

***He was run-over by a bus*: Passive – but not pseudo-passive – sentences are rated as more acceptable when the subject is highly affected. New data from Hebrew, and a meta-analytic synthesis across English, Balinese, Hebrew, Indonesian and Mandarin**

Ben Ambridge, University of Manchester, ESRC International Centre for Language and Communicate Development (LuCID), UK, ben.ambridge@manchester.ac.uk

Inbal Arnon, Hebrew University of Jerusalem, IL, inbal.arnon@gmail.com

Dani Bekman, Hebrew University of Jerusalem, IL, dani.bekman@mail.huji.ac.il

Several recent experimental studies have investigated the hypothesis that the passive construction is associated with the semantics “[B] (mapped onto the surface [passive] subject) is in a state or circumstance characterized by [A] (mapped onto the by-object or an understood argument) having acted upon it”. (Pinker et al., 1987, p. 249). In the present work, we extend this method to a new language, Hebrew, and conduct a Bayesian mixed-effects meta-analytic synthesis which draws together the findings of similar studies conducted for English, Indonesian, Mandarin and Balinese. For Hebrew, we found that native adult speakers’ ($N = 60$) acceptability ratings for passives with each of 56 different verbs were significantly predicted by verb-semantic affectedness ratings provided by a separate group of 16 native adult speakers. Both for Hebrew and across languages, we found that these semantic affectedness ratings always predict verbs’ acceptability in passive constructions, but not – in most cases – in active and pseudo-passive constructions. However, a meta-analytic synthesis revealed only suggestive evidence that this semantic effect is, across languages, *larger* for passives than actives and pseudo-passives; due probably to a lack of power when collapsing across languages ($N = 5$). These findings support – if only tentatively – the possibility that a passive construction that denotes undergoer-affectedness may approach the status of a semantic universal.



1. Introduction

1.1 The importance of the passive

Passive constructions have long played a central role in linguistic theorizing and psycholinguistic research. On the one hand, passives have been used to argue for the autonomy of syntax (see, e.g., Alexiadou et al., 2018, for a historical review), since they reverse canonical participant role-ordering (i.e., PATIENT ACTION AGENT), yet still maintain the language's canonical word order in terms of syntactic (as opposed to semantic) roles (e.g., SUBJECT VERB OBJECT for English). This line of argument dates as far back as at least Chomsky (1957), but also holds for non-Chomskyan “simpler syntax” accounts (e.g., Culicover & Jackendoff, 2005; Pollard & Sag, 1994) that posit a “syntactic level of representation [that] includes syntactic category information but not semantic information...or lexical content” (Branigan & Pickering, 2017, p. 8). Indeed, some frameworks go so far as to assume that the passive construction enjoys no independent existence, and that passives are derived from other sentence types or from basic principles:

Constructions such as...[the] passive remain only as taxonomic artifacts, collections of phenomena explained through the interaction of the principles of UG, with the values of the parameters fixed. (Chomsky, 1993, p. 4).

On the other hand, passives have been used as evidence for a tight link between syntax and function/semantics. For example, passives are used more often when only one of the noun phrases is animate, when the verb is of the *theme-experiencer* type (e.g., *John was startled by a loud noise...*; Ferreira, 1994) and when it allows an easily-retrievable (given) referent to be in subject position (...*but he was reassured by his travelling companion*; MacDonald, 2013).

In terms of the semantics of the (English) passive, one particularly influential proposal (Pinker et al., 1987; see also Pinker, 1989; Pullum, 2014) is as follows:

“[B] (mapped onto the surface [passive] subject) is in a state or circumstance characterized by [A] (mapped onto the by-object or an understood argument) having acted upon it”. (p.249).

Thus, a passive such as *The pedestrian was run over by a bus* is prototypical, while a passive such as **\$5 was cost by the book* is ungrammatical, since it is not possible to construe \$5 as having been “acted upon” – even metaphorically – by the book. Theoretically speaking, the notion of a semantically-based passive construction is perhaps most strongly associated with cognitive-linguistic approaches, which view semantically-meaningful constructions in general (i.e., not just the passive) as a central tenet (e.g., Croft & Cruse, 2004; Goldberg, 1995; Hilpert, 2014). However, functional/semantics-based approaches to the passive exist in other theoretical frameworks too; and not only in the related approach of Lexical Functional Grammar (e.g., Pinker, 1989). For

example, Aissen (1999) sets out a theory of subject choice in Optimality Theory, while Paolazzi et al. (2022) locate the (apparent) semantic difficulty with *experiencer-theme* passives (e.g., *John was seen by Bill*) in event structure, rather than syntax *per se* (specifically, the difficulty arises when coercing a state into receiving an eventive interpretation). Similarly, Nguyen and Pearl (2021) situate this effect – focussing on children – in lexical semantics.

While the theoretical implications remain debatable – and debated – the finding of something like Pinker’s semantic constraint on the (English) passive appears to be well supported. In the past decade, a decent number of adult studies have tested this proposal, and generally found evidence in its favour (Ambridge et al., 202; Bidgood et al., 2020; Jones et al., 2021; though for partial null findings, see Darmasetiyawan et al., 2012; Messenger et al., 2012; for a meta-analysis of child studies, see Nguyen & Pearl, 2021).

Most relevant to the present work is the English study of Ambridge, Bidgood, Pine, Rowland and Freudenthal (2016): A group of native adult speakers were asked to rate each of 72 verbs for the extent to which each exhibits each of 10 semantic properties related to Pinker et al.’s (1987) notion of “affectedness” (e.g., *A is doing something to B; B changes state or circumstances; the action adversely (negatively) affects B*). These ratings were then combined using Principal Components Analysis to create a composite “affectedness” score for each verb. (Note that here and throughout this paper, we are using the term “affectedness” simply as a shorthand for the cluster of semantic properties that are characteristic of the passive construction, even though some are only tangentially related to affectedness *per se*; an issue to which we return in more detail in 2.1). The researchers then obtained, from a separate group of adult speakers, grammatical acceptability judgments of each verb in passive and, as a control, active sentences (e.g., *Bob was kicked by Wendy; Wendy kicked Bob*). For passive sentences, a large and significant correlation between semantic affectedness and sentence grammatical acceptability was observed. For example, the verbs *hit, kick, push* and *shake* all received close to the maximum possible semantic affectedness score, and close to the maximum possible acceptability scores in passive sentences (e.g., *Bob was hit/kicked/pushed/shaken by Wendy*). Conversely, the verbs *like, remember, know* and *believe* received low semantic affectedness scores, and lower acceptability judgment scores in passive sentences (e.g., *Bob was liked/remembered/known/believed by Wendy*).

Complicating the picture somewhat, a similar – though significantly smaller – correlation was observed for active sentences. That is, participants also displayed a small preference for sentences such as *Wendy hit/kicked/pushed/shook Bob* over sentences such as *Wendy liked/remembered/knew/believed Bob*. There are two possible explanations for this semantic affectedness effect for actives; though the two are not mutually exclusive. The first is that the active transitive construction is – like the passive – also prototypically associated with a cluster of semantic properties related to affectedness (e.g., Ambridge et al., 2014; Hopper & Thompson, 1980;

1984; Ibbotson et al., 2012; Næss, 2007; Talmy, 1985). For example, Ibbotson et al. (2012, p. 1270) characterize the semantics of the active transitive construction as an “agent intentionally instigating an action that directly results in the patient being affected” (Ibbotson et al., 2012, p. 1270). If this possibility is correct, then we would expect to see semantic affectedness effects observed for acceptability judgments with *both* passives and actives; and possibly even similar-size effects for the two constructions (i.e., no interaction of semantic affectedness by sentence-type passive/active).

That said, at least assuming a construction-based/functionalist approach, languages clearly “need” a two-participant construction that does not (necessarily) denote high affectedness (e.g., to convey a scenario in which Wendy pushed Bob, but so gently that he barely noticed). Thus, for every language that shows a semantic affectedness effect for passives, we would expect to see a smaller, and close to zero, semantic affectedness effect for at least one of (a) actives and (b) pseudo-passives (defined in more detail below).

The second possible explanation for this semantic affectedness effect for actives is that it is an experimental artefact resulting from some kind of general dispreference and/or processing difficulty for verbs that score low on affectedness. For example, higher-affectedness verbs (e.g., *push, bite, hit*) tend to show higher scores on measures of concreteness, imageability or perceptual strength (e.g., Connell & Lynott, 2012) than lower-affectedness verbs (e.g., *know, miss, remember*), which tend to be more abstract, and generally refer to mental states rather than actions. If this possibility is correct, then even if the active transitive construction has no particular semantics, we would expect to see an (apparent) semantic affectedness effect for actives, merely by dint of the fact that low-affectedness verbs receive lower grammatical acceptability ratings across the board. To rule out this possibility, we would need to see, for every language that shows a semantic affectedness effect for passives, a smaller semantic affectedness effect for at least one of (a) actives and (b) pseudo-passives (defined in more detail below).

The original English study of Ambridge et al. (2016) did find evidence for an interaction of affectedness by construction type, such that the (continuous) semantic affectedness effect was greater for passives than actives, even when controlling for the frequency of each verb in passive and active constructions (no pseudo-passive construction was included). Likewise, in the present study, we conclude that the passive construction has the meaning of affectedness *only when we see a significantly larger affectedness effect for passive than pseudo-passive and/or active constructions*.

1.2 Replications of Ambridge et al. (2016, English) in Indonesian, Mandarin and Balinese

Subsequently, three replication studies of Ambridge et al.’s (2016) English study have been conducted for different languages. First, for Indonesian, Aryawibawa and Ambridge (2018) found

semantic affectedness effects for both passives (1) and actives (2), but not for a noncanonical “pseudo-passive” construction (3) that shares passive word order, but not the morphosyntactic hallmarks of the passive construction, specifically the verbal passive prefix *di-* and the preposition *oleh*, ‘by’.

- (1) Ayah **di-**tendang **oleh** ibu. (Passive)
 father PASS-kick by mother
 ‘The father was kicked by the mother.’
- (2) Ibu **men-**endang ayah. (Active)
 mother ACTIVE-kick father
 ‘The mother kicked the father.’
- (3) Ayah, ibu tending. (Pseudo-passive)
 father mother kicked
 ‘As for the father, the mother kicked him.’

Second, for Mandarin, Liu and Ambridge (2021) found semantic affectedness effects for passives (4) and BA-actives ((5); a dedicated affectedness construction), but not for standard actives (6), or noncanonical “pseudo-passives”/“notional passives” (7), which again share passive word order, but lack the passive morphosyntactic marker, *bei* (*le* marks completive aspect).

- (4) Lisi **bei** Zhangsan jiu le. (Passive)
 Lisi PASS Zhangsan save COMPL
 ‘Lisi was saved by Zhangsan.’
- (5) Zhangsan ba Lisi jiu le. (BA-Active).
 Zhangsan ACTIVE Lisi save COMPL
 ‘Zhangsan saved Lisi.’
- (6) Zhangsan jiu Lisi le. (Active)
 Zhangsan save Lisi COMPL
 ‘Zhangsan saved Lisi.’
- (7) Lisi Zhangsan jiu le.
 Lisi Zhangsan save COMPL
 ‘As for Lisi, Zhangsan saved her.’

Third, for Balinese, Darmasetiyawan and Ambridge (2022) found semantic affectedness effects for actives (8) and three different types of passives (*-a-*, *ka-* and *ma-* passives; (9–11)), but again not for noncanonical “pseudo-passives”/“basic passives” (12), which lack passive morphosyntax (*-a/ka-/ma-*), but do include the preposition *teken*, ‘by’.

- (8) Nak muani ento **n-ulud** nak luh ento. (Active)
 person male that ACTIVE-push person female that
 ‘The man pushed the woman.’
- (9) Nak luh ento tulud-**a** **teken** nak muani ento. (Passive)
 person female that push-PASS by person male that
 ‘The woman was pushed by the man.’
- (10) Nak luh ento **ka-tulud** **teken** nak muani ento. (Passive)
 person female that PASS-push by person male that
 ‘The woman was pushed by the man.’
- (11) Nak luh ento **ma-tulud** **teken** nak muani ento. (Passive)
 person female that PASS-push by person male that
 ‘The woman was pushed by the man.’
- (12) Nak luh ento tulud **teken** nak muani ento. (Pseudo-Passive)
 person female that push by person male that
 ‘As for the woman, the man pushed her.’

The differences between *-a*, *ka-* and *ma-* passives (essentially whether the agent is volitional, non-volitional or unimportant) are not relevant for our purposes here; we treat them all as bona-fide morphosyntactic (as opposed to pseudo-) passives.

Thus, a crosslinguistic pattern seems to be emerging, whereby semantic affectedness effects are always observed for genuine passives, but are attenuated – or absent altogether – for “pseudo-passives” (also called “noncanonical” or “basic” passives in the original papers) – topicalization constructions that share passive word order but lack passive morphosyntax (e.g., Keenan & Dryer, 2007) – and, to a lesser extent, actives (which generally seem to share some degree of affectedness semantics with passives).

In the present study, we extend this paradigm to Hebrew, and thereby to a fourth language family (Indonesian and Balinese are both Austronesian languages). We then conduct a Bayesian mixed-effects meta-analytic synthesis to investigate whether, across languages, semantic affectedness are (i) observed when looking only at passives (a relatively lenient criterion) and – crucially – (ii) bigger for passives than at least one of (a) pseudo-passives and (b) actives (a strict criterion; see Section 1).

2. Study 1: Hebrew

Although Hebrew has a non-passive topicalization construction, it is used only for cases of contrast or correction (Shlonsky, 2014), and would therefore be infelicitous in the context of the present study, which involves rating isolated sentences that are not part of a larger discourse. As

in the English study of Ambridge et al. (2016) – and unlike the Mandarin, Indonesian and Balinese studies discussed above – we therefore consider only the active and passive constructions, which are analogous to their English equivalents, as per the following examples from Ravid and Vered (2017, p. 1313)

- (13) ha-texnay hiklit et ha-re'ayon. (Active)
 DEF-technician recorded[ACTIVE] ACC DEF-interview
 'The technician recorded the interview.'
- (14) ha-re'ayon huklat al-yedey ha-texnay. (Passive)
 DEF-interview recorded[PASS] by-hand DEF-technician
 'The interview was recorded by (the hand of) the technician.'

As this example shows, the active and passive forms of the verb (*hiklit*, *huklat*) vary according to the binyan system that is characteristic of Hebrew (and other Semitic languages). These details are of minor importance for our purposes, since the present study focusses on construction semantics rather than verbal morphology. In brief, each verb root (**l-m-d** 'learn', **s-b-r** 'explain', **t-p-l** 'take care of') can appear in different binyanim (templates) with accordingly different semantics. There are several active binyanim (*q'al*, *hif'il* and *pi'el*) and several passive binyanim (*nif'al*, *huf'al* and *pu'al*). Not all active verbs have passive counterparts, and a given root can appear in multiple active binyanim with different (but related) meanings. The passive form is created by using a root with one of the passive binyanim. Thus, in the past tense, the active/passive forms of the example roots above (from Ravid & Vered, 2017, p. 1315) are *lamad/nilmad* ('learned/was learned'), *hisbir/husbar* ('explained/was explained') and *tipel/tupal* ('took care of/was taken care of'). This morphological system is semi-productive, and some roots may be coerced into nonstandard binyanim for special discourse-pragmatic effects (or erroneously by children; adultlike command of the system is achieved only by adolescence; Ravid & Saban, 2008). But since this would constitute a confound in the present study, we use – for any given verb – only the active binyan form and the passive binyan form that gives the “core” meaning.

2.1 Method (Study 1: Hebrew)

2.1.1 Participants

Sixty native speakers of Hebrew completed the grammaticality judgment task. This task took place online using the Gorilla platform with participants recruited using Sona Systems. A further 16 native speakers of Hebrew completed the semantic rating task, which involved completing an Excel spreadsheet offline, via email (simply because this was more convenient for the experimenters than coding an online task). All participants gave informed consent, and the study was approved by ethics committees at the University of Liverpool and the Hebrew University of Jerusalem.

2.1.2 Verbs

Fifty-six Hebrew translation equivalents of 72 “core” verbs from Ambridge et al. (2016) were chosen for use in the study (16 were removed because they did not have unique translations in Hebrew). These verbs are shown in **Table 1**, which also shows the active and passive binyan for each, as well as the total counts for each binyan. The same 56 verbs were used across the semantic rating and grammaticality judgment tasks described below.

Table 1: Hebrew verbs and binyanim used in the present study.

English translation	Active	Active (Romanized)	Active Binyan	Passive	Passive (Romanized)	Passive Binyan
kiss	נישק	nishek	Piel	נושק	nushak	Pual
entertain	בידר	bider	Piel	בודר	budar	Pual
believe	האמין	heemin	Hifil	הואמן	huaman	Hufal
admire	העריץ	heerits	Hifil	הוערץ	huarats	Hufal
sadden	העציב	he'etsiv	Hifil	הועצב	huatsav	Hufal
disturb	הפריע	hifria	Hifil	הופרע	hufra	Hufal
frighten	הפחיד	hifxid	Hifil	הופחד	hufxad	Hufal
know	הכיר	hekir	Piel	הוכר	hukar	Hufal
hit	הכה	hika	Piel	הוכה	hukar	Pual
tease	הקניט	hiknit	Hifil	הוקנט	huknat	Hufal
dress	הלביש	hilibish	Hifil	הולבש	hulbash	Hufal
drop	הפיל	hepil	Hifil	הופל	hupal	Hufal
calm	הרגיע	hirgia	Hifil	הורגע	hurga	Hufal
irritate	הרגיז	hurgiz	Hifil	הורגז	hurgaz	Hufal
distract	הסיח	hesiix	Piel	הוסח	husax	Hufal
bother	הטריד	hitrid	Hifil	הוטרד	hutrad	Hufal
lead	הוביל	hovil	Hifil	הובל	huval	Hufal

(Contd.)

English translation	Active	Active (Romanized)	Active Binyan	Passive	Passive (Romanized)	Passive Binyan
understand	הבין	hevin	Piel	הובן	huvan	Hufal
terrify	הבעית	hivit	Hifil	הובעת	huvat	Hufal
notice	הבחין	hivxin	Hifil	הובחן	huvxan	Hufal
hold	החזיק	hexzik	Hifil	הוחזק	huxzak	Hufal
teach	לימד	limed	Piel	לומד	lumad	Pual
pat	ליטף	lited	Piel	לוטף	lutaf	Pual
love	אהב	ahav	Paal	נאהב	neehav	nifal
follow	עקב	akav	Paal	נעקב	neekav	nifal
eat	אכל	axal	Paal	נאכל	neexal	nifal
help	עזר	azar	Paal	נעזר	neezar	nifal
cut	חתך	xatax	Paal	נחתך	nextax	nifal
amaze	הדהים	hidhim	Hifil	נדהם	nidham	nifal
push	דחף	daxaf	Paal	נדחף	nidxaf	nifal
disgust	הגעיל	higil	Hifil	נגעל	nigal	nifal
pull	משך	mashax	Paal	נמשך	nimshax	nifal
crush	מחץ	maxats	Paal	נמחץ	nimxats	nifal
bite	נשך	nashax	Paal	ננשך	ninshax	nifal
see	ראה	raa	Paal	נראה	niraa	nifal
chase	רדף	radaf	Paal	נרדף	nirdaf	nifal
forgot	שכח	shachax	Paal	נשכח	nishkax	nifal
dislike	סלד	salad	Paal	נסלד	nislad	nifal
hate	שנא	sana	Paal	נשנא	nisna	nifal
lookat	הסתכל	histakel	hitpaal	נסתכל	nistakel	niftael

(Contd.)

English translation	Active	Active (Romanized)	Active Binyan	Passive	Passive (Romanized)	Passive Binyan
carry	סחב	saxav	Paal	נסחב	nisxav	nifal
ignore	התעלם	hitalem	hitpael	נתעלם	nitalem	niftael
miss	התגעגע	hitga'agea	hitpael	נתגעגע	nitgaagea	niftael
watch	צפה	tsafa	Paal	נצפה	nitspa	nifal
kick	בעט	baat	Paal	נבעט	nivaat	nifal
scare	הבהיל	hivhil	Hifil	נבהל	nivhal	nifal
trust	בטח	batax	Paal	נבטח	nivtax	nifal
remember	זכר	zaxar	Paal	נזכר	nizkar	nifal
cheerup	עודד	oded	Paal	עודד	oded	Pual
please	ריצה	ritsa	Piel	רוצה	rutsa	Pual
amuse	שיעשע	shiashea	Piel	שועשע	shuasha	Pual
annoy	עיצבן	itsben	Piel	עוצבן	utsban	Pual
hug	חיבק	xibek	Piel	חובק	xubak	Pual
like	חיבב	xibev	Piel	חובב	xubav	Pual
shock	זיעזע	ziazea	Piel	זועזע	zuaza	Pual
recognize	זיהה	zihaa	Piel	זוהה	zuhaa	Pual
		Total Piel	15		Total Pual	13
		Total Hifil	18		Total Hufal	18
		Total Paal	20		Total nifal	22
		Total hitpael	3		Total niftael	3

2.1.3 Semantic rating task

Participants were given (via email) an Excel spreadsheet that listed, in random order, each verb, in active form only, and each of 10 semantic properties (based on Ambridge et al., 2016; based in turn on Pinker, 1989; Pinker et al., 1987):

(a) It would be possible for A to deliberately [VERB] B, (b) B undergoes a change in state or circumstances, (c) B is responsible, (d) B is affected in some way, (e) B is adversely (negatively) affected, (f) A makes physical contact with B, (g) A is responsible, (h) A is doing something to B, (i) A enables or allows the change/event, (j) A causes (or is responsible for) some effect/change that happens to B.

The instructions, in translation, read: “On the following sheet is a list of verbs. Each describes an event involving two people (or things, ideas, etc.), denoted by A and B. For example, if the verb is *damage*, the event would be *A damaged B*.” Note that, as in Ambridge et al. (2016), we obtained ratings for each of the (putative) semantic properties of the passive listed by Pinker (1989); not only those that are *prima facie* related to affectedness. Whether these properties reflect one or more underlying semantic dimensions we treat as an empirical question to be determined by Principal Components Analysis (PCA); again the same approach taken in Ambridge et al. (2016). Participants completed the spreadsheet with their ratings of the extent to which each verb exhibits each property, on a scale of 1–9. Importantly, participants completing the semantic rating task were not informed that the study related to passive sentences; all verbs were presented in active form only. For each semantic property, we took the mean rating across all 16 participants. We then combined these scores into a single “affectedness” score for each verb, using Principal Components Analysis (principal’ from the R package ‘psych’; Revelle, 2018). As noted by an anonymous reviewer, some of the original semantic properties are only tangentially related to affectedness per se (e.g., It would be possible for A to deliberately [VERB] B; A enables or allows the change/event; A causes (or is responsible for) some effect/change that happens to B). Nevertheless, a scree plot revealed a sharp inflection point after the first component (which explained 70% of variance), with the second component explaining only an additional 12.5%. Thus, the ratings data clearly suggest that the semantic properties of these verbs – at least insofar as they are captured by the 10 original semantic predictors – reflect a single underlying semantic dimension (which, for convenience, we refer to simply as *affectedness*).

2.1.4 Grammaticality judgment test

Each verb was presented in a single active sentence ($N = 56$) and a single passive sentence ($N = 56$) of the form *The woman pushed the man/The man was pushed by the woman*. All sentences had *the man* as the subject and *the woman* as the object, or vice versa: the arguments were reversed for two counterbalance versions. For example, half of the participants rated *The woman pushed the man* and *The man was pushed by the woman*, while half rated *The man pushed the woman* and *The woman was pushed by the man* (of course, in Hebrew translation). Therefore, just as in all the previous studies discussed in Section 1, both arguments were (human) animates. We can only speculate as to whether the present findings would generalize to non-animates; however, note

that the wider English study reported in Ambridge et al. (2016) – as opposed to just the “core set” of verbs discussed here – found broadly the same pattern for verbs that take inanimate arguments. Participants gave their ratings on a 5-point smiley face scale (originally designed for use with children). The scale had no text or numerical labels; instead, participants completed practice trials (with feedback), consisting of sentences that generally receive ratings of “perfect” (i.e., 5/5), “terrible” (i.e., 1/5), “good but not perfect” (4/5), “bad but not terrible” (2/5), and “OK; in the middle” (3/5). All sentences were presented in random order, via the online Gorilla platform. No filler sentence were used, as we saw no reason to hide the true purpose of the study: obtaining acceptability ratings for the same verbs in active and passive sentences. Since the participants had no particular expertise in linguistics, we can safely assume they were not aware of the hypothesis of a link between affectedness and passive – but not active – sentence acceptability. As in the previous studies discussed in Section 1, each sentence was accompanied by an animation showing the relevant action (these can be viewed at <https://app.gorilla.sc/openmaterials/426411>). The purpose of the animations was to make clear the intended meaning of each sentence and to maintain a comparable method across studies (some previous studies included children, for whom the animations were included to make the study more engaging). It is certainly the case that the verbs and animations vary with regard to concreteness and imageability, and in a way that could potentially affect sentence acceptability judgments. This is another reason why a strong test of the semantics hypothesis requires a *larger* semantic effect for passives than for pseudo-passives or actives, which function as control sentences with regard to between-verb factors such as concreteness, imageability and (as noted above) canonical versus noncanonical linking (e.g., *The woman feared [c.f., frightened] the man*).

2.1.5 Corpus counts

Previous studies using this paradigm (except Balinese, for which no corpus exists) have obtained corpus counts of each verb in active and passive sentences, to allow frequency to be “controlled-out” statistically. This notion is somewhat dubious, first because the frequency distributions in part reflect semantics (i.e., speakers tend not to use verbs in semantically incompatible constructions); second because predictors can only be “controlled-out” if they are measured perfectly (Westfall & Yarkoni, 2016). Nevertheless, for comparability with previous studies, we obtained corpus counts in order to create a verb-bias measure, reflecting the extent to which – relative to other verbs in the sample – each verb is biased towards or away from passive versus active uses.

Counts were obtained from the Hebrew Blog Corpus (<https://github.com/TalLinzen/hebrew-blog-corpus>), using a web-based search tool (which seems to have since been retired). Searches were performed by entering a root and a binyan (here, the relevant active and passive binyanim), which yields separate frequency counts for that root + binyan combination in all tense/person/

number/gender forms. Thus, we were able to exclude adjectival passives that, in some cases, take the same form as passive verbs. However, it is important to note that, due to the nature of this search procedure, the counts obtained are of active and passive verb *forms* only, and there was no way to take into account the wider syntactic structure of the relevant sentence (which would have required a fully-parsed corpus and a much more complex search). These counts are perfectly adequate for the present purposes, in that they reflect genuine passive and active frequency, but they preclude the possibility of using any finer-grained counts (as suggested by an anonymous reviewer). We must acknowledge, however, that – as noted by another anonymous reviewer – this counting method did not allow us to exclude potentially irrelevant senses (e.g., we included *bider/budar*, meaning ‘entertain’, but it also has the rarer meaning of ‘scatter’). The extent to which this is a problem depends on whether the different senses are unrelated homonyms or related polysemes (studies summarized by Rice et al., 2019, suggest that polysemes should be counted, but homonyms excluded). Thus, given the absence of objective ratings of homonymy versus polysemy, even laborious hand-coding would not necessarily yield any cleaner counts than the present automated method. Since the study was conducted in written form with literate adults, the use of a written, as opposed to spoken, corpus does not constitute a problem, and is perhaps even an advantage.

Following Liu and Ambridge (2020) and Stefanowitsch and Gries (2003), we calculated for each verb a chi-square statistic which measures the extent to which – relative to the other 55 verbs in the sample – it is biased towards (+) or away from (–) the passive construction versus all non-passive constructions. As an example, the calculation for *push* is shown in **Table 2**.

Table 2: Example calculation of the chi-square frequency predictor for *push*.

	Passive	Non-Passive
push	A (3253)	B (17,165)
55 other verbs (summed)	C (341,991)	D (2,542,114)

$$\frac{(A*D - B*C)^2 * (A + B + C + D)}{(A + C) * (B + D) * (A + B) * (C + D)}$$

This chi-square value (not the associated *p* value) – for *push*, 321.327 – is then log-transformed (5.776) and set to + or – depending on whether it is biased towards or away from passives (versus non-passives) compared to other verbs. For *push*, the final value was +5.776, reflecting a strong bias *towards* passives. Conversely for – for example – *believe*, the final value was –9.345, reflecting an even stronger bias *away from* passives. This method is not perfect, in that it assumes (incorrectly, of course) that our 56 verbs are a representative sample of the language in general.

Note, however, that, to some extent, all linguistic experiments suffer from a version of this problem, since all necessarily study just a small sample of the language.

2.2 Results and discussion (Study 1: Hebrew)

All data and analysis scripts are available at <https://osf.io/hwb26/>. **Figure 1** plots the raw sentence acceptability ratings on the 5-point scale (Y axis) against the scaled and centred continuous semantic predictor (X axis) for the present Hebrew data (and – for comparative purposes – for Balinese, English, Hebrew and Mandarin). Inspection of this plot suggests that, consistent with some of these previous studies, a semantic affectedness effect is observed for passives, but not for actives.

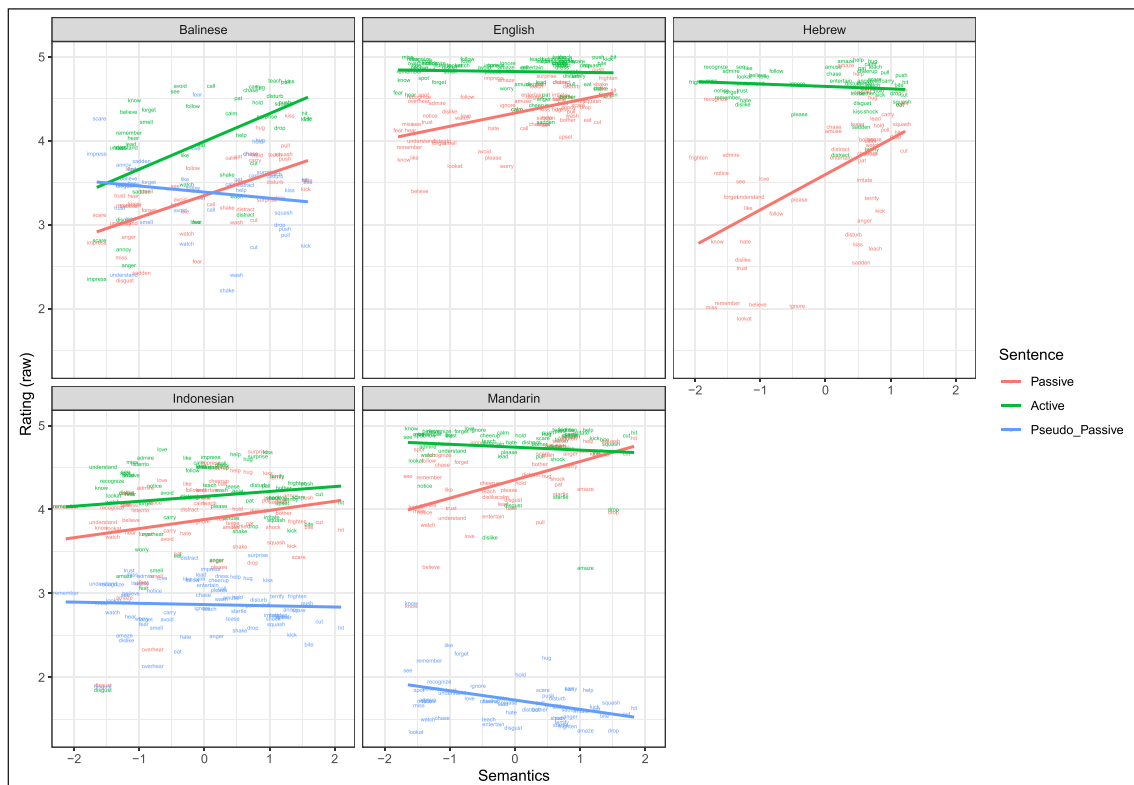


Figure 1: Semantics effects for Hebrew (also Balinese, English, Indonesian and Mandarin).

In order to test this observation statistically, we ran a series of mixed effects models with verb and participant as random factors. The main analysis used Bayesian models (brms package; Bürkner, 2018), in R (R Core Team, 2022), with a cumulative link function [family = cumulative()], reflecting the fact that the acceptability judgement data are on an ordinal scale (this reflects a methodological improvement over the previous papers reviewed above, which

treated the judgment data as continuous). Following Bürkner and Vuorre (2019), we used a wide, flat, Gaussian prior of (0,5), with all predictors centred and scaled (there is an argument for instead using a Dirichlet prior based on the anticipated distribution of ratings across the five response categories, but this is not implemented in brms). Following the suggestion of an anonymous reviewer, we adopted fully-maximal random effects structure (in the sense of Barr et al., 2013) for all models, in order to ensure maximum generalizability (particularly for the subsequent meta-analysis). That is, we included all random intercepts and slopes that are justified given the design. In any case, frequentist model selection using the MixedModels.jl package (Bates, 2016; Bates et al., 2016) in Julia (Bezanson et al., 2012) revealed that, for the main semantics-only model, a fully-maximal model was indeed optimal.

$$\text{Response} \sim \text{Sentence_Type} * \text{Semantics} + (1 + \text{Sentence_Type} * \text{Semantics} | \text{Participant}) + (1 + \text{Sentence_Type} | \text{verb})^1$$

where “Response” is participants’ ratings on the 5-point scale, “Sentence_Type” is the sentence being rated [active = 1, passive = 0] and “Semantics” is the scaled and centred continuous semantic affectedness predictor. (Note that in the actual R syntax and output tables reproduced below, Sentence_Type is abbreviated to S_Type). However, for the secondary semantics-and-frequency model, frequentist model comparison revealed that a slightly simpler model was preferred according to both the AICc and BIC criteria (Matuschek et al., 2017):

$$\text{Response} \sim \text{Sentence_Type} * \text{Semantics} + \text{Sentence_Type} * \text{Frequency} + (1 + \text{Sentence_Type} * \text{Semantics} | \text{Participant}) + (1 + \text{Sentence_Type} | \text{verb})$$

(where “Frequency” is the scaled and centred continuous chi-square verb bias predictor). However, as noted above, to ensure maximum generalizability we did not run this model, but the maximal model with a by-participant random slope for the interaction of Sentence_Type by frequency:

$$\text{Response} \sim \text{Sentence_Type} * \text{Semantics} + \text{Sentence_Type} * \text{Frequency} + (1 + \text{Sentence_Type} * \text{Semantics} + \text{Sentence_Type} * \text{Frequency} | \text{Participant}) + (1 + \text{Sentence_Type} | \text{verb})$$

¹ We also ran a version of the model that controlled for the Active and Passive binyan of each verb form, by including a by-binyan random slope for the effect of semantics (i.e., adding the random effects terms $(1 + \text{Semantics} | \text{Heb_Active_Binyan}) + (1 + \text{Semantics} | \text{Heb_Passive_Binyan})$ to the model shown in the main text). Note that we did not allow Sentence Type (Active/Passive) – or its interaction with Semantics – to vary by binyan, since binyan is perfectly correlated with Sentence Type (since there is no overlap between active binyanim and passive biyanim). However, the model that controlled for binyan provided a worse fit on both criteria (AICc = 17768, BIC = 17932) than the model without (AICc = 17755, BIC = 17878). Thus, we reverted to the simpler model.

All Bayesian models were run for 5,000 iterations with a 2,000 iteration warm-up, with 16 chains running on 16 cores, with `adapt_delta` set to 0.99.

The main semantics-only model and the secondary semantics-and-frequency models are shown in Tables 3 and 4, respectively. The credible intervals shown are 95% intervals for the main model terms (which are two-sided) and 90% intervals for the one-sided hypothesis tests (shown in bold and explained below). Posterior probabilities are shown in the column “ $B < 0$ ”. Although an advantage of the Bayesian approach is that it discourages dichotomous statistical thinking, readers who want to ask “Is this effect significant?” may interpret these values as “Bayesian p values” by mentally subtracting them from 1. Indeed (and perhaps somewhat ironically given the “statistics wars” between frequentist and Bayesian approaches), with a completely uniform prior and a one-sided test, Bayesian posterior probabilities ($1-X$) and frequentist p values are numerically identical (Marsman & Wagenmakers, 2017).

Table 3: Semantics-only model for Hebrew.

Covariate	Estimate	Est.Error	CI_lower	CI_upper	$B < 0$
Intercept[1]	-3.12	0.24	-3.59	-2.66	1
Intercept[2]	-1.7	0.23	-2.16	-1.24	1
Intercept[3]	-0.48	0.23	-0.93	-0.03	0.98
Intercept[4]	0.78	0.23	0.33	1.23	1
S_TypeActive	3.05	0.27	2.53	3.58	1
Semantics	0.79	0.19	0.42	1.16	1
S_TypeActive:Semantics	-0.81	0.2	-1.21	-0.42	1
Semantics_Slope_for_Active	-0.03	0.17	-0.3	0.25	0.43
Semantics_Slope_for_Passive	0.79	0.19	0.48	1.09	1

Table 4: Semantics-and-frequency model for Hebrew.

Covariate	Estimate	Est.Error	CI_lower	CI_upper	$B < 0$
Intercept[1]	-3.26	0.25	-3.75	-2.77	1
Intercept[2]	-1.83	0.25	-2.32	-1.36	1
Intercept[3]	-0.62	0.24	-1.1	-0.14	0.99

(Contd.)

Covariate	Estimate	Est.Error	CI_lower	CI_upper	B < 0
Intercept[4]	0.64	0.24	0.16	1.12	0.99
S_TypeActive	2.82	0.28	2.27	3.38	1
Semantics	0.76	0.18	0.41	1.12	1
Frequency	0.37	0.2	-0.02	0.76	0.97
S_TypeActive:Semantics	-0.78	0.19	-1.15	-0.41	1
S_TypeActive:Frequency	-0.28	0.32	-0.91	0.34	0.82
Semantics_Slope_for_Active	-0.02	0.17	-0.29	0.25	0.45
Semantics_Slope_for_Passive	0.76	0.18	0.46	1.06	1
Frequency_Slope_for_Active	0.09	0.17	-0.2	0.38	0.7
Frequency_Slope_for_Passive	0.37	0.2	0.05	0.7	0.97

Since we used treatment coding (active = 1, passive = 0), the estimate for “Semantics” represents the estimated effect of the semantic affectedness predictor for passive sentences. The estimate for the “S_TypeActive:Semantics” interaction represents the *difference* between the effect of the semantics predictor for actives versus passives. This term is large and in the predicted (negative) direction for both models, indicating that the effect of semantics is greater for passive than active sentences. Note that this term represents the *difference* between the semantics effects for passives versus actives, *not* an estimate of the semantics effect for either passives or actives independently. In order to obtain these estimates, we used the “hypothesis” function of brms to perform a directional (i.e., one sided) test of the hypothesis that the effect of semantics is greater than zero for (a) actives and (b) passives (although this is just a one-sided version of the estimate for passives already described; note that the estimates are identical). Conceptually, this is equivalent to the more common practice of breaking down a significant interaction by running separate follow-up models for – in this case – (a) active sentences only and (b) passive sentences only. However, the present method of estimating the relevant contrast directly from the main model is preferable (e.g., Schad et al., 2020) for two reasons. First, estimation of variance and random effects terms is more stable compared to fitting individual models to subsets of the data (Schad, personal communication), since – in the latter case – each model is fitted to only half of the data, which decreases the reliability of parameter estimates (e.g., Kerkhoff & Nussbeck, 2019). Second, running separate models on subsets of the data increases the number of tested effects. Thus, unless we are rigorous about ignoring any effect that was not specifically predicted, we run the risk of turning up – and seeking to interpret – spurious chance findings. These values are shown in the bold rows of **Tables 3–4**. This analysis revealed that the semantics slope is

large and in the predicted direction for both the semantics-only (**Table 3**) and semantics-and-frequency (**Table 4**) models (i.e., whether or not frequency is included in the model). Indeed, the correlation between the frequency predictor and the semantic affectedness predictor was only $r = 0.1$, which explains why controlling out this factor makes little difference.

Following the same logic, this analysis also revealed that the effect of frequency (see Table 4) is (a) moderate for passives (row “Frequency” for the two-sided test; row “Frequency_Slope_for_Passive” for the one-sided test), (b) marginally smaller for actives than passives (row “S_TypeActive:Frequency”), and (c) essentially zero for actives (“Frequency_Slope_for_Active”, one-sided test).

Before summing up, we pause to consider a potential objection; that our semantics manipulation is confounded with concreteness – or perhaps perceptual strength (e.g., Connell & Lynott, 2012) – which affects ease of processing. It is certainly true that our high-affectedness verbs (e.g., *push*, *bite*, *hit*) would also score highly on measures of concreteness, imageability or perceptual strength, while our low-affectedness verbs (e.g., *know*, *miss*, *remember*) would not. However, this shows the value of including active sentences (and, for some other languages, pseudo-passive sentences) as a control condition. For Hebrew, the correlation between affectedness and active sentence acceptability is not even in the direction one would expect if affectedness were a proxy for concreteness-related ease of processing. Thus, it does not seem that low-affectedness verbs show low acceptability in the passive merely because their low concreteness induces processing difficulties.

In summary, the findings for Hebrew were, if anything, even more clear-cut than for the languages previously studied: Verb-by-verb semantic affectedness ratings predict verbs’ acceptability in passive, but not active, sentences; and attempting to control for verbs’ relative frequency in the passive construction makes very little difference. To investigate the extent to which this holds true across languages, we now turn to our meta-analytic synthesis.

3. Study 2: Meta-analytic synthesis

Before we begin, some terminological clarification is in order. The reason we call Study 2 a “meta-analytic synthesis” rather than simply a “meta-analysis” is that it lacks some of the key elements of the latter, such as an exhaustive literature search and publication-bias analyses. Since the raw trial-level data for all of the previous studies included in this synthesis are publicly available online, we do not need to resort to the more familiar – but less precise – meta-analytic technique of aggregating across studies using an effect-size measure. Instead, we are able to build mixed-effects models that include the raw data from all languages, treating language as a random effect. Following the standard logic of mixed-effects models, this allows us to investigate whether any semantic (and frequency) effects observed hold across the sampled languages (and, hence, may be potentially generalizable to other languages with similar constructions).

3.1 Individual languages model

Before proceeding to this all-languages model, however, we first built a separate model for each language. This allows us to “correct” for some small differences between the analyses in the published papers such as different priors, different ways of estimating separate effects for different sentence types, different instantiations of the frequency predictor and (for Balinese) the use of a 10-point, rather than a 5-point, scale (the ratings are re-scaled for the analyses presented here). We also used a cumulative link function throughout, reflecting the fact that the acceptability judgement data are on an ordinal scale (a methodological improvement over the original analyses, which treated the judgment data as continuous). As already noted above, we used maximal random-effects structure throughout, in order to ensure maximum generalizability. Treatment coding was used throughout with Passive designated as the reference category (i.e., 0). This allows us to test whether, as compared with Passives, the effect of Semantics (or Frequency) is smaller for (a) Actives and (b) Pseudo Passives; via the individual interaction terms *S_TypeActive:Semantics* and *S_TypePseudo_Passive:Semantics*, respectively (for Frequency, *S_TypeActive:Frequency* and *S_TypePseudo_Passive:Frequency*). In order to allow for comparable analyses across languages and sentence types, we collapsed Balinese *-a*, *ka-* and *ma-* passives as Passive (a supplementary model retaining the different sentence types is available on the OSF project site).

Table 5: Semantics-only models for each language separately.

(a) Balinese					
Covariate	Estimate	Est.Error	CI_lower	CI_upper	B <> 0
Intercept[1]	-1.46	0.12	-1.7	-1.22	1
Intercept[2]	-0.91	0.12	-1.15	-0.66	1
Intercept[3]	-0.25	0.12	-0.49	-0.01	0.98
Intercept[4]	0.63	0.12	0.38	0.87	1
<i>S_TypeActive</i>	1.16	0.16	0.85	1.47	1
<i>S_TypePseudo_Passive</i>	0.03	0.17	-0.3	0.37	0.58
Semantics	0.38	0.08	0.23	0.53	1
<i>S_TypeActive:Semantics</i>	0.33	0.13	0.08	0.59	1
<i>S_TypePseudo_Passive:Semantics</i>	-0.53	0.13	-0.79	-0.27	1

(Contd.)

Semantics_Slope_for_Active	0.71	0.16	0.45	0.97	1
Semantics_Slope_for_Passive	0.38	0.08	0.25	0.51	1
Semantics_Slope_for_Basic_Passive	-0.15	0.12	-0.34	0.05	0.11
(b) English					
Covariate	Estimate	Est.Error	CI_lower	CI_upper	B <math>\diamond</math> 0
Intercept[1]	-6.51	0.38	-7.26	-5.77	1
Intercept[2]	-4.13	0.31	-4.74	-3.52	1
Intercept[3]	-2.58	0.3	-3.18	-1.99	1
Intercept[4]	-0.22	0.3	-0.82	0.36	0.78
S_TypeActive	2.54	0.26	2.04	3.05	1
Semantics	0.53	0.11	0.33	0.74	1
S_TypeActive:Semantics	-0.54	0.18	-0.89	-0.19	1
Semantics_Slope_for_Active	-0.01	0.18	-0.31	0.29	0.48
Semantics_Slope_for_Passive	0.53	0.11	0.36	0.7	1
(c) Hebrew					
Covariate	Estimate	Est.Error	CI_lower	CI_upper	B <math>\diamond</math> 0
Intercept[1]	-3.12	0.24	-3.59	-2.66	1
Intercept[2]	-1.7	0.23	-2.16	-1.24	1
Intercept[3]	-0.48	0.23	-0.93	-0.03	0.98
Intercept[4]	0.78	0.23	0.33	1.23	1
S_TypeActive	3.05	0.27	2.53	3.58	1
Semantics	0.79	0.19	0.42	1.16	1
S_TypeActive:Semantics	-0.81	0.2	-1.21	-0.42	1
Semantics_Slope_for_Active	-0.03	0.17	-0.3	0.25	0.43
Semantics_Slope_for_Passive	0.79	0.19	0.48	1.09	1

(Contd.)

(d) Indonesian					
Covariate	Estimate	Est.Error	CI_lower	CI_upper	B <math>\lt; \gt; 0</math>
Intercept[1]	-5.17	0.22	-5.59	-4.74	1
Intercept[2]	-3.33	0.22	-3.76	-2.91	1
Intercept[3]	-1.25	0.21	-1.67	-0.82	1
Intercept[4]	1.29	0.21	0.87	1.71	1
S_TypeActive	0.88	0.16	0.57	1.21	1
S_TypeNon_Canonical_Passive	-2.55	0.28	-3.1	-2.01	1
Semantics	0.27	0.17	-0.06	0.6	0.95
S_TypeActive:Semantics	-0.07	0.08	-0.23	0.09	0.82
S_TypeNon_Canonical_Passive:Semantics	-0.34	0.1	-0.53	-0.15	1
Semantics_Slope_for_Active	0.2	0.17	-0.08	0.48	0.88
Semantics_Slope_for_Passive	0.27	0.17	-0.01	0.55	0.95
Semantics_Slope_for_Non_Canonical_Passive	-0.07	0.11	-0.25	0.11	0.27
(e) Mandarin					
Covariate	Estimate	Est.Error	CI_lower	CI_upper	B <math>\lt; \gt; 0</math>
Intercept[1]	-6.58	0.25	-7.08	-6.09	1
Intercept[2]	-4.46	0.25	-4.95	-3.98	1
Intercept[3]	-2.59	0.24	-3.08	-2.12	1
Intercept[4]	-0.5	0.24	-0.98	-0.03	0.98
S_TypeActive	1.82	0.25	1.33	2.31	1
S_TypeBA_Active	-2.11	0.21	-2.53	-1.69	1
S_TypeNotional_Passive	-6.95	0.36	-7.67	-6.25	1
Semantics	0.8	0.15	0.5	1.1	1

(Contd.)

Covariate	Estimate	Est.Error	CI_lower	CI_upper	B $\diamond 0$
S_TypeActive:Semantics	-0.8	0.21	-1.21	-0.38	1
S_TypeBA_Active:Semantics	1.15	0.18	0.79	1.51	1
S_TypeNotional_Passive:Semantics	-1.09	0.18	-1.45	-0.73	1
Semantics_Slope_for_Active	0	0.22	-0.35	0.36	0.5
Semantics_Slope_for_Passive	0.8	0.15	0.55	1.05	1
Semantics_Slope_for_Notional_Passive	-0.29	0.12	-0.49	-0.09	0.99
Semantics_Slope_for_BA_Actives	1.95	0.23	1.57	2.32	1

Table 6: Semantics-and-frequency models for each language separately.

(a) English					
Covariate	Estimate	Est.Error	CI_lower	CI_upper	B $\diamond 0$
Intercept[1]	-6.51	0.38	-7.26	-5.78	1
Intercept[2]	-4.13	0.31	-4.73	-3.52	1
Intercept[3]	-2.58	0.3	-3.16	-1.99	1
Intercept[4]	-0.22	0.29	-0.8	0.36	0.78
S_TypeActive	2.55	0.25	2.06	3.06	1
Semantics	0.42	0.1	0.23	0.61	1
Frequency	0.4	0.09	0.22	0.58	1
S_TypeActive:Semantics	-0.44	0.18	-0.81	-0.08	0.99
S_TypeActive:Frequency	-0.48	0.22	-0.9	-0.05	0.99
Semantics_Slope_for_Active	-0.02	0.19	-0.33	0.28	0.45

(Contd.)

Semantics_Slope_for_Passive	0.42	0.1	0.26	0.58	1
Frequency_Slope_for_Active	-0.08	0.18	-0.36	0.21	0.33
Frequency_Slope_for_Passive	0.4	0.09	0.25	0.55	1
(b) Hebrew					
Covariate	Estimate	Est.Error	CI_lower	CI_upper	B < 0
Intercept[1]	-3.26	0.25	-3.75	-2.77	1
Intercept[2]	-1.83	0.25	-2.32	-1.36	1
Intercept[3]	-0.62	0.24	-1.1	-0.14	0.99
Intercept[4]	0.64	0.24	0.16	1.12	0.99
S_TypeActive	2.82	0.28	2.27	3.38	1
Semantics	0.76	0.18	0.41	1.12	1
Frequency	0.37	0.2	-0.02	0.76	0.97
S_TypeActive:Semantics	-0.78	0.19	-1.15	-0.41	1
S_TypeActive:Frequency	-0.28	0.32	-0.91	0.34	0.82
Semantics_Slope_for_Active	-0.02	0.17	-0.29	0.25	0.45
Semantics_Slope_for_Passive	0.76	0.18	0.46	1.06	1
Frequency_Slope_for_Active	0.09	0.17	-0.2	0.38	0.7
Frequency_Slope_for_Passive	0.37	0.2	0.05	0.7	0.97
(d) Indonesian					
Covariate	Estimate	Est.Error	CI_lower	CI_upper	B < 0
Intercept[1]	-5.18	0.22	-5.62	-4.75	1
Intercept[2]	-3.34	0.22	-3.78	-2.92	1
Intercept[3]	-1.26	0.22	-1.69	-0.83	1
Intercept[4]	1.28	0.22	0.85	1.7	1
S_TypeActive	0.88	0.16	0.57	1.2	1

(Contd.)

S_TypeNon_Canonical_Passive	-2.55	0.28	-3.11	-2.01	1
Semantics	0.27	0.17	-0.06	0.6	0.95
Frequency	-0.05	0.14	-0.32	0.22	0.64
S_TypeActive:Semantics	-0.07	0.08	-0.23	0.09	0.81
S_TypeNon_Canonical_Passive:Semantics	-0.34	0.1	-0.53	-0.15	1
S_TypeActive:Frequency	0.03	0.08	-0.12	0.18	0.65
S_TypeNon_Canonical_Passive:Frequency	0	0.09	-0.18	0.18	0.5
Semantics_Slope_for_Active	0.2	0.17	-0.08	0.49	0.88
Semantics_Slope_for_Passive	0.27	0.17	0	0.55	0.95
Semantics_Slope_for_Non_Canonical_Passive	-0.07	0.11	-0.25	0.11	0.27
Frequency_Slope_for_Active	-0.02	0.14	-0.26	0.22	0.45
Frequency_Slope_for_Passive	-0.05	0.14	-0.27	0.18	0.36
Frequency_Slope_for_Non_Canonical_Passive	-0.05	0.08	-0.18	0.08	0.27
(e) Mandarin					
Covariate	Estimate	Est.Error	CI_lower	CI_upper	B <> 0
Intercept[1]	-6.59	0.25	-7.09	-6.09	1
Intercept[2]	-4.47	0.25	-4.96	-3.98	1
Intercept[3]	-2.6	0.25	-3.09	-2.12	1
Intercept[4]	-0.51	0.24	-0.99	-0.03	0.98
S_TypeActive	1.77	0.26	1.27	2.28	1
S_TypeBA_Active	-1.9	0.22	-2.34	-1.47	1
S_TypeNotional_Passive	-6.93	0.37	-7.65	-6.22	1
Semantics	0.69	0.17	0.37	1.02	1

(Contd.)

Frequency	0.23	0.13	-0.03	0.49	0.96
S_TypeActive:Semantics	-0.65	0.23	-1.11	-0.2	1
S_TypeBA_Active:Semantics	0.94	0.2	0.54	1.34	1
S_TypeNotional_Passive:Semantics	-0.98	0.19	-1.36	-0.61	1
S_TypeActive:Frequency	-0.16	0.21	-0.56	0.24	0.78
S_TypeBA_Active:Frequency	0.49	0.23	0.03	0.95	0.98
S_TypeNotional_Passive:Frequency	-0.1	0.19	-0.48	0.26	0.71
Semantics_Slope_for_Active	0.04	0.24	-0.35	0.43	0.58
Semantics_Slope_for_Passive	0.69	0.17	0.42	0.97	1
Semantics_Slope_for_Notional_Passive	-0.29	0.12	-0.49	-0.09	0.99
Semantics_Slope_fpr_BA_Actives	1.64	0.25	1.23	2.04	1
Frequency_Slope_for_Active	0.06	0.15	-0.19	0.31	0.67
Frequency_Slope_for_Passive	0.23	0.13	0.01	0.44	0.96
Frequency_Slope_for_Notional_Passive	0.12	0.14	-0.11	0.35	0.82
Frequency_Slope_fpr_BA_Actives	0.72	0.23	0.35	1.09	1

As for the Hebrew models described earlier, we again ran separate semantics-only and semantics-and-frequency models. The logic of including frequency here is to control for its effect, in order to allow us to investigate the effect of semantic affectedness above and beyond any effect of frequency (at least in principle; see Westfall & Yarkoni, 2016, for caveats). In order to be maximally conservative in this regard, we always include frequency in each semantics-and-frequency model, regardless of whether or not it reaches any particular statistical criterion. However, the semantics-and-frequency models must be interpreted with caution, since – in principle – there is likely to be collinearity between the semantic affectedness and frequency predictors (i.e., speakers tend to use verbs in only semantically compatible constructions). In practice, this correlation is worryingly high only for Mandarin ($r = 0.48$) and, to some extent,

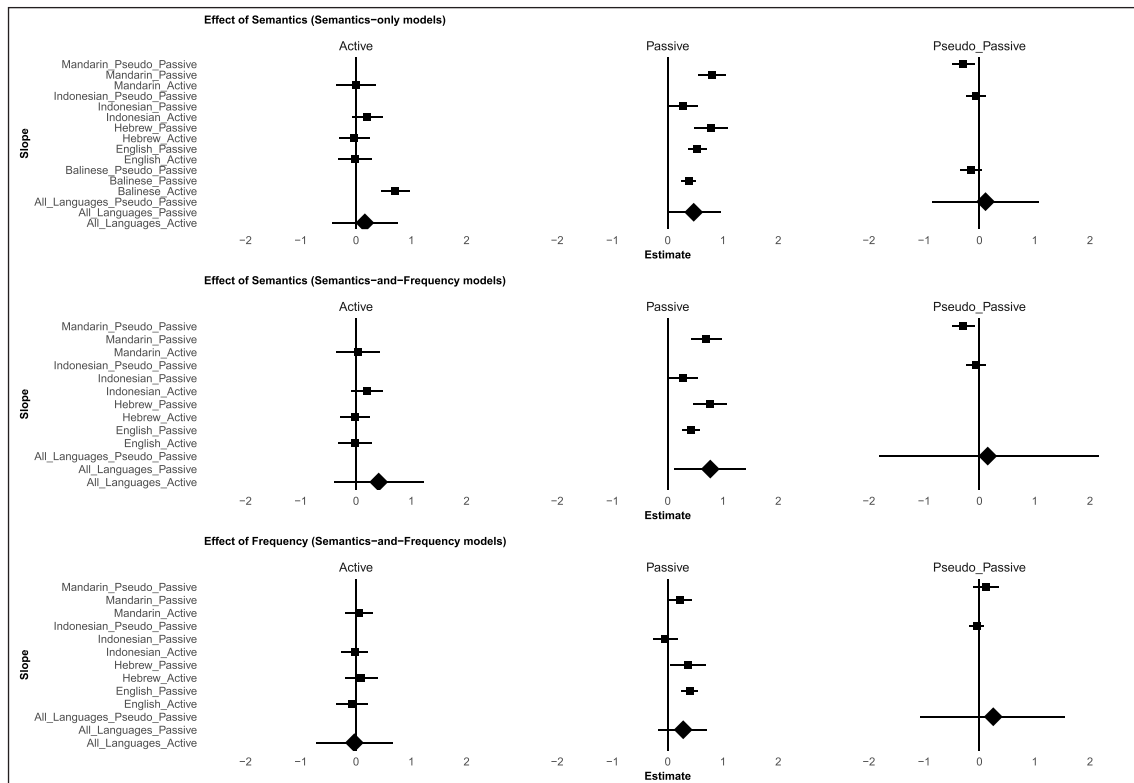


Figure 2: Semantics (and frequency) effects by language and sentence type.

English ($r = 0.26$). Only a marginal degree of collinearity is present for Hebrew ($r = 0.10$), and none for Indonesian ($r = -0.01$), with the caveat that the frequency counts for Indonesian are sparse and likely rather unreliable.

The findings of these analyses are summarized in **Figure 2** (for full models, see **Tables 5–6**). **Figure 2** summarizes the hypothesis tests which – analogous to those performed for Hebrew above – estimate the effect (i.e., slope) of semantics for each sentence type individually. Despite the differences noted above, it is clear that **Figure 2** generally recapitulates the findings outlined in the original papers, with non-zero (“significant”) effects of semantics observed for all bona-fide passives (i.e., not pseudo-passives) in all languages. Indeed, for every language, the confidence interval for the effect of semantics for passive sentences always excluded zero, and the $B < 0$ value was always 1 (i.e., *all* posterior samples were greater than zero), except for Indonesian, which was close ($CI = [-0.01, 0.55]$; $B < 0 = 0.95$; $CI = [0, 0.55]$; $B < 0 = 0.95$; for the analyses excluding and including frequency, respectively).

For pseudo-passives (i.e., Balinese “basic” passives, Indonesian “noncanonical” passives and Mandarin “notional” passives), the slope was always in the opposite direction to that for passives (see **Figure 1**). Indeed, as shown in **Tables 5–6**, the interaction term `S_TypePseudo_Passive:Semantics`

was always negative and non-zero (i.e., “significant”), indicating that the effect of semantics was always smaller for pseudo- than genuine passives.

For actives, the estimate of the semantics slope was – with the exception of Balinese – always (a) very close to zero and (b) “significantly” smaller for actives than passives (see the interaction term *S_TypeActive:Semantics*), though only marginally so for Indonesian. However, Balinese bucked the trend by showing a “significantly” larger effect for actives than passive.

Before moving on to the all-languages meta-analytic synthesis, it is important to clarify the extent to which the findings of the present by-languages reanalyses differ from – or are similar to – those presented in the original papers (though differences in the statistical modeling approaches mean that there is rarely a one-to-one correspondence *per se*).

- For Balinese, Darmasetiyawan and Ambridge (2022) found no evidence that the effect of semantics differed between actives (the reference category) and any type of passive (*-a*, *ka-* and *ma-*; the three were analyzed separately). In contrast, the present reanalysis found that the effect of semantics was (counter to expectations) *bigger* for actives than for (pooled) passives. The present reanalysis also found that the effect of semantics was (as expected) bigger for (pooled) passives than for pseudo-passives. No corresponding analysis was included in Darmasetiyawan and Ambridge (2022) (since actives were always the reference category). Looking at individual sentence types, Darmasetiyawan and Ambridge (2022) found semantic effects for actives and *-a*, *ka-* and *ma-* passives (the three were analyzed separately); but not pseudo passives. These findings were replicated in the present analyses, both when pooling *a*, *ka-* and *ma-* passives, and (in the supplementary analysis available at the OSF project site) analysing each separately.
- For English, Ambridge et al. (2016, Study 3) found a significant interaction, such that the effect of semantics was bigger for passive than active sentences; just as in the present reanalysis. The present reanalysis found a non-zero (“significant”) semantics effect for passive sentences but not for active sentences; no corresponding analysis was included in Ambridge et al. (2016, Study 3).
- For Indonesian, the present reanalysis yields an identical pattern of findings to the analysis reported in Aryawibawa and Ambridge (2018); even the numerical estimates and B values are broadly similar.
- For Mandarin, Liu and Ambridge (2021) found a significant interaction such that, compared to (“normal”; i.e., non-BA-) actives (the reference category), the effect of semantics was bigger for passives, bigger for BA-actives, and smaller for pseudo-passives. The present study found an interaction such that compared to passives (the reference category), the effect of semantics was smaller for actives and pseudo-

passives, and bigger for BA-actives. Thus, although the two analyses are not directly comparable, both found a bigger effect for passives than (“normal”) actives. Looking at individual sentence types, both the original analysis and the present reanalysis found positive effects of semantics for passives and BA-actives, and a negative effect of semantics for pseudo-passives. The present study additionally yielded an estimate of approximately zero for the effect of semantics for (“normal”) active sentences; no corresponding analysis was included in Liu and Ambridge (2021).

In summary, there are no major *discrepancies* between the reanalyses reported here and those in the original studies, though the original analyses do show some important omissions. For this reason, and because the present analyses contain several improvements (most notably, the use of ordinal regression models), they should be considered as more definitive estimates of the relevant effects.

To sum up, then, the individual by-languages analyses generally found (a) evidence for a non-zero effect of passives in the predicted direction, and (b) evidence that this effect was bigger for passives than (always) pseudo-passives and (mostly) actives.

3.2 All-languages (meta-analytic synthesis) model

In order to investigate whether this pattern holds across languages, we built final semantics-only and semantics-and-frequency models including Language as a random intercept (and with maximal random slopes); i.e.,

All-languages semantics-only-model: $\text{Response} \sim \text{S_Type} * \text{Semantics} + (1 + \text{S_Type} * \text{Semantics} | \text{Participant}) + (1 + \text{S_Type} * \text{Semantics} | \text{verb}) + (1 + \text{S_Type} * \text{Semantics} | \text{Language})$

All-languages semantics-and-frequency model: $\text{Response} \sim \text{S_Type} * \text{Semantics} + \text{S_Type} * \text{Frequency} + (1 + \text{S_Type} * \text{Semantics} + \text{S_Type} * \text{Frequency} | \text{Participant}) + (1 + \text{S_Type} | \text{verb}) + (1 + \text{S_Type} * \text{Semantics} + \text{S_Type} * \text{Frequency} | \text{Language})$

In order to allow these models to converge within the maximum time allowed by the server (7 days), it was necessary to reduce adapt delta from 0.99 to 0.95 and the number of iterations from [2000,5000] to [1500,3000] (all Rhat values remained at 1.0, indicating no convergence problems).

In order to allow for comparable analyses across languages and sentence types, we (a) collapsed Balinese *-a*, *ka-* and *ma-* passives as “Passive”, (b) collapsed Balinese “basic” passives, Indonesian “noncanonical” passives and Mandarin “notional” passives as “Pseudo_Passive” and (c) excluded Mandarin BA-Actives. The reason for this latter decision is that (garden variety) active sentences are included as a control sentence type that – at least on a strict standard (see

Section 1) – allows us to rule out the possibility that apparent semantic affectedness effects are spurious, and reflect a *general* dispreference for verbs that show noncanonical linking (e.g., *The woman feared [cf. frightened] the man*), *regardless of sentence type*. We can rule out (or, at least, provide evidence against) this possibility by showing that semantic affectedness effects do not hold for (garden variety) actives (or, at least, are smaller for such actives than for passives). That is, the active functions as a control sentence type, precisely because it does NOT show the semantics of affectedness (or, at least, to a lesser extent than does the passive). Thus, a special type of active that has dedicated semantics of affectedness (i.e., the Mandarin BA-Active) clearly does not function as a control sentence type in this way.

Finally, note also that Balinese was excluded from the semantics-and-frequency model, since frequency counts were unavailable for Balinese. In order to tease apart the effects of (a) adding frequency and (b) excluding Balinese, we also ran a supplementary semantics-only model excluding Balinese. This model (available at the project OSF site) was more similar to the semantics-and-frequency model than to the semantics-only model that included Balinese. Thus, it seems that (as for the individual by-language models) adding frequency as a predictor has very little effect on the estimates of semantics, and that *apparent* differences between the all-languages semantics-only and all-languages semantics-and-frequency models are largely spurious, and reflect the exclusion of Balinese.

3.2.1 Semantics-only model

The all-languages semantics-only model is shown in **Table 7**. Following the same logic outlined above with regard to the equivalent models for Hebrew (**Tables 3–4**) and other languages (**Tables 5–6**), the evidence suggests that semantic effects are numerically bigger for passives ($M = 0.47$, $SE = 0.31$, $B = 0.95$) than for either pseudo-passives ($M = 0.11$, $SE = 0.66$, $B = 0.60$) or for actives ($M = 0.16$, $SE = 0.37$; $B = 0.69$), though the posterior probabilities for the comparison (i.e., for the interaction terms “S_TypePseudo_Passive:Semantics” and “S_TypeActive:Semantics”, respectively) do not provide particularly strong evidence ($B = 0.81$ and 0.89 , respectively). That is, the probability that, collapsing across languages, the effect of semantics is bigger for passives than (a) pseudo-passives and (b) actives is only 81% and 89%, respectively. That is, there is certainly *some* evidence for this claim, but it does not meet the threshold if we “translate” the frequentist cut-off of $p < 0.05$ into a Bayesian posterior of $B > 0.95$. On the other hand, if we consider the sentence types individually, the semantics slope for passives arguably meets this criterion ($B = 0.95$ exactly), with the 95% credible interval including zero by the very narrowest of margins ($[-0.01, 0.96]$). For pseudo-passives and actives, however, the credible interval straddles zero.

3.2.2 Semantics-and-frequency model

Before discussing this model, it is important to reemphasize that it differs from the semantics-only model above not only in the addition of frequency, but in the exclusion of Balinese, for which frequency counts were not available. The exclusion of Balinese might be expected to overestimate the extent to which the effect of semantics is bigger for passives than actives, since Balinese was the only language for which this was not the case. In fact, the results (see **Table 8**) were very similar overall to those of the semantics-only model: again, semantic effects are numerically bigger for passives ($M = 0.77$, $SE = 0.44$) than for either pseudo-passives ($M = 0.15$, $SE = 1.25$) or for actives ($M = 0.41$, $SE = 0.54$), though again the posterior probabilities for the comparison (i.e., for the interaction terms “S_TypePseudo_Passive:Semantics” and “S_TypeActive:Semantics”, respectively) do not provide particularly strong evidence (0.80 and 0.91, respectively). That is, the probability that, collapsing across languages, the effect of semantics is bigger for passives than (a) pseudo-passives and (b) actives is only 80% and 91%, respectively (again, failing to meet a frequentist-inspired threshold of $B > 0.95$). On the other hand, if we consider the sentence types individually, the semantics slope for passives narrowly meets this criterion ($B = 0.97$), with the 95% credible interval excluding zero ([0.12, 1.41]). For pseudo-passives and actives, however, the credible interval straddles zero.

Table 7: Semantics-only models for all languages combined.

Covariate	Estimate	Est.Error	CI_lower	CI_upper	B \diamond 0
Intercept[1]	-2.75	0.6	-3.92	-1.48	1
Intercept[2]	-1.79	0.6	-2.96	-0.52	0.99
Intercept[3]	-0.67	0.6	-1.83	0.6	0.89
Intercept[4]	0.81	0.6	-0.36	2.08	0.93
S_TypeActive	1.69	0.6	0.41	2.87	0.99
S_TypePseudo_Passive	-1.94	1.56	-5.03	1.35	0.91
Semantics	0.47	0.31	-0.12	1.07	0.95
S_TypeActive:Semantics	-0.31	0.28	-0.85	0.23	0.89
S_TypePseudo_Passive:Semantics	-0.36	0.65	-1.69	1.01	0.81

(Contd.)

Covariate	Estimate	Est.Error	CI_lower	CI_upper	B \diamond 0
Semantics_Slope_for_Active	0.16	0.37	-0.42	0.76	0.69
Semantics_Slope_for_Pseudo_Passive	0.11	0.66	-0.85	1.08	0.6
Semantics_Slope_for_Passive	0.47	0.31	-0.01	0.96	0.95

Table 8: Semantics-and-frequency models for all languages combined.

Covariate	Estimate	Est.Error	CI_lower	CI_upper	B \diamond 0
Intercept[1]	-5.02	0.81	-6.53	-3.23	1
Intercept[2]	-3.28	0.81	-4.8	-1.5	1
Intercept[3]	-1.55	0.81	-3.07	0.23	0.96
Intercept[4]	0.56	0.81	-0.96	2.35	0.79
S_TypeActive	1.79	0.8	0.07	3.33	0.98
S_TypePseudo_Passive	-4.13	2.2	-8.29	0.66	0.96
Semantics	0.77	0.44	-0.11	1.62	0.97
Frequency	0.28	0.3	-0.32	0.88	0.89
S_TypeActive:Semantics	-0.36	0.36	-1.05	0.35	0.91
S_TypePseudo_Passive:Semantics	-0.62	1.17	-3.1	2	0.8
S_TypeActive:Frequency	-0.31	0.37	-1.03	0.4	0.87
S_TypePseudo_Passive:Frequency	-0.04	0.84	-1.88	1.81	0.56
Semantics_Slope_for_Active	0.41	0.54	-0.4	1.22	0.84
Semantics_Slope_for_Pseudo_Passive	0.15	1.25	-1.81	2.16	0.57

(Contd.)

Covariate	Estimate	Est.Error	CI_lower	CI_upper	B < 0
Semantics_Slope_for_Passive	0.77	0.44	0.12	1.41	0.97
Frequency_Slope_for_Active	-0.03	0.46	-0.72	0.66	0.47
Frequency_Slope_for_Pseudo_Passive	0.25	0.89	-1.06	1.55	0.43
Frequency_Slope_for_Passive	0.28	0.3	-0.16	0.71	0.89

4. General discussion

Recent experimental studies of (chronologically) English, Indonesian, Mandarin and Balinese have found support for the hypothesis that the passive construction is associated with the semantics “[B] (mapped onto the surface [passive] subject) is in a state or circumstance characterized by [A] (mapped onto the by-object or an understood argument) having acted upon it”. (Pinker et al., 1987). In each of these studies, participants’ ratings of verbs’ semantic affectedness significantly predicted other participants’ acceptability ratings for those verbs in passive sentences, but not – or to a lesser extent – pseudo-passive and/or active sentences. In Study 1, we replicated this finding for a fifth language, Hebrew. In Study 2, we conducted a meta-analytic synthesis which confirmed that – when looking across languages – a semantic effect is found for passives, but not pseudo-passives or actives. However, we found only weak crosslinguistic evidence (posterior probabilities of 80%–91%) that this semantic effect is *larger* for passives than pseudo-passives or actives.

This is particularly surprising, given that – in general – these interactions are seen in the individual by-language models (Tables 5–6). How can these interaction terms (i.e., Sentence_TypeActive:Semantics and Sentence_TypePseudo_Passive:Semantics) be “significant” for each individual language, but not when collapsing across languages? Presumably this somewhat counterintuitive pattern arises because – with just five languages – the critical interactions are underpowered with respect to the random effect of language (see Brysbaert & Stevens, 2018; Judd et al., 2017; Kerkhoff & Nussbeck, 2019, for evidence that power in mixed-effects models is a function of all the random effects in the model). As an analogy, consider the more familiar case where Participant is modelled as a random effect. We would not be surprised if – in an $N = 5$ study – each individual participant showed a significant effect of some factor, yet no significant effect of this factor were observed in the study overall. We understand intuitively that the statistical model needs more than five participants before it is happy to say that there is an effect that generalizes across participants. The same is true here for languages.

On a more positive note, it remains the case that – despite being underpowered with regard to languages – the present synthesis did find a crosslinguistic semantic effect for passive sentences (though not pseudo-passive or active sentences), when analyzed individually. However, it is important to acknowledge that – as noted in Section 1 – observing a semantic affectedness effect when looking at passive sentences only is a relatively lenient criterion for concluding that we have evidence of such an effect, since we cannot rule out the possibility that it is caused – at least in part – by the fact that many low-affectedness verbs also reverse canonical linking (e.g., *The woman feared [cf. frightened] the man*), which generally decreases perceived grammaticality across the board. A more definitive basis on which to conclude that we are seeing a crosslinguistic effect of passive semantics would be the finding that this semantic effect is bigger for passives than pseudo-passives or actives. As we have seen, although the present study generally found this pattern for each language individually, we do not yet have sufficient evidence that it generalizes *across* languages.

These findings therefore raise – though do not yet provide conclusive evidence for – the possibility that a passive construction that denotes undergoer-affectedness may approach the status of a semantic universal. To be clear, we are not suggesting that every language in the world *has* a passive construction; in fact, only around half do so: $162/373 = 43\%$, according to the World Atlas of Language Structures online (Siewierska. 2013). What we are suggesting, however, is that far from reflecting “mere” derivation from active structures or general principles, the passive construction – in languages that have one – might always denote a scenario in which the passive subject is highly affected in some way. Of course, this suggestion goes far beyond what we have been able to demonstrate empirically, since the present analysis includes only five languages. However, the fact that the effect holds so consistently within individual languages from four unrelated families augers well for the *possibility* that it holds true across languages in general. Why might this be the case? Given that these languages are unrelated (apart from Balinese and Indonesian), it cannot be due to crosslinguistic borrowing. Instead, an evolutionary explanation must be in order: Whether or not they have at their fingertips a non-passive topicalization/fronting construction, speakers apparently find it very useful to have a construction that both topicalizes the undergoer and highlights the extent to which it was affected. Interestingly, Rissman et al. (2020) show that agent-backgrounding constructions such as the passive are present in Nicaraguan Sign Language, but not home sign, presumably because they first require the establishment of a canonical means of expressing *agent/patient/theme/experiencer* roles, which the passive(like) constructions can then reverse.

From a theoretical standpoint, the present findings could potentially be explained by a variety of different frameworks. For example, for a constructivist, exemplar approach (e.g., Ambridge, 2020a, 2020b; Croft & Cruse, 2004; Goldberg, 1995; Hilpert, 2014;) the implication is that models of adult linguistic competence should include (for the relevant languages) some notion of a free-standing passive construction with the relevant semantics (rather than merely deriving

the passive from clauses with canonical word order, or using basic principles). Under more traditional linguistic approaches which assume autonomy of syntax (e.g., Chomsky, 1993), the implication would be that the semantic constraints observed here are the purview of discourse pragmatics (Aissen, 1999), event structure (e.g., Paolazzi et al., 2022), lexical semantics (Nguyen & Pearl, 2021) or some other factor outside of the realm of syntax. One possible argument for the first position is the finding of Ambridge et al. (2016, Study 2) that, for English, verbs with very low scores on the same semantic affectedness measure discussed here (from an extended verb set not used in the crosslinguistic studies) receive such low acceptability ratings that they, in effect, do not passivize at all (e.g., **\$5 was cost by this book*; **Four people are slept by this tent*; **He is resembled by Winston Churchill*; **The movie was left by us*). This finding constitutes a potential problem for autonomy-of-syntax approaches which assume that delineating the possible and impossible sentences of a language is very much a job for syntax.

A further theoretical Implication is that, although we have here treated frequency as very much a nuisance variable to be “controlled out” when investigating semantic effects, the observed frequency effects – which were almost as pervasive as semantic effects – are potentially theoretically important in their own right. To recap, when collapsing across languages, the slope for frequency was “significant” (i.e., $pMCMC > 0.95$) for passives, and the same was true for each language individually, except for Balinese (where no frequency counts were available) and Indonesian (where the frequency counts were sparse and unreliable). Recall that the (chi-square-based) frequency predictor reflects the extent to which – relative to the other verbs in the sample – each verb is biased towards (+) or away from (–) the passive construction versus all non-passive constructions. Thus, what the present findings show, in simple terms, is that the more a verb is biased to appear in a passive construction in the input, the more acceptable it is deemed in the passive construction. Again, two alternative broad classes of explanations of this finding are possible. Autonomy-of-syntax accounts could explain these frequency effects as resulting from discourse pragmatics: The passive sentences that our participants find somewhat awkward are the same types of passive sentences that our corpus speakers avoid in production, for whatever discourse-pragmatic reason disfavors the use of passive subjects that are low in affectedness. Construction-based accounts would seek instead to incorporate these effects into the grammar, by assuming that the grammar itself is learned probabilistically from the input. Differentiating these explanations will not be straightforward, particularly since construction-based accounts would not seem to rule out discourse-pragmatic effects on top.

To sum up, while their theoretical implications remain a matter for debate, what the present findings have shown is that across Balinese, English, Hebrew, Indonesian and Mandarin, semantic affectedness effects are always observed for bona-fide passives, but not for non-passive topicalization constructions that share passive word order, and only sporadically for actives. These findings will need to be incorporated by accounts of language evolution, adult representation, and child acquisition alike.

Data accessibility statement

All raw data and analysis code can be downloaded from <https://osf.io/hwb26/>. Hebrew stimuli are available from Gorilla Open Materials at <https://app.gorilla.sc/openmaterials/426411>.

Ethics and consent

Ethics approval was obtained from the Research Ethics Committees of The University of Liverpool (RETH001041) and The Hebrew University of Jerusalem (“Cross-Linguistic Acquisition of Sentence Structure: Integrating Experimental and Computational Approaches”) prior to recruitment. Participants gave informed written consent.

Acknowledgements

Ben Ambridge is Professor in the Division of Psychology, Communication and Human Neuroscience at The University of Manchester. The support of the Economic and Social Research Council [ES/L008955/1] is gratefully acknowledged. This project has received funding from the European Research Council (ERC) under the European Union’s research and innovation programme (grant agreement no. 681296: CLASS: Crosslinguistic Acquisition of Sentence Structure).

Competing interests

The authors have no competing interests to declare.

Authors’ contributions

Ben Ambridge was responsible for Conceptualization, Data curation, Formal analysis, Funding acquisition, Methodology, Project administration, Supervision, Visualization, Writing – original draft, Writing – review & editing. Inbal Arnon was responsible for Methodology, Project administration, Supervision, Writing – review & editing. Dani Bekman was responsible for Investigation, Methodology, Writing – review & editing.

References

- Aissen, J. (1999). Markedness and subject choice in Optimality Theory. *Natural Language & Linguistic Theory*, 17(4), 673–711. DOI: <https://doi.org/10.1023/A:1006335629372>
- Alexiadou, A., Anagnostopoulou, E., & Schäfer, F. (2018). Passive. *Syntactic Structures after 60 years: The impact of the Chomskyan revolution in linguistics* (pp. 403–425). DOI: <https://doi.org/10.1515/9781501506925-407>
- Ambridge, B., Bidgood, A., & Thomas, K. (2021). Disentangling syntactic, semantic and pragmatic impairments in ASD: Elicited production of passives. *Journal of Child Language*, 48(1), 184–201 DOI: <https://doi.org/10.1017/S0305000920000215>

- Ambridge, B., Bidgood, A., Pine, J., Rowland, C., & Freudenthal, D. T. (2016). Is passive syntax semantically constrained? Evidence from adult grammaticality judgment and comprehension studies. *Cognitive Science*, *40*(6), 1435–1459. DOI: <https://doi.org/10.1111/cogs.12277>
- Ambridge, B., Noble, C., & Lieven, E. V. M. (2014). The semantics of the transitive causative construction: Evidence from a forced-choice pointing study with adults and children. *Cognitive Linguistics*, *25*(2), 293–311. DOI: <https://doi.org/10.1515/cog-2014-0012>
- Arywibawa, I. N., & Ambridge, B. (2018). Is syntax semantically constrained? Evidence from a grammaticality judgment study of Indonesian. *Cognitive Science*, *42*(8), 3135–3148. DOI: <https://doi.org/10.1111/cogs.12697>
- Barr, D., Levy, R., Scheepers, C., & Tily, H. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. DOI: <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D. (2016). Linear mixed-effects models in Julia. <https://github.com/dmbates/MixedModels.jl>
- Bates, D., Kelman, T., Simon, A., Noack, A., Hatherly, M., & Bouchet-Valat, M. (2016). *Dmbates/Mixedmodels.Jl: Drop Julia V0.4.X And Earlier Support*. Zenodo. Available at: <https://doi.org/10.5281/ZENODO.162045>.
- Bezanson, J., Karpinski, D., Shah, V., & Edelman, A. (2012). *Julia: A fast dynamic language for technical computing*. <http://julia-lang.org/images/julia-dynamic-2012-tr.pdf>
- Bidgood A., Rowland, C. F., Pine, J. M., & Ambridge, B. (2020). Syntactic representations are both abstract and semantically constrained: Evidence from children’s and adults’ comprehension and production/priming of the English passive. *Cognitive Science*, *44*(9), ee12892. DOI: <https://doi.org/10.1111/cogs.12892>
- Branigan, H. P., & Pickering, M. J. (2017). An experimental approach to linguistic representation. *Behavioral and Brain Sciences*, *40*, e282. DOI: <https://doi.org/10.1017/S0140525X16002028>
- Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1*(1), 9. DOI: <https://doi.org/10.5334/joc.10>
- Bürkner, P. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, *10*(1), 395–411. DOI: <https://doi.org/10.32614/RJ-2018-017>
- Bürkner, P. C., & Vuorre, M. (2019). Ordinal regression models in psychology: A tutorial. *Advances in Methods and Practices in Psychological Science*, *2*(1), 77–101. DOI: <https://doi.org/10.1177/2515245918823199>
- Chomsky, N. (1957). *Syntactic Structures*. Mouton. DOI: <https://doi.org/10.1515/9783112316009>
- Chomsky, N. (1993). A minimalist program for linguistic theory. In K. Hale & S. J. Keyser (Eds.), *The view from building 20* (pp. 1–52). MIT Press.
- Connell, L., & Lynott, D. (2012). Strength of perceptual experience predicts word processing performance better than concreteness or imageability. *Cognition*, *125*(3), 452–465. DOI: <https://doi.org/10.1016/j.cognition.2012.07.010>

- Croft, W., & Cruse, A. (2004). *Cognitive linguistics*. Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511803864>
- Culicover, P. W., & Jackendoff, R. (2005). *Simpler syntax*. Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780199271092.001.0001>
- Darmasetiyawan, I. M. S., & Ambridge, B. (2022). Syntactic representations contain semantic information: Evidence from Balinese passives. *Collabra: Psychology*, 8(1), 33133. DOI: <https://doi.org/10.1525/collabra.33133>
- Darmasetiyawan, I. M. S., Messenger, K., & Ambridge, B. (2022). Is Passive priming really impervious to verb semantics? A high-powered replication of Messenger et al. (2012). *Collabra: Psychology*, 8(1), 31055. DOI: <https://doi.org/10.1525/collabra.31055>
- Ferreira, F. (1994). Choice of passive voice is affected by verb type and animacy. *Journal of Memory and Language*, 33(6), 715–736. DOI: <https://doi.org/10.1006/jmla.1994.1034>
- Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. University of Chicago Press.
- Hilpert, M. (2014). *Construction Grammar and its application to English*. Edinburgh University Press.
- Hopper, P. J., & Thompson, S. A. (1980). Transitivity in grammar and discourse. *Language*, 56, 281–299. DOI: <https://doi.org/10.1353/lan.1980.0017>
- Hopper, P. J., & Thompson, S. A. (1984). The discourse basis for lexical categories in universal grammar. *Language*, 60, 703–752. DOI: <https://doi.org/10.1353/lan.1984.0020>
- Ibbotson, P., Theakston, A. L., Lieven, E. V., & Tomasello, M. (2012). Semantics of the transitive construction: Prototype effects and developmental comparisons. *Cognitive Science*, 36(7), 1268–1288. DOI: <https://doi.org/10.1111/j.1551-6709.2012.01249.x>
- Jones, S. D., Dooley, M., & Ambridge, B. (2021). Passive sentence reversal errors in autism: Replicating Ambridge, Bidgood, and Thomas (2020). *Language Development Research*. DOI: <https://doi.org/10.31234/osf.io/mz9ut>
- Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology*, 68, 601–625. DOI: <https://doi.org/10.1146/annurev-psych-122414-033702>
- Keenan, E. L., & Dryer, M. S. (2007). Passive in the world's languages. In T. Shopen (Ed.), *Language typology and syntactic description: Vol. II, Clause structure* (pp. 325–361). Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511619427.006>
- Kerkhoff, D., & Nussbeck, F. W. (2019). The influence of sample size on parameter estimates in three-level random-effects models. *Frontiers in Psychology*, 10, 1067. DOI: <https://doi.org/10.3389/fpsyg.2019.01067>
- Liu, L., & Ambridge, B. (2021). Balancing information-structure and semantic constraints on construction choice: building a computational model of passive and passive-like constructions in Mandarin Chinese. *Cognitive Linguistics*, 32(3), 349–388. DOI: <https://doi.org/10.1515/cog-2019-0100>

- MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology, 4*, 226. DOI: <https://doi.org/10.3389/fpsyg.2013.00226>
- Marsman, M., & Wagenmakers, E. J. (2017). Three insights from a Bayesian interpretation of the one-sided P value. *Educational and Psychological Measurement, 77*(3), 529–539. DOI: <https://doi.org/10.1177/0013164416669201>
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, R. H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language, 94*, 305–315. DOI: <https://doi.org/10.1016/j.jml.2017.01.001>
- Messenger, K., Branigan, H., McLean, J., & Sorace, A. (2012). Is young children's passive syntax semantically constrained? Evidence from syntactic priming. *Journal of Memory and Language, 66*(4), 568–587. DOI: <https://doi.org/10.1016/j.jml.2012.03.008>
- Næss, A. (2007). *Prototypical transitivity*. Typological Studies in Language, 72. John Benjamins. DOI: <https://doi.org/10.1075/tsl.72>
- Nguyen, E., & Pearl, L. (2021). The link between lexical semantic features and children's comprehension of English verbal be-passives. *Language Acquisition, 28*(4), 433–450. DOI: <https://doi.org/10.1080/10489223.2021.1939354>
- Paolazzi, C. L., Grillo, N., Cera, C., Karageorgou, F., Bullman, E., Chow, W. Y., & Santi, A. (2022). Eyetracking while reading passives: An event structure account of difficulty. *Language, Cognition and Neuroscience, 37*(2), 135–153. DOI: <https://doi.org/10.1080/23273798.2021.1946108>
- Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*. MIT Press.
- Pinker, S., Lebeaux, D., & Frost, L. A. (1987). Productivity and constraints in the acquisition of the passive. *Cognition, 26*(3), 195–267. DOI: [https://doi.org/10.1016/S0010-0277\(87\)80001-X](https://doi.org/10.1016/S0010-0277(87)80001-X)
- Pollard, C., & Sag, I. A. (1994). *Head-driven phrase structure grammar*. University of Chicago Press.
- Pullum, G. (2014). Fear and loathing of the English passive. *Language and Communication, 37*, 60–74. DOI: <https://doi.org/10.1016/j.langcom.2013.08.009>
- R Core Team. (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. www.R-project.org/
- Ravid, D., & Saban, R. (2008). Syntactic and meta-syntactic skills in the school years: A developmental study in Hebrew. In I. Kupferberg & A. Stavans (Eds.), *Language education in Israel: Papers in honor of Elite Olshtain* (pp. 75–110). Magnes Press.
- Ravid, D., & Vered, L. (2017). Hebrew passive morphology in later language development: The interface of verb temporality and register. *Journal of Child Language, 44*, 1309–1336. DOI: <https://doi.org/10.1017/S0305000916000544>
- Revelle, W. (2018). Psych: Procedures for personality and psychological research, Northwestern University. <https://CRAN.R-project.org/package=psych> Version = 1.8.3.
- Rice, C. A., Beekhuizen, B., Dubrovsky, V., Stevenson, S., & Armstrong, B. C. (2019). “A comparison of homonym meaning frequency estimates derived from movie and television subtitles, free association, and explicit ratings.” *Behavior research Methods, 51*, 139–1425. DOI: <https://doi.org/10.3758/s13428-018-1107-7>

- Rissman, L., Horton, L., Flaherty, M., Senghas, A., Coppola, M., Brentari, D., & Goldin-Meadow, S. (2020). The communicative importance of agent-backgrounding: Evidence from homesign and Nicaraguan Sign Language. *Cognition*, 203, 104332. DOI: <https://doi.org/10.1016/j.cognition.2020.104332>
- Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, 110, 104038. DOI: <https://doi.org/10.1016/j.jml.2019.104038>
- Shlonsky, U. (2014) Topicalization and focalization: A preliminary exploration of the Hebrew left periphery. In: A. Cardinaletti, G. Cinque, and Y. Endo (Eds.). *Peripheries*. Hituzi Syobo (pp. 327–341).
- Siewierska, A. (2013). Passive constructions. In: M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online*. Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/107>
- Stefanowitsch, A., & Gries, S. (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, 8(2), 209–243. DOI: <https://doi.org/10.1075/ijcl.8.2.03ste>
- Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and syntactic description, vol. III: Grammatical categories and the lexicon* (pp. 57–149). Cambridge University Press.
- Westfall, J., & Yarkoni, T. (2016). Statistically controlling for confounding constructs is harder than you think. *PLoS ONE*, 11(3), e0152719. DOI: <https://doi.org/10.1371/journal.pone.0152719>

