. Mary hit John because ... *she* was very angry.

As for the presuppositional *a-p* verbs, explanations of simple cause type should be highly dispreferred for *a-p* verbs such as *hit* in general since they always involve a dimension of intentionality or volitionality. This leaves us with only internal and external reasons as possible explanation types, contributing to the proportion of NP1 and NP2 bias, respectively. It may be noted that the continuation in (18a) constitutes an external reason, whereas the continuation in (18b) constitutes an internal reason.

It follows from our theory that as long as the argument of a verb is not underlyingly propositional in nature or, alternatively, associated with a presupposition, there is no missing content of relevance to IC bias which could trigger a specific bias pattern. Thus, on the face of it, our theory predicts all *a-p* verbs to display a balanced bias (unless there should exist independent factors favoring e.g. internal

over external reasons). Obviously, this is not the case as witnessed for instance with the strongly NP1-biased *a-p* verb *telephone* (cf. e.g. Stewart et al. 2000). This is an area where world knowledge, i.e. knowledge about stereotypical aspects of events, comes into play. Who is responsible for an action, who is taking an active part in it? For instance, such verbs may involve special intentional profiles as to the goal of the action, role hierarchies or gender stereotypes. Importantly, though, if we are correct that IC bias reflects specific explanation strategies, a clear bias pattern for (non-presuppositional) *a-p* verbs should correspond to clear preferences for specific ways of explaining the action or state of affairs described by the verb. Given that only internal and external reasons are possible for e.g. NP1-biased *telephone*, we expect this verb to preferably trigger NP1-associated internal reasons, instead of NP2-associated external reasons (cf. Sect. 3.1).

Predictions for (Non-presuppositional) *a-p* Verbs We predicted the class of *a-p* verbs to display a balanced bias overall. Clear bias patterns of single verbs should be reflected in specific explanation patterns. Verbs in this class with a strong NP1 bias should trigger internal reasons, whereas verbs with a strong NP2 bias should trigger external reasons as the preferred explanation type. Conversely, *a-p* verbs with a balanced explanation profile (i.e. the number of external reasons equals the number of internal reasons) were expected to be unbiased. Put differently, the bias of any *a-p* verb should be fully predictable by its ratio of external and internal reasons.

3.4 General Considerations and Predictions for the Relation Between IC Bias and Preferences for Explanations

In this section, we have presented a semantic theory of IC, contending that IC reflects a general preference for not leaving missing causal content unspecified. We also claimed that IC bias is but an epiphenomenon of preferences for *specific explanations* which are triggered by semantic properties of the verb and its arguments. We identified two main factors triggering IC: i) Underspecified, propositional stimulus arguments of psych verbs, and ii) yet-to-be-justified presuppositions associated with one of the arguments of *a-p* verbs. IC bias shows up in *because* clauses and independent sentences since they may both be used to specify these underspecified properties. From these general considerations we may derive general predictions for verbs displaying either a pronounced or a balanced IC bias.

Predictions w.r.t. Proportions of ‘Full Stop Explanations’ According to our analysis, the stimulus argument in both *s-e* and *e-s* psych verbs serves as a placeholder for some unspecified property which can be elaborated on in an explanation. Similarly, unexpressed presupposed causal content in presuppositional *a-p* verbs can be made explicit in an explanation. All three kinds of verbs should thus trigger a higher-than-average proportion of explanations after a full stop. We expected explanations to be the default for prompts with these verbs. By contrast, *a-p* verbs without

a presupposition do not lack causal content. We thus expected less explanations in these cases than for the three other classes.

Predictions Concerning Explanation Types The classes discussed above clearly differ w.r.t. the explanation types they allow for. Agentive verbs involve a subject with an intention to act and hence should be constrained to external or internal reasons. This is different for psych-verbs where a stimulus argument is not in (volitional) control of evoking an experience. In the latter case, we should thus only observe simple causes or explanations of background type (see Sect. 3.1). The only verbs allowing for all types of explanations are ambiguous verbs like *disturb* which can have both an *a-p* as well as an *s-e* interpretation.

Bias Predictions for *because* and ‘Full Stop’ Continuations Since we take IC bias to depend on the explanation profile of a particular verb, we followed Kehler et al. (2008) in predicting that, as long as only explanatory ‘full stop’ continuations are included, we should expect verbs to display an identical bias in *because* and ‘full stop’ conditions.

Predictions w.r.t. Cross-linguistic Comparison Returning finally to the cross-linguistic comparison between German and Norwegian, the semantics of the selected verbs and of *weil/fordi* (‘because’) should be the same in German and Norwegian. Thus, we don’t expect any differences in explicit explanations prompted by *weil/fordi*. At the most, the two languages should exhibit an across-the-board difference in their proportions of explanations after a full stop, but the relative differences between synonymous verbs should be stable across languages.

Differences could show up with non-marked explanations, though. As outlined in Sect. 1, translation studies showed that Norwegian discourse displays a paratactic structure with no marking whereas German discourse is more hypotactic with explicit marking of discourse relations. This led us to expect the proportion of implicit explanations to be higher in Norwegian than in German. The predicted difference should be most pronounced for “ordinary” *a-p* verbs without a presupposition. The need for further causal elaboration in the semantically “deficitary” cases (*e-s*, *s-e* and presuppositional *a-p* verbs) may however lead to a uniform pattern in both languages, although this implies that German native speakers realize the explanation in an independent sentence which they would try to avoid under normal circumstances.

4 Methods

To test our predictions we conducted a comparative discourse continuation study in which Germans and Norwegians provided a total of 10,100 discourse continuations. We prompted continuations for a sample of 101 verbs in both languages and annotated the resulting German/Norwegian corpus of discourse continuations with respect to anaphoric relations, discourse relations and the proposed typology of explanation types. We should mention that, originally, the study included 102 verbs,

but that one verb was left out of the analysis because after closer inspection it turned out to be used only marginally as an interpersonal verb in one of the languages.

4.1 Selection of Verbs

We selected 101 verbs which have very close lexical meaning(s) in German and Norwegian. See the Appendix for the complete set of verbs. These comprised 53 unambiguous *a-p* verbs like *kill*, (*a*)*wake* or *congratulate* (Tables 6 and 7), 14 verbs allowing both an *a-p* and a *s-e* reading like *disturb*, *hurt* or *provoke* (Table 5), 16 unambiguous *s-e* verbs like *fascinate* or *inspire* (Table 3), and 18 *e-s* verbs, out of which 17 were unambiguous like *love* or *hate* (Table 4A) whereas one allowed an *a-p* interpretation, *übersehen/overse* ('overlook'; Table 4B). The unambiguous *a-p* verbs were further divided: 10 of them presupposed a preceding event, e.g. *praise* and *apologize to* (Table 6), whereas the rest (43 verbs) did not (Table 7).

We applied a number of tests to classify the verbs into these classes. To test for agenthood of the external argument we modified the transitive verb frames (*Mary VERB-ed Peter*) with the adverbial *absichtlich/med vilje* ('deliberately'). Verbs that passed this test were annotated to allow for agentive subjects. Furthermore, since stimulus arguments that are realized as proper names can be replaced by a *that* clause (Sect. 3.2.1), we used a *that* clause replacement test (see Sect. 3.2.1) to assert that NP1 or NP2 can serve as placeholders for a property of NP1 or NP2, respectively. The presupposition verbs were classified on the basis of the standard presupposition tests (e.g. embedding under negation, cf. Sect. 3.2.2).

4.2 Design of the Study

The experiments both employed a 2x2 factorial design with the two within factors *explicit/implicit marking* (two levels: *because* vs. *full stop*) and *gender* (two levels: *NP1: masc./NP2: fem.* vs. *NP1: fem./NP2: masc.*). The latter factor was manipulated to investigate (and control for) potential influences of gender information. (19) shows a sample item in all four conditions.

- (19) a. Victoria faszinierte Steven. ...
 Marie faszinierte Fredrik. ...
 '*NP1_{fem.} fascinated NP2_{masc.}*'
 b. Steven faszinierte Victoria. ...
 Fredrik faszinierte Marie. ...
 '*NP1_{masc.} fascinated NP2_{fem.}*'
 c. Victoria faszinierte Steven, weil ...
 Marie faszinierte Fredrik fordi ...
 '*NP1_{fem.} fascinated NP2_{masc.} because*'

- d. Steven faszinierte Victoria, weil ...
 Fredrik fascinerte Marie fordi ...
 ‘*NP1_{masc.} fascinated NP2_{fem.} because*’

4.3 Participants and Procedure

52 native Germans (mean age 25.6 years; 44 female) and 48 native Norwegians (mean age 27.0; 34 female) participated in the study for monetary compensation. The procedure was exactly the same in the German and the Norwegian version of the experiment. Participants were instructed in their native language.

In each trial, a text prompt was displayed to the participants for which they had to provide the first continuation that came to their mind. In case they felt this was impossible they were instructed to press a “not possible” button. This option was chosen only 1% of the time and less than 10% for each individual verb, suggesting that all verbs were well-known. The “not possible” cases were treated as missing values in the analysis.

An experimental session asked participants to continue ‘mini-stories’ freely, yet coherently, and was divided into two blocks, each with its own set of instructions. In the first block, participants were asked to continue discourses consisting of one sentence that always ended in a full stop (e.g. *John fascinated Mary. ...*) with a single sentence. In this first block, participants never encountered a discourse marker. This was done to avoid any bias towards a particular discourse relation. In the second block, participants were given a second set of instructions asking them to complete *because* clauses (e.g. *John fascinated Mary because ...*). The second block consisted exclusively of *because* sentences. A latin square design with four lists was used to ensure that each participant saw each verb only once. An average session took about 45 minutes.

4.4 Corpus Annotation

We excluded all trials with “not possible” button presses and nonsensical or ungrammatical continuations. This affected 2.7% of the data. The resulting corpus of discourse continuations was annotated by the authors with respect to 1) IC bias, 2) discourse relations and 3) explanation types. For each part of the annotation we determined the inter-rater agreement. Agreement was determined in a random sample of 300 continuations which were independently annotated by both authors.

IC Bias We annotated whether the continuation contained an anaphor and which argument served as its antecedent. With respect to the occurrence of an anaphor, we distinguished the following cases: 1) The continuation contained no anaphoric expression, cf. (20a), or a reciprocal expression, cf. (20b); it was then coded as a missing value, 2) it contained a single anaphoric expression, cf. (20c) or 3) there

were multiple anaphoric elements, cf. (20d). The latter cases behaved exactly as the unambiguous cases w.r.t. IC bias. Therefore, we did not analyze them separately. IC bias was then coded as the referent (subject vs. object) of the first anaphoric expression in the continuation sentence (for convenience, all examples are in English):

- (20) a. Linda admired Peter because empathy is what you wouldn't expect in such a case.
 b. Mary loved Frank. **They** understood **each other** without words.
 c. Peter admired Linda because **this girl** always knew what to do.
 d. Linda admired Peter because **Peter** had read a beautiful poem to **her**.

We also annotated whether the participant had interpreted the prompt with a 'subject before object' or an 'object before subject' word order. This is important because German (in contrast to English or Norwegian) displays a relatively free word order and proper names are not marked for case information in any of the languages. In the 'object before subject' word order, the NP1/NP2 bias annotation must be reversed to make it compatible with the constant 'subject before object' order in Norwegian (and English). A sentence was annotated as 'object before subject' in case the continuation only made sense with this word order. In these cases (0.6%) the bias was adjusted accordingly. Taken together, for this part of the annotation scheme the inter-rater agreement was very high ($\kappa = .93$).

Discourse Relations For the purposes of the present paper, we only wanted to distinguish continuations that were explanations from continuations that instantiated some other relation. Therefore, we limited ourselves to a partial annotation of discourse relations. As a coarse filter we made use of the fact that causes temporally precede/overlap with their effects. In a first step we annotated whether the production described an eventuality that followed or preceded/overlapped with the prompted eventuality. The latter cases were subjected to an insertion test. Those productions that allowed the insertion of *weil/fordi* ('because') without a change in meaning were counted as explanations (this was also the method applied in Kehler et al. 2008, Hannah Rohde, p.c.). Again, the inter-rater agreement was rather high ($\kappa = .83$). (21) provides sample discourses and their discourse annotation, respectively.

- (21) a. *Temporal overlap/precedence: NO; Explanation: NO*
 John killed Julie. He didn't show any sign of regret.
 b. *Temporal overlap/precedence: YES; Explanation: NO*
 John killed Julie. He stabbed her in the back.
 c. *Temporal overlap/precedence: YES; Explanation: YES*
 Julie killed John. He had abused her when she was a teenager.
 d. *Temporal overlap/precedence: YES; Explanation: YES*
 Martin frightened Mary. He was talking to himself all the time.

Causal Annotation Explanations were annotated according to the causal typology described in Sect. 3.1. They were classified as simple causes, external reasons or internal reasons, respectively. A fourth category were explanations of background

or criterial type, cf. (8b) in Sect. 3.1. However, since they will not play a role in the analyses of the next section, we will treat these explanations as missing values, if not otherwise mentioned. In total, causal backgrounds accounted for less than 10% of the explanations. The inter-rater reliability for this part of the annotation scheme was satisfactorily high ($\kappa = .80$), too.

4.5 Data Analysis

For statistical analysis, we computed logit mixed effects models (cf. Jäger 2008) on 1. proportions of explanations and elaborations, 2. the proportion of causal categories of the proposed causal typology and 3. proportions of NP1 vs. NP2 continuations. The fixed effects of *language*, *connective* (*explicit relation: weil/fordi* vs. *implicit relation: full stop*), *verb class* (*a-p* vs. *e-s* vs. *s-e*), *presupposition* (*presupposition verb* vs. *verb without a presupposition*) and their interactions served as predictors. Note that these factors depend on each other (i.e. all presupposition verbs were *a-p* verbs). We therefore included the predictors not in a single model but computed multiple models to test the particular predictions always including the highest interaction of fixed effects. Into each of these models the random intercepts and random slopes of participants and items were included. The random slopes included the highest interaction of the within factors (participant- and/or item-wise). *Verb class* was thus left out of the random effects structure by items.

5 Results and Discussion

In the following presentation of results, we follow the predictions laid out in Sect. 3.

5.1 Cross-linguistic Uniformity

Germans and Norwegians did not differ in how often they continued with an explanation after a full stop. Across the board, German participants continued with an explanation 50.8% of the time, whereas Norwegians produced numerically slightly less explanations (50.0%). A χ^2 -test revealed that the difference was not significant ($\chi^2(1)=.33$; $p=.56$). This was contrary to our original expectations. Based on comparative translation studies, we predicted a higher proportion of explanations in Norwegian than in German. However, the need for specification of yet unspecified causal content may be sufficiently strong for Germans to adopt a discourse structure they would otherwise disprefer.

5.1.1 Proportion of Explanations After a Full Stop

We expected the proportion of explanations to be higher after prompts with an underspecified argument (stimulus of *e-s* or *s-e* verbs) or with a presupposition verb than after prompts with a regular *a-p* verb. The findings fully supported this prediction. While ordinary *a-p* verbs had a mean proportion of 39.1% explanations, presuppositional *a-p* verbs had 61.5% explanations and *e-s* and *s-e* verbs had 63.2% and 67.6% explanations, respectively. On average, the proportion of explanations was 64.3% in the latter three verb classes and was thus 25.2% higher than in the ordinary *a-p* verbs. The patterns were alike in German (64.5% vs. 39.7%) and in Norwegian (64.1% vs. 38.6%). Accordingly, a logit mixed effects model including the fixed effects of *language*, *verb class* (*e-s/s-e/presupposition verbs* vs. *ordinary a-p verbs*) and their interactions revealed a significant effect of *verb class* (estimate=1.18; $z=5.6$; $p<.01$), but neither a reliable main effect of *language* ($z<1$) nor a reliable interaction of *language* and *verb class* ($z<1$).

5.1.2 Explanation Profiles

We expected specific explanation profiles for different types of verbs. These explanation profiles were predicted to be stable across languages.

Psych Verbs We predicted *s-e* and *e-s* verbs with animate arguments to require the specification of some property of the stimulus argument. Since the elaboration of the stimulus is equivalent to providing a simple cause, simple causes should be preferred over internal and external reasons. This is exactly what we found. *e-s* verbs had 87.9% simple causes (78.7% specifying a property of NP2 and 9.2% of the background type), whereas *s-e* verbs gave rise to 74.7% simple causes (68.8% specifying a property of NP1 and 5.9% of the background type). For comparison, *a-p* verbs only received 20.2% simple causes and only 6.4% simple causes of the non-background type. The overall trends were the same in both languages but there were slight differences in connection with *e-s* verbs. The proportion of simple causes after an *e-s* verb was 93.5% in German, whereas Norwegians continued with a simple cause continuation only 82.0% of the time. We analyzed the proportions of simple causes in a logit mixed effects model with the fixed effects of *language* and *verb class* (*a-p* vs. *e-s* vs. *s-e*) and their interaction. The analysis revealed significant contrasts between *e-s* verbs vs. all other verbs (estimate=5.8; $z=12.9$; $p<.01$) and *s-e* verbs vs. all other verbs (estimate=3.7; $z=10.1$; $p<.01$). Surprisingly, the fixed effect of *language* (estimate=.5; $z=2.9$; $p<.01$) and its interactions with *e-s* (estimate=-1.7; $z=-3.0$; $p<.01$) and *s-e* verbs also reached significance (estimate=-.5; $z=-2.1$; $p<.05$). These effects were due to the fact that Norwegian *e-s* verbs gave rise to a lower proportion of simple causes than their German counterparts. A possible explanation might be that the psych verbs in the two languages had different ambiguity profiles. Although the verbs had the same interpretational variants (e.g. *s-e/a-p* or *e-s/a-p*), the two interpretations were not equally prominent in German

and Norwegian. Consider e.g. *übersehen/overse* ('overlook'), for which the agentive interpretation (roughly corresponding to *ignore*) is preferred for Norwegian *overse*, but highly dispreferred for German *übersehen*. For details, the reader is referred to Case study II in Sect. 5.2.

Presuppositional *a-p* Verbs Presupposition verbs were predicted to trigger a higher proportion of external reasons than *a-p* verbs without a presupposition. Again, the pattern was predicted to be the same in both languages. This prediction was fully borne out. After a presupposition verb participants continued with an external reason in 76.6% of the cases (German: 76.4%; Norwegian: 76.9%), whereas after an *a-p* verb without a presupposition they produced external reasons only 31.4% of the time (German: 33.9%; Norwegian: 28.7%). A logit mixed effects model analyzing the proportions of external reasons of *a-p* verbs including the fixed effects of *language*, *presupposition* and their interaction yielded a reliable effect of *presupposition* (estimate=2.6; $z=6.6$; $p<.01$), but no significant main effect of *language* and no significant interaction ($|z| < 1.5$).

***a-p* Verbs Without a Presupposition** According to our analysis, *a-p* verbs are only compatible with external and internal reasons.¹ To what degree explanations specify one or the other type of reason depends on world knowledge and on social psychological factors as, for instance, role hierarchies or gender stereotypes (see Sect. 3.3). This class of verbs should therefore display the greatest variability of ratios between external and internal reasons, but at the same time obey the proposed ontological restrictions and disallow simple causes. The corpus data support these claims. First of all, there were only very few instances of simple causes (overall: 8.8%; German: 8.0%; Norwegian: 9.6%). Further inspection of the data revealed that the majority (201 out of 250) of these cases were from ambiguous *a-p* verbs which also had a *s-e* interpretation and all of these examples were psych verb uses as illustrated in (22).

- (22) Louisa verletzte Tom, weil sie fremdgegangen war.
 Louisa hurt Tom because she cheated-on had.
 'Louisa hurt Tom because she had cheated on him.'

On average, this kind of verbs triggered 39.1% external reasons (German: 40.4%; Norwegian: 37.6%) and 50.4% internal reasons (German: 48.8%; Norwegian: 50.8%), respectively. To investigate whether German and Norwegian differed significantly from each other, we computed two logit mixed effects models with the fixed effect of *language*. In the first model we analyzed the proportions of external reasons; in the second we analyzed the proportions of internal reasons. Both models revealed no reliable effects of *language* ($|z| < .75$) indicating that the explanation profiles were cross-linguistically stable.

¹ Background explanations were also possible (20.3% of all explanations for non-presuppositional *a-p* verbs were backgrounds). Statistical analyses including these explanations revealed the same pattern of effects.

Fig. 1 The interdependence of causal profiles and IC biases of all non-presupposition *a-p* verbs in the study. A difference score of external vs. internal reasons of -1 corresponds to a verb with only internal reasons; a value of $+1$ corresponds to a verb that only had external reasons. A bias score of -1 corresponds to 100% NP2 and a score of $+1$ corresponds to 100% NP1 continuations.

We then considered the explanation profiles of individual verbs. For each verb we filtered on external and internal reasons and computed a difference score by subtracting the proportion of internal reasons from the proportion of external reasons. A value of 1 indicates that the verb had only external reasons, whereas a value of -1 indicates that it only had internal reasons. The difference scores are depicted in Fig. 1. We also plotted the IC bias scores (a value of -1 corresponding to 100% NP2 continuations and $+1$ to 100% NP1 continuations) in the figure, showing that the ratio between external and internal reasons are strongly correlated with IC bias. In fact, a linear regression model predicting the bias on the basis of the computed ratios accounted for 75% of the total variance (corrected $R^2=.75$). Also in line with our predictions, the figure shows a gradient transition from 100% external reasons to 96% internal reasons which offers a straightforward explanation for the observed variability of IC bias in this class of verbs.

Verbs that License all Three Kinds of Explanations The stimulus argument of a psych verb has no intentional control over the property leading to the experience. Hence, we shouldn't be able to find external and internal reasons. *a-p* verbs, on the other hand, do not allow for simple causes. This made us expect that only ambiguous verbs should be compatible with an explanation profile displaying all three kinds of

Table 1 Verbs (German version) that prompted all three explanation types plus their mean proportions of simple causes, external reasons (*ext. reason*) and internal reasons (*int. reason*). Verb classes as they were originally assigned when we designed the study (see the Appendix).

	<i>simple cause</i>	<i>ext. reason</i>	<i>int. reason</i>	<i>verb class</i>
<i>beleidigen</i> ‘insult’	25.0%	21.4%	53.6%	<i>ap/se</i>
<i>erschrecken</i> ‘frighten’	47.2%	9.4%	43.4%	<i>ap/se</i>
<i>kränken</i> ‘offend’	46.3%	22.2%	31.5%	<i>ap/se</i>
<i>peinigen</i> ‘torment’	15.4%	20.0%	64.6%	<i>ap/se</i>
<i>provozieren</i> ‘provoke’	47.0%	9.1%	43.9%	<i>ap/se</i>
<i>quälen</i> ‘bother’	21.0%	19.4%	56.5%	<i>ap/se</i>
<i>überraschen</i> ‘surprise’	56.9%	27.6%	15.5%	<i>se</i>
<i>übersehen</i> ‘overlook’	52.5%	5.1%	42.4%	<i>ap/es</i>
<i>verletzen</i> ‘hurt’	62.5%	7.1%	26.8%	<i>ap/se</i>

explanations. We examined the explanation profiles of the verbs and selected those that had at least 10 % explanations of the simple cause type as well as at least 10% explanations of either reason type. Table 1 presents the nine verbs that met this criterion.

Eight out of the nine verbs were cases which had been judged to be ambiguous beforehand. The remaining case that at first sight violated the ambiguity criterion for displaying the full range of explanation types was *surprise* (*überraschen/overraske*). Upon closer inspection, *surprise* is clearly ambiguous, arguably also displaying an *a-p* interpretation. Note, for instance, that it can be modified by *intentionally/deliberately*. Thus, all of the verbs exhibiting the full range of explanation types were ambiguous.

5.1.3 IC Bias

How does this affect IC bias? We restricted the analysis to continuations that were explanations and ignored the other productions in the full stop condition. The two languages displayed an almost perfect match in IC bias (Pearson’s $r=.92$; $p<.01$) which supports our claim that the lexical semantics of the tested verbs and the semantics of *weil* and *fordi* are essentially the same in both languages. Furthermore, the explicit explanations in the *weillfordi* (‘because’) conditions showed more or less the same IC bias pattern as the implicit explanations after a full stop (Pearson’s $r=.85$; $p<.01$). Separate correlation analyses for the two languages revealed that the correlation was substantial in German (Pearson’s $r=.76$; $p<.01$) and in Norwegian (Pearson’s $r=.80$; $p<.01$). The correlation between the full stop condition and *weillfordi* gets substantially weaker if the continuations in the full stop condition are not conditioned on explanations (Pearson’s $r=.70$; $p<.01$). This fits the results of Stevenson et al. (2000) and lends support to Kehler et al. (2008) who claimed that coreferential preferences can only be properly understood in case coherence relations are considered.

Fig. 2 Mean NP1 bias in the three psych verb classes. Error bars indicate 95% confidence intervals. *ap* = *a-p* verbs; *es* = *e-s* verbs; *se* = *s-e* verbs

Psych Verbs Figure 2 depicts the mean IC bias scores of the three psych verb classes. The graph shows that the IC biases were approximately the same in German and Norwegian.

e-s verbs exhibited a strong NP2 preference. Only 14.3% of the explanations had a (first) pronoun referring back to NP1. Proportions were similar in the full stop condition (17.4% NP1) and in the *weillfordi* condition (12.5% NP1).

By contrast, *s-e* verbs showed the opposite pattern with an average NP1 bias of 78.5% NP1. To our surprise, there was an across the board 20% bias difference between full stop explanations (65.6% NP1) and *weillfordi* (87.4% NP1). We suspected that this difference was due to the fact that a considerable proportion of full stop explanations were also compatible with a background interpretation (e.g. *John disturbed Mary. Actually, she had to prepare class.*). These were coded as background explanations. To see whether our assumption was on the right track, we excluded all background-type explanations and recomputed the bias of the remaining *s-e* cases. Now, the difference in bias was much smaller, namely 82.1% NP1 in the full stop explanations versus 93.7% NP1 bias in the *weillfordi* explanations.

Finally, we analyzed the bias of verbs that were ambiguous between a *s-e* and an *a-p* interpretation. Overall, these verbs had 78.2% NP1 continuations, i.e. approximately the same amount as the unambiguous *s-e* verbs. The proportions were similar in explanations after a full stop (80.2% NP1) and after a *weillfordi* (77.5% NP1) prompt.

We analyzed the IC bias in all three types in a logit mixed effects model with the fixed effects of *language*, *connective* (*full stop* vs. *weillfordi*), *verb class* (*es* vs. *se* vs. *ap/se*) and their interactions. None of the effects involving *language* reached significance (all $z < 1.7$) indicating that German and Norwegian behaved the same way. Both contrasts involving *verb classes* were reliable. *s-e* verbs differed from the other two verb classes (estimate=1.1; $z=2.7$; $p < .01$) and so did *e-s* verbs (estimate=-

Fig. 3 Mean NP1 bias of *a-p* verbs with (+ *psp*) and without (– *psp*) a presupposition. Error bars indicate 95% confidence intervals

3.4; $z=-8.6$; $p<.01$). Furthermore, the relative difference between NP1 bias of *s-e* verbs depending on the connective was reflected by a significant interaction between *connective* and *verb class* (estimate=-2.0; $z=-4.8$; $p<.01$). No other effects were significant.

The IC biases of the three verb classes met our predictions in almost every respect. *e-s* verbs triggered NP2 continuations, *s-e* verbs showed the opposite pattern and ambiguous *s-e/a-p* verbs weren't stronger than the unambiguous cases. In the *weillfordi* condition the NP1 preference in the unambiguous *s-e*-verbs was 9.2% stronger than in the ambiguous cases. This is what is expected because – under their *a-p* interpretation – the latter are compatible with external reasons whereas the former do not allow this interpretation.

***a-p* Verbs** According to our theory, *a-p* verbs with a presupposition should exhibit a bias towards the associated referent. This should always be the object with the only exception of *apologize* (*entschuldigen sich/unnskyldde seg*; cf. Sect. 3.2.2) which therefore will be analyzed separately. Figure 3 shows the mean bias scores of *a-p* verbs with and without a presupposition.

The presupposition verbs showed a clear NP2 bias. These verbs had only 12.5% NP1 continuations in the *weillfordi* condition (German: 10.7%; Norwegian: 14.4%) and 20.1% NP1 continuations in the explanations after a full stop (German: 27.7%; Norwegian: 9.8%). By contrast, *a-p* verbs without a presupposition were balanced. On average they led to 55.0% NP1 continuations in the *weillfordi* condition (German: 53.9%; Norwegian: 56.1%) and to 47.5% NP1 continuations after a full stop (German: 46.6%; Norwegian: 48.7%). Again, the findings fully supported our predictions with the only exception of the relatively weak NP2 bias of presupposition verbs in the German full stop condition. Yet, we have to be cautious about over-interpreting this apparent difference. Note that it is not necessarily the case that what we took to be an explanation, i.e. a continuation compatible with *because*, was in-

Table 2 Number of NP1 vs. NP2 continuations for *apologize* and *forgive*

	NP1	NP2	total
<i>apologize</i>	77 (98.7%)	1 (1.3%)	78 (100%)
<i>forgive</i>	40 (59.7%)	27 (40.3%)	67 (100%)
<i>total</i>	113	28	145

tended as an explanation by the participant. This is a severe limitation of the present study because the continuations after a full stop generally contained no discourse markers which would have objectified the annotation of discourse relations.

To analyze the bias of the two classes of unambiguous *a-p* verbs we computed a logit mixed effects model including the fixed effects of *presupposition* (unambiguous *a-p* verb with vs. without a presupposition), *language*, *connective* (*full stop* vs. *weillfordi*) and their interactions. The analysis revealed a significant main effect of *presupposition* (estimate=-2.6; $z=-6.1$; $p<.01$) and a significant two-way interaction of *presupposition* and *connective* (1.4; $z=4.4$; $p<.01$) as well as a significant three-way interaction of the three factors (estimate=-1.4; $z=-2.7$; $p<.01$). Apart from these effects no other effects were reliable.

Case Study I: *Apologize* vs. *Forgive* These two verbs offer an interesting test case for our theory because although they clearly are semantically related (similar to e.g. *fear* vs. *frighten*), they were predicted to show an idiosyncratic pattern. The presupposition verb *apologize* (*entschuldigen sich/unnskylda seg*) should trigger external reasons, but, in contrast to all other presupposition verbs in the study, these reasons should be associated with NP1 and the verb should show a strong NP1 bias opposite to the other presupposition verbs. *Forgive* (*verzeihen/tilgi*), on the other hand, also carries a presupposition about a preceding event, but this presupposition cannot be targeted by an explanation (see Sect. 3.2.2). Therefore, *forgive* should pattern like a verb without a presupposition and display no clear bias.

Table 2 shows the distribution of NP1 vs. NP2 continuations for both verbs aggregated over languages and sentence connectives. The pattern was the same across languages and across implicit and explicit explanations. The verbs patterned exactly as predicted. *Apologize* had a clear (85.4%) NP1 bias, whereas *forgive* had no clear bias (59.6% NP1). A χ^2 -test revealed that this bias difference was significant ($\chi^2(1) = 13.9$; $p<.01$). Furthermore, binomial tests revealed a significant difference from a 50% bias for *apologize* ($p<.01$) but not for *forgive* ($p=.14$).

5.1.4 The Link between Explanation Profiles and IC Bias

Figure 4 shows the proportions of NP1 and NP2 continuations of all verb classes and their explanation profiles. It is obvious that particular explanation types were linked to certain kinds of verbs. Simple causes were only licensed by psych verbs and were always associated with the stimulus argument. *a-p* verbs were only compatible

Fig. 4 Proportions of NP1 and NP2 explanations and their explanation profiles. Verb classes: *ap* = *a-p* verbs; *+/- psp* = with/without a presupposition; *es* = *e-s* verbs; *se* = *s-e* verbs

with reasons or explanations of the background type. Their bias was determined by the distribution of external and internal reasons. Presupposition verbs, last but not least, in general received external reasons. The figure illustrates that even though the explanation profiles of two verb classes are completely different, their IC bias may (under special circumstances) be the same. This was the case in unambiguous *s-e* versus ambiguous *s-e/a-p* verbs. The two verb classes had the same IC bias, but in the unambiguous cases the NP1 bias resulted from simple causes, whereas in the ambiguous cases NP1 continuations were due to simple causes and internal reasons.

5.2 Cross-linguistic Differences

All analyses revealed a close correspondence between German and Norwegian. We even observed an equal proportion of explanations in the full stop condition of the German and the Norwegian part of the production study. This finding suggests that IC solely depends on the lexical semantics of the verb and the semantics of the discourse relation of explanation without being influenced by cross-linguistic differences in discourse segmentation. In the great majority of cases we included only verbs that were very close in meaning in German and Norwegian.

We should, however, be able to find micro-variation in verbs with no exact match in lexical semantics. We could identify two verbs in the corpus which superficially look like exact synonyms but slightly differ in meaning.

Case Study II: *provozieren/provosere* and *übersehen/overse* The verb *provoke* (*provozieren/provosere*) has both an *a-p* (= NP1 deliberately aims to achieve a change of state in NP2) and an *s-e* interpretation (= something about NP1 leads to the experience of provocation in NP2). Interestingly, the preferences for either interpretation clearly differed between German and Norwegian. Whereas in German, NP1 *provozierte* NP2 preferably received an *a-p* interpretation, the Norwegian verb *provosere* was preferably interpreted psychologically, i.e. as an *s-e* verb. The observed explanation profiles were thus rather different in the two languages: while Germans overwhelmingly produced internal reasons (62.9%), Norwegians predominantly produced simple causes (68.8%). Although the explanation profiles were different ($\chi^2(1)=18.7$; $p<.01$), the bias was the same (German: 83% NP1; Norwegian: 94% NP1; $\chi^2(1)=1.9$; $p=.17$), since both simple causes of an *s-e* verb and internal reasons are associated with NP1.

The second case in question is the verb *overlook* (*übersehen/overse*). The German verb was preferably interpreted as a perception verb of *e-s* type, whereas for the Norwegian counterpart the *a-p* reading was preferred. In contrast to what was the case for *provoke*, this should thus have implications for both the explanation profile and IC bias. Whereas Germans preferably produced simple causes (81.6%) associated with NP2 (84.2% NP2 bias), Norwegians continued with internal reasons (75.8%) which are linked to NP1 (75.7% NP1 bias). χ^2 tests revealed a significant difference in explanation profiles between the two languages ($\chi^2(1)=18.7$; $p<.01$) and in IC bias ($\chi^2(1)=25.8$; $p<.01$).

This case study illustrates cross-linguistic differences in explanation profiles which show up whenever ambiguous *s-e/a-p* verbs differ cross-linguistically in their preferred interpretations. This can, cross-linguistically, lead to exactly opposite IC biases. The observed shift in bias for *overlook*, but not for *provoke*, strongly depended on the explanation profile.

6 General Discussion and Conclusions

We have proposed a semantic analysis which allowed us to derive detailed predictions about the expected discourse relation in explanatory discourse. We made use of a typology of explanations independently developed by Solstad (2010) and showed how these explanation types interact with missing semantic content of particular interpersonal verbs with two proper name arguments. This way, the idiosyncratic coreference preferences of verbs with different IC biases could be treated as an epiphenomenon of the expected explanation type. Our analysis is essentially a coherence based theory in the spirit of Hobbs (1979) and Kehler (2002), but extends the proposal of Kehler et al. (2008) in significant ways.

First of all, Kehler et al. (2008, p. 35) noticed that “IC verbs” are more likely to evoke explanation continuations than “non-IC verbs” without offering any explanation for this. They remarked that “this suggests that the lexical semantics of IC verbs create a stronger-than-usual expectation for an explanation.” Our account can

explain why this should be the case. If there is causal content which can be specified by an explanation in the form of a *because* clause or an independent sentence, then providing this explanation should be the default strategy in language processing. Otherwise interpreters would have to accommodate the missing information, a cognitively taxing operation which should thus be dispreferred. Second, our account does not only predict which verbs attract an explanation, it also predicts the kind of explanation likely to come. And since these kinds of explanations are closely associated with specific coreferential preferences, the explanation profile also predicts the IC bias of any verb. Finally, we consider our account a cognitively more plausible model of coherence driven processing. Instead of having to assume computationally difficult abductive inference as the driving force of discourse interpretation, our theory assumes a much easier operation, namely specification of yet unspecified content, i.e. on a par with the cataphoric justification of a presupposition. The mechanisms are basically of the same kind as when predicting yet missing structure in syntactic processing. This is not to say that abduction never plays a role. In fact, abduction may be the only available way to go in case the interpretation system is facing integration difficulty between locally incompatible discourse segments.

In a recent proposal, Hartshorne & Snedeker (2012) claim that the bias is due to the fact that *because* clauses specify the causing entity in a causal relation provided by the verb. On the face of it, this proposal may seem to be compatible with our analysis. However, we contend that Hartshorne & Snedeker's claim cannot be correct as it stands. Rather, it can at most be valid for psych verbs. To see this, consider the *s-e* verb *annoy* and the causative *a-p* verb *kill*. They may both be assumed to involve a sub-lexical CAUSE relation between two entities. Consequently, Hartshorne & Snedeker (2012) predict that both verbs should display a NP1 bias. However, *kill* and other causative *a-p* verbs show an overall balanced bias pattern (see Bott & Solstad 2012). A crucial factor why causative *a-p* verbs (i.e. non-psychological verbs) come out differently, is displayed in (23):

- (23) a. John annoyed Mary because he sang loudly.
 b. #John killed Mary because he stabbed her in the back.
 c. John killed Mary by stabbing her in the back.

In contrast to what is the case for *s-e* verbs, cf. (23a), it seems pragmatically odd to specify the causing entity in a causative *a-p* verb, i.e. a simple cause, by means of a *because* clause, cf. (23b). The preferred mode of expression for causes of causative *a-p* verbs is modification by a *by* phrase (cf. (23c); see e.g. Sæbø 2008, and references therein; #=pragmatically anomalous on simple cause reading). This is another argument in favour of emphasizing the discourse perspective on IC and not only the causal properties of the verb. It is crucial that the explanation may be given in the form of a *because* clause or a sentence.

Another point to be made concerning the fine-grained semantic analysis of IC verbs, is that our approach allows us to make precise predictions for individual verbs, cf. the discussion of the differences between *apologize* and *forgive* in Case study I on the one hand, and between *apologize* and other presuppositional *a-p* verbs such as e.g. *punish*, *congratulate*, and *thank* (see Sect. 3.2.2) on the other.

Turning now to the debate over focusing vs. integration, what sets our theory apart from focusing accounts proposed in the psycholinguistic literature (McKoon et al. 1993; Greene & McKoon 1995; Stevenson et al. 1994; McDonald & MacWhinney 1995; Stevenson et al. 2000; Koornneef & van Berkum 2006; van Berkum et al. 2007; Featherstone & Sturt 2010; Pyykkönen & Järvikivi 2010; Cozijn et al. 2011)? Our account clearly involves predicting certain explanation types and therefore focusing of the involved event participants. Nevertheless, we do not take IC to be one factor among others that leads to direct activation of some argument role. Instead, on our account IC activates an event participant only indirectly via the prediction of a particular explanation type. Therefore, we take IC bias to be merely correlated with certain thematic roles. We have discussed various cases in which the correlation clearly breaks down: *a-p* verbs with a presupposition clearly differ from *a-p* verbs without a presupposition. Furthermore, the presupposed event can be associated with either argument. Cases in question were *apologize to* as compared to *punish* (Sect. 3.2.2), where the former was associated with NP1, whereas the latter was associated with NP2. We think that our account is conceptually advantageous over a mere activation account since an explanation constitutes a relation between something propositional in nature, and not individuals. Moreover, since we take explanations to relate propositions or facts, not only the verb but also the linguistic realization of the arguments should matter, a fact that has been observed in social psychology (e.g. Corrigan 2001) but sometimes neglected in psycholinguistics (for discussion, see e.g. Ferstl et al. 2011).

In fact, we have recently tested our compositional account of implicit causality in a number of additional continuation experiments which show that IC bias can be manipulated in a predictable way if we add optional material to the matrix clause (Bott & Solstad 2012, in prep.). Consider (24a) and (24b) where the prepositional phrases in italics elaborate on the bias referent, much like a *because* clause:

- (24) a. Peter thanked Mary *for all the support before the exam*. . .
 b. Peter impressed Mary *with his aggressive play*. . .

Note that the PP is about the support by NP2 in (24a) and the aggressive play of NP1 in (24b), i.e. the bias argument. From the perspective of multiple interactive constraints we would expect the bias – if anything at all – to be even stronger in cases where a modifying PP elaborates on the bias referent because this could add activation to it. On our account, however, it is obvious that the modifier provides the missing information which triggers the bias in the first place, filling the empty slot. Hence, interpreters can be expected to choose a different discourse relation. If they nevertheless provide an explanation, they can be expected to select a different kind of explanation which should in turn lead to a clear shift in bias. This is exactly what our data show. Crucially, this finding corroborates our claim that in order to properly understand the phenomenon of implicit causality we need a compositional theory that is able to take into account the lexical semantics of the verb, the semantic properties of its arguments, interactions with (adverbial) modifiers, the semantics of the explanation relation as well as contextual influences.

Another merit of our account is that focusing and integration (as proposed e.g. by Garnham et al. 1996; Stewart et al. 2000) are not viewed as mutually exclusive alternatives, but that besides focusing effects (as recently demonstrated by Koornneef & van Berkum 2006; van Berkum et al. 2007; Featherstone & Sturt 2010; Pykkönen & Järvikivi 2010; Cozijn et al. 2011) integration effects are clearly expected. Whenever the discourse continues in an unpredicted way, the unexpected continuation has to be integrated. For instance, it is certainly possible that a continuation is congruent with the pronoun bias, but violates the expected type of explanation. This is illustrated in (25a) where instead of a simple cause the *because* clause specifies an external reason. By contrast, example (25b) illustrates a case where the *because* clause violates the bias, but the expected kind of explanation still appears in an embedded clause. Hence integration should be easy, even though the processor is dealing with a bias violation.

- (25) a. John disturbed Mary because he was angry at her.
 b. John congratulated Mary because he thought that she had deserved to win the race.

Taken together, our theory synthesizes the different facets of implicit verb causality that have been debated in the literature. It meets all the desiderata which we formulated in Sect. 2. Moreover, it offers a detailed account of predictive and integrative discourse processing at a level as abstract as discourse relations. Our theory constitutes the first account of IC bias connecting verb semantics, discourse structure, and coreference.

Acknowledgements We would like to thank Cathrine Fabricius-Hansen, Barbara Hemforth, Juhani Järvikivi, Piritä Pykkönen-Klauck, Barbara Schmiedtová as well as two anonymous reviewers for comments on an earlier draft; furthermore Bergljot Behrens, Stefan Engelberg, Lyn Frazier, Robin Hörnig, Katja Jasinskaja, Hans Kamp, Elena Karagjosova, Arnout Koornneef, Katrin Petodnig, Anna Pryslopska, Arndt Riester, Antje Roßdeutscher, Kjell Johan Sæbø, Henk Zeevat, as well as the audiences at the following conferences and workshops for discussing these issues with us: “Discourse Representation, Comprehension and Production in a Cross-linguistic Perspective”, “5th Linguistic Evidence 2012”, “25th CUNY conference”, “11th International Conference of Nordic and General Linguistics”, “Semantik und Pragmatik im Südwesten 4”, “Informationsstruktur als Teil der menschlichen Kognition”, and “Discourse Expectations: Theoretical, Experimental, and Computational Perspectives” (DETEC2013). This research was made possible by grants from the Deutsche Forschungsgemeinschaft to project B1 of the SFB 833 “The construction of meaning – the dynamics and adaptivity of linguistic structures” at the University of Tübingen, and project B4 of the SFB 732 “Incremental specification in context” at the University of Stuttgart. We also gratefully acknowledge the financial support from the Centre for Advanced Study, the Norwegian Research Council (NFR project IS-DAAD 216850), and the German Academic Exchange Service (DAAD). Last but not least, we would like to thank Cathrine Fabricius-Hansen for inviting us to spend some inspiring months at the Centre for Advanced Study (Oslo, Norway) during the winter of 2010/2011.

References

- Abelson, R., & Kanouse, D. (1966). Subjective acceptance of verb generalizations. In S. Feldman (Ed.) *Cognitive consistency* (pp. 173–199). San Diego, CA: Academic Press.
- Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, 30(3), 191–238.
- Au, T. (1986). A verb is worth a thousand words: the causes and consequences of interpersonal events implicit in language. *Journal of Memory and Language*, 25(1), 104–122.
- Baumann, P., Konieczny, L., & Hemforth, B. Conversational implicatures in anaphora resolution: alternative constructions and referring expressions, this volume.
- Bott, O., & Solstad, T. (2012). The mechanics of implicit verb causality. Poster presented at the 25th CUNY conference, New York.
- Bott, O., & Solstad, T. The mechanics of causal interpretation: Explaining the implicit causality bias. Manuscript in preparation.
- Brown, R., & Fish, D. (1983). The psychological causality implicit in language. *Cognition*, 14(3), 237–273.
- Caramazza, A., Grober, E., Garvey, C., & Yates, J. (1977). Comprehension of anaphoric pronouns. *Journal of Verbal and Learning & Verbal Behavior*, 16(5), 601–609.
- Carreiras, M., Garnham, A., & Oakhill, J. (1996). Understanding anaphora: the role of superficial and conceptual information. In M. Carreiras, J. García-Albea, & N. Sebastián-Gallés (Eds.) *Language processing in Spanish* (pp. 241–274). Mahwah: Erlbaum.
- Colonna, S., Schimke, S., & Hemforth, B. Information structure and pronoun resolution in German and French: Evidence from the visual-world paradigm, this volume.
- Corrigan, R. (2001). Implicit causality in language: Event participants and their interactions. *Journal of Language and Social Psychology*, 20(3), 285–320.
- 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.
- Crinean, M., & Garnham, A. (2006). Implicit causality, implicit consequentiality and semantic roles. *Language and Cognitive Processes*, 21(5), 636–648.
- Ehrlich, K. (1980). Comprehension of pronouns. *The Quarterly Journal of Experimental Psychology*, 32(2), 247–255.
- Fabricsius-Hansen, C. (1999). Information packaging and translation: aspects of translational sentence splitting (German – English/Norwegian). In M. Doherty (Ed.) *Sprachspezifische Aspekte der Informationsverteilung* (pp. 175–214). Berlin: Akademie-Verlag.
- Fabricsius-Hansen, C., Ramm, W., Solfeld, K., & Behrens, B. (2005). Coordination, discourse relations, and information packaging - cross-linguistic differences. In *Proceedings of the Symposium on the Exploration and Modeling of Meaning (SEM-05)* (pp. 85–93). Biarritz, France.
- 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.
- 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.
- Garnham, A. (2001). *Mental models and the interpretation of anaphora*. Hove: Psychology Press.
- Garnham, A., & Oakhill, J. (1985). On-line resolution of anaphoric pronouns: effects of inference making and verb semantics. *British Journal of Psychology*, 76(3), 385–393.
- Garnham, A., Oakhill, J., & Cruttenden, H. (1992). The role of implicit causality and gender cue in the interpretation of pronouns. *Language & Cognitive Processes*, 7(3-4), 231–255.
- 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.
- Garvey, C., & Caramazza, A. (1974). Implicit causality in verbs. *Linguistic Inquiry*, 5(3), 459–464.
- Garvey, C., Caramazza, A., & Yates, J. (1974). Factors influencing assignment of pronoun antecedents. *Cognition*, 3(3), 227–243.

- Gernsbacher, M. A., & Hargreaves, D. (1988). Accessing sentence participants: The advantage of first mention. *Journal of Memory and Language*, 27(6), 699–717.
- 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.
- Greene, S. B., & McKoon, G. (1995). Telling something we can't know: Experimental approaches to verbs exhibiting implicit causality. *Psychological Science*, 6(5), 262–270.
- Grosz, B., Joshi, A. K., & Weinstein, S. (1995). Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2), 203–225.
- Guerry, M., Gimenes, M., Caplan, D., & Rigalleau, F. (2006). How long does it take to find a cause? an online investigation of implicit causality in sentence production. *The Quarterly Journal of Experimental Psychology*, 59(9), 1535–1555.
- Hartshorne, J. K., & Snedeker, J. (2012). Verb argument structure predicts implicit causality: The advantages of finer-grained semantics. *Language and Cognitive Processes*. doi:10.1080/01690965.2012.689305.
- Hartshorne, J. K., Sudo, Y., & Uruwashii, M. (2013). Are implicit causality pronoun resolution biases consistent across languages and cultures? *Experimental Psychology*, 60(3), 179–96.
- Hobbs, J. R. (1979). Coherence and coreference. *Cognitive Science*, 3(1), 67–90.
- Jackendoff, R. (1990). *Semantic Structures*. Cambridge, MA: MIT Press.
- Jäger, T. (2008). Categorical Data Analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language*, 59(4), 434–446.
- Kamp, H., van Genabith, J., & Reyle, U. (2011). Discourse representation theory. In D. Gabbay, & F. Guenther (Eds.) *Handbook of Philosophical Logic* (pp. 125–394). Dordrecht/Heidelberg/London/New York: Springer.
- Kehler, A. (2002). *Coherence, reference, and the theory of grammar*. Stanford, CA: CSLI Publications.
- Kehler, A., Kertz, L., Rohde, H., & Elman, J. L. (2008). Coherence and coreference revisited. *Journal of Semantics*, 25(1), 1–44.
- Kipper-Schuler, K. (2006). *VerbNet: a broad-coverage, comprehensive verb lexicon*. Ph.D. thesis, University of Pennsylvania.
- Koornneef, A. W., & van Berkum, J. J. A. (2006). On the use of verb-based implicit causality in sentence comprehension: Evidence from self-paced reading and eye tracking. *Journal of Memory and Language*, 54(4), 445–465.
- Landau, I. (2010). *The Locative Syntax of Experiencers*. Cambridge, MA: MIT Press.
- Long, D., & Ley, L. D. (2000). Implicit causality and discourse focus: the interaction of text and reader characteristics in pronoun resolution. *Journal of Memory and Language*, 42(4), 545–570.
- Malle, B. (2002). Verbs of interpersonal causality and the folk theory of mind and behavior. In M. Shibata (Ed.) *The Grammar of Causation and Interpersonal Manipulation* (pp. 57–83). Amsterdam: John Benjamins.
- McDonald, J. L., & MacWhinney, B. (1995). The time-course of anaphor resolution: Effects of implicit verb causality and gender. *Journal of Memory and Language*, 34(4), 543–566.
- McKoon, G., Greene, S. B., & Ratcliff, R. (1993). Discourse models, pronoun resolution, and the implicit causality of verbs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(5), 1040–1052.
- Pesetsky, D. (1995). *Zero Syntax: Experiencers and Cascades*. Cambridge, MA: MIT Press.
- Pickering, M., & Majid, A. (2007). What are implicit causality and implicit consequentality? *Language and Cognitive Processes*, 22(5), 780–788.
- Pyykkönen, P., & Järviö, J. (2010). Activation and persistence of implicit causality information in spoken language comprehension. *Experimental Psychology*, 57(1), 5–16.
- Ramm, W. (2011). *Satzgrenzenveränderungen in der Übersetzung: Satzverbindung und lokale Diskursorganisation im Norwegischen und Deutschen*. Ph.D. thesis, University of Oslo.
- Rudolph, U. (1997). Implicit verb causality: verbal schemas and covariation information. *Journal of Language & Social Psychology*, 16(2), 132–158.

- Rudolph, U., & Försterling, F. (1997). The psychological causality implicit in verbs: a review. *Psychological Bulletin*, 121(2), 192–218.
- Sæbø, K. J. (2008). The structure of criterion predicates. In J. Dölling, T. Heyde-Zybatow, & M. Schäfer (Eds.) *Event Structures in Linguistic Form and Interpretation* (pp. 127–147). Berlin: Mouton de Gruyter.
- Solfjeld, K. (1996). Sententiality and translation strategies German-Norwegian. *Linguistics*, 34(3), 567–590.
- Solstad, T. (2010). Some new observations on ‘because (of)’. In M. Aloni, H. Bastiaanse, T. de Jager, & K. Schulz (Eds.) *Logic, Language and Meaning: 17th Amsterdam Colloquium*, (pp. 436–445). Berlin: Springer.
- Stevenson, R., Knott, A., Oberlander, J., & McDonald, S. (2000). Interpreting pronouns and connectives: Interactions among focusing, thematic roles and coherence relations. *Language and Cognitive Processes*, 15(3), 225–262.
- Stevenson, R. J., Crawley, R. A., & Kleinman, D. (1994). Thematic roles, focus and the representation of events. *Language and Cognitive Processes*, 9(4), 519–548.
- Stewart, A. J., Pickering, M. J., & Sanford, A. J. (2000). The time course of the influence of implicit causality information: Focusing versus integration accounts. *Journal of Memory and Language*, 42(3), 423–443.
- van Berkum, J. J. A., Brown, C. M., & Hagoort, P. (1999). Early referential context effects in sentence processing: evidence from event-related potentials. *Journal of Memory and Language*, 41(2), 147–182.
- van Berkum, J. J. A., Koornneef, A. W., Otten, M., & Nieuwland, M. S. (2007). Establishing reference in language comprehension: An electrophysiological perspective. *Brain Research*, 1146, 158–171.
- van der Sandt, R. A. (1992). Presupposition projection as anaphora resolution. *Journal of Semantics*, 9(4), 333–377.
- Vonk, W. (1985). On the purpose of reading and the immediacy of processing pronouns. In A. Gale, & F. Johnson (Eds.) *Theoretical and applied aspects of eye movement research* (pp. 207–215). Amsterdam: North Holland.
- Zeevat, H. (2000). The asymmetry of optimality theoretic syntax and semantics. *Journal of Semantics*, 17(3), 243–262.

Appendix: Verbs tested in the study (+ IC bias scores)

Verbs appear as classified when we designed the study. Some of the initial decisions should be reconsidered given the corpus evidence (see the discussion in Sect. 5). All values are relative frequencies NP1; Ø=overall bias, G=German, N= Norwegian.

Table 3 unambiguous *s-e* verbs (N=16)

<i>G verb</i>	<i>N verb</i>	<i>translation</i>	$\emptyset$	<i>G</i>	<i>N</i>
amüsieren	more	'amuse'	0.88	0.92	0.83
ängstigen	uroe	'distress'	0.81	0.81	0.81
beeindrucken	imponere	'impress'	0.84	0.75	0.92
bezaubern	sjarmere	'charm'	0.91	0.88	0.94
enttäuschen	skuffe	'disappoint'	0.65	0.67	0.63
entzücken	begeistre	'ravish'	0.91	0.95	0.86
erstaunen	forbause	'astonish'	0.63	0.62	0.63
faszinieren	fascinere	'fascinate'	0.80	0.77	0.84
gefallen	behage	'please'	0.91	0.95	0.87
inspirieren	inspirere	'inspire'	0.87	0.80	0.94
irritieren	irritere	'irritate'	0.83	0.79	0.89
langweilen	kjede	'bore'	0.85	0.88	0.82
sjockieren	sjokkere	'shock'	0.76	0.76	0.76
überraschen	overraske	'surprise'	0.57	0.57	0.56
verblüffen	forbløffe	'amaze'	0.65	0.64	0.67
verwirren	forvirre	'confuse'	0.75	0.78	0.71

Table 4 A: unambiguous *e-s* verbs (N=17), B: ambiguous *e-s/a-p* verb (N=1)

<i>G verb</i>	<i>N verb</i>	<i>translation</i>	$\emptyset$	<i>G</i>	<i>N</i>
A					
beneiden	misunne	'envy'	0.02	0.04	0.00
betrauern	sørge over	'mourn'	0.12	0.08	0.16
bewundern	beundre	'admire'	0.05	0.04	0.05
entdecken	oppdage	'spot'	0.20	0.27	0.13
fürchten	frykte	'fear'	0.11	0.09	0.14
hassen	hate	'hate'	0.09	0.05	0.13
lieben	elske	'love'	0.07	0.10	0.04
misstrauen	mistro	'distrust'	0.19	0.17	0.21
Mitleid haben mit	føle med	'feel sympathy for'	0.20	0.09	0.33
mögen	like	'like'	0.01	0.00	0.03
respektieren	respektere	'respect'	0.04	0.02	0.06
sorgen sich um	bekymre seg for	'worry about'	0.14	0.19	0.09
verabscheuen	avsky	'detest'	0.10	0.11	0.10
verachten	forakte	'despise'	0.16	0.20	0.13
verdächtigen	mistenke	'suspect'	0.28	0.23	0.33
vergöttern	forgude	'adore'	0.09	0.09	0.09
vermissen	savne	'miss'	0.36	0.45	0.25
B					
übersehen*	overse	'overlook', 'ignore'	0.44	0.16	0.76

* See case study II in Sect. 5.2 for details on the discrepancy in bias.

Table 5 ambiguous *a-p/se* verbs (N=14)

<i>G verb</i>	<i>N verb</i>	<i>translation</i>	$\emptyset$	<i>G</i>	<i>N</i>
beleidigen	fornærme	'insult'	0.72	0.66	0.80
beruhigen	berolige	'calm (down)'	0.43	0.33	0.59
berühren	berøre	'touch'	0.78	0.83	0.72
erschrecken	skremme	'frighten'	0.82	0.79	0.85
foltern	torturere	'torture'	0.63	0.69	0.57
hypnotisieren	hypnotisere	'hypnotize'	0.60	0.61	0.59
kränken	krenke	'offend'	0.69	0.72	0.67
necken	erte	'tease'	0.56	0.77	0.35
peinigen	pine	'torment'	0.76	0.73	0.79
provizieren	provosere	'provoke'	0.88	0.83	0.94
quälen	plage	'bother'	0.72	0.70	0.74
stören	forstyrre	'disturb'	0.69	0.68	0.69
überzeugen	overbevise	'convince'	0.84	0.83	0.84
verletzen	såre	'hurt'	0.88	0.94	0.82

Table 6 *a-p* verbs with a presupposition (N=10)

<i>G verb</i>	<i>N verb</i>	<i>translation</i>	$\emptyset$	<i>G</i>	<i>N</i>
belønne	belønne	'reward'	0.12	0.12	0.13
berichtige	rette på	'correct'	0.23	0.29	0.13
bestrafte	straffe	'punish'	0.16	0.15	0.17
danke	takke	'thank'	0.14	0.11	0.17
entschuldigen sich bei	unnskylde seg overfor	'apologize to'	0.99	1.00	0.97
gratulieren	gratulere	'congratulate'	0.03	0.02	0.05
kritisieren	kritisere	'criticize'	0.27	0.31	0.21
loben	rose	'praise'	0.06	0.02	0.11
rächen sich an	hevne seg på	'get back at'	0.26	0.37	0.12
trösten	trøste	'comfort'	0.22	0.32	0.10

Table 7 unambiguous *a-p* verbs (N=43)

<i>G verb</i>	<i>N verb</i>	<i>translation</i>	$\emptyset$	<i>G</i>	<i>N</i>
antworten	svare	'answer'	0.41	0.43	0.39
attackieren	angripe	'attack'	0.40	0.38	0.42
bandagieren	bandasjere	'bandage'	0.18	0.08	0.28
behindern	hindre	'impede'	0.95	0.97	0.92
belehren	belære	'teach'	0.42	0.30	0.55
belügen	lyve til	'lie to'	0.91	0.92	0.91
beschützen	beskytte	'protect'	0.45	0.41	0.49
besiegen	beseire	'defeat'	0.78	0.85	0.71
betrügen	bedra	'deceive'	0.68	0.66	0.70
denunzieren	angi	'denounce'	0.64	0.60	0.68
entlarven	avsløre	'expose'	0.42	0.43	0.42
ermorden	myrde	'murder'	0.60	0.65	0.54
erschlagen	slå i hjel	'batter to death'	0.51	0.46	0.57
feuern	avskjedige	'fire'	0.05	0.00	0.12
helfen	hjelp	'help'	0.36	0.29	0.44
krachen in	smelle inn i	'crash into'	0.84	0.83	0.86
meiden	unngå	'avoid'	0.55	0.47	0.66
misshandeln	mishandle	'mistreat'	0.76	0.68	0.83
mobben	mobbe	'bully'	0.42	0.50	0.32
operieren	operere	'operated on'	0.22	0.15	0.30
pflügen	pleie	'care for'	0.21	0.25	0.17
rempeln gegen	støte mot	'barge into'	0.73	0.72	0.76
schlagen	slå	'hit'	0.45	0.32	0.60
schubsen	dytte	'push'	0.46	0.45	0.46
täuschen	lure	'mislead'	0.82	0.90	0.73
töten	drepe	'kill'	0.41	0.39	0.43
überfahren	kjøre over	'run over'	0.55	0.57	0.52
überlisten	overliste	'outwit'	0.61	0.59	0.63
überreden	overtale	'persuade'	0.77	0.86	0.67
unterbrechen	avbryte	'interrupt'	0.59	0.59	0.58
verfolgen	forfølge	'follow'	0.80	0.79	0.81
verführen	forføre	'seduce'	0.84	0.84	0.84
verhaften	arrestere	'arrest'	0.16	0.18	0.13
verhöhnern	håne	'mock'	0.39	0.39	0.40
verhören	avhøre	'examine'	0.47	0.51	0.41
verjagen	jage	'chase away'	0.46	0.44	0.48
verprügeln	banke	'bash'	0.38	0.23	0.54
verraten	forråde	'betray'	0.75	0.74	0.77
verspotten	gjøre narr av	'deride'	0.20	0.18	0.21
verteidigen	forsvare	'defend'	0.57	0.50	0.65
verwunden	skade	'injure'	0.85	0.87	0.84
verzeihen	tilgi	'forgive'	0.60	0.53	0.68
wecken	vekke	'wake up'	0.27	0.22	0.33