Why Algebraic systems aren't sufficient for syntax: Experimental evidence from passives in English, Mandarin, Indonesian, Balinese and Hebrew.

Ben Ambridge, University of Manchester, UK

Ben.Ambridge@Manchester.ac.uk

Ben Ambridge is Professor in the International Centre for Language and Communicative Development (LuCiD) 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).

Why Passives?

If ever there was a construction that seems particularly well suited to algebraic approaches to language, it is the passive. Even the hackiest programmer could, in a matter of minutes, bash out some code to automatically generate a sentence like *Bob was kicked by Wendy* from the corresponding active *Wendy kicked Bob* (as in the heyday of transformational approaches), or from the VP *kicked Bob*, adding *by Wendy* as an adjunct (as under minimalist approaches). The algorithm is so simple that when I need to generate passive sentences for my experiments, I do so automatically with an Excel spreadsheet. And provided it's furnished with a list of passive participles (e.g., it's *eaten*, not *ate*), this single-line formula *almost never* fails to generate well-formed English passives. The same is true for passives in – to pick just the languages I happen to have worked on – Indonesian, Mandarin, Balinese and Hebrew.

But appearances can be deceptive. In this chapter I will summarize evidence that, for all of these languages, algebraic rules that operate without reference to the semantics of the construction cannot explain the patterns we see in passive experiments with flesh-and-blood speakers.

Why not? Remember I said that my single-line Excel formula *almost* never fails to generate well-formed passives. Well, those occasional failures hold the key:

This tent sleeps three people → *Three people are slept by this tent.

Homer resembled Grandpa → *Grandpa was resembled by Homer.

The leak dribbled water → *Water was dribbled by the leak.

The house has four windows → *Four windows are had by the house.

Now, before you try and tell me that some of these passives aren't *that* bad, I asked people to rate them in an acceptability judgment study, and they all scored well below the midpoint of the scale (Ambridge, Bidgood, Pine, Rowland & Freudenthal, 2016). What went wrong with my trusty spreadsheet? Well unbeknown to Excel, these passives fall foul of a semantic constraint on the passive that was – building on the work of many others of course – neatly summarized by a pre-pop-sci-era Steven Pinker:

> [B] (mapped onto the surface subject [of a passive]) is in a state or circumstance characterized by [A] (mapped onto the *by*-object or an understood argument) having acted upon it. Pinker, Lebeaux and Frost (1987: 249; see also Pinker, 1989)

The passive examples above are bad, because they don't satisfy what I will call this "affectedness" constraint. If *Bob was kicked by Wendy*, it's fair to say that he is in a state or circumstance (probably quite a painful one) characterized by *Wendy* having acted upon him. But there is no sense in which *four windows* or *three people* are in a state or circumstance caused by *the house* or *this tent* having acted upon them.

"No problem", you might say, "we can just give the algorithm a list of these very rare exception verbs (there are probably less than 50 in the entire English language, at least based on Pinker,

1989) and go home". The problem is, the same *affectedness* constraint that rules out these rather extreme examples predicts participants' performance with – in a binary sense – "well formed" passives, across production, priming, comprehension and judgment tasks. We certainly *can* say, for example, *Bob was seen by Wendy*. But, compared to *Bob was hit by Wendy*, we are less likely to say it (production), to be syntactically primed by it (priming), to rapidly select a matching picture or animation (comprehension) or to give it a high acceptability rating (judgment). Unless, that is, the sentence is presented in a context in which Bob *is* highly affected by being seen by Wendy (e.g., Bob was trying very hard to avoid Wendy, because he owed her a huge amount of money he couldn't pay back. Then…).

Before we take a look at this evidence, let's clear up a couple of potential objections. First, am I straw-manning here: Does anyone actually propose "algebraic" rules of passive formation with no role for semantics? Ladies and gentlemen, I give you Noam Chomsky (1993: 4)

> 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.

OK, so Chomsky doesn't explicitly mention semantics here. But he's going even further: He's ruling out *any* special rules or processes that apply to the passive altogether – whether semantics-based or otherwise. Similarly, other Chomskyan approaches to the passive (e.g., Boeckx, 1998; Collins, 2005; Carnie, 2007), while not specifically *ruling out* semantic factors, propose detailed

rules and procedures for passive formation that make no mention of them.

At least one (non-Chomskyan) approach goes further, explicitly positing that syntax consists of a single "level of representation [that] includes syntactic category information but not semantic information (e.g., thematic roles) or lexical content" (Branigan & Pickering, 2017: 8). Or, to put it even more starkly, "syntactic representations do not contain semantic information" (Branigan & Pickering, 2017: 8). Are they talking about passives here? Definitely. When they say that "Like adults, 3- and 4-year-olds appear to have abstract syntactic representations that are not specified for lexical or thematic content" (p.16), two of the four studies they cite as evidence are studies of the passive (Bencini & Valian, 2008; Messenger et al., 2012).

On to the second possible objection: Why does a sematic constraint rule out an algebraic rule? Why can't we have both: an algebraic passive rule with a semantic constraint sitting on top; perhaps somewhere outside of syntax, such as in the discourse-pragmatics system, or as a phenomenon of performance rather than competence? In all fairness, yes, if this is your preferred option, the data I'm about to summarize don't rule it out (though it's difficult to imagine what possible data *could* rule it out). What I *will* argue (in the rather-pretentiously-titled "Denouement" section) is that we should prefer an account which explains all of the relevant phenomena in a single mechanism without the need for various add-ons (your mileage – as the Americans say – may vary). What we need, I will argue, is an account under which this probabilistic affectedness constraint emerges in speakers' grammars, as a result of hearing thousands of passives with exactly that type of meaning. I will try to sketch such an account at the end of this chapter. First, here's the evidence that motivates it.

Act I: Early studies with children

Our story starts with a pair of studies conducted in the 1980s. Maratsos, Fox, Becker and Chalkley (1985) presented children with passive sentences (e.g., *Superman was held by Batman*) and asked them verbally "Who did it" or – in separate experiments – to choose the correct picture from a pair (e.g., Batman holding Superman; Superman holding Batman). The manipulation of interest was the type of verb: Four verbs – *find, hold, wash* and *shake* – were "physical action verbs" (p.170); or, as they later came to be known in the passives literature, AGENT-PATIENT verbs. Eight verbs – *remember, forget, know, like, miss, see, hear*, and *watch* – were "mental state verbs" (p.170); later called EXPERIENCER-THEME verbs. Although Pinker et al (1987) was still a couple of years away, Maratsos et al (1985) clearly had a similar kind of semantic constraint in mind (which they date to Bever, 1970 and Jackendoff, 1972): The actional (AGENT-PATIENT) type verbs all describe events where the PATIENT is clearly affected (i.e., "in a state or circumstance characterized by…the *by*-object having acted upon it") in some way (e.g., *Superman was found / held / washed / shaken by Batman*). This is not the case for the mental-state (EXPERIENCER-THEME) verbs (e.g., *Superman was remembered, forgotten, known, liked, missed, seen, heard, watched by Batman*).

So, what happened? Exactly as predicted by a semantic affectedness constraint, children were worse with the mental-state passives than the actional passives. Like, *much* worse: In the "Whodunnit?" study, 4-5 year-olds scored 40% on mental-state passives and 67% on actional

passives. And crucially, the deficit for mental-state verbs was specific to passives. Children were also given active sentences as a control, and scored 89% and 91% for mental-state and actional actives respectively. The contrast in the picture-choice study was even more stark: 4-year-olds scored just 34% on mental-state passives, but 85% on actional passives (92% and 97% on actives). Importantly from our perspective, this mental-state-passives deficit was still evident for 7- and even 9-year-olds, at least numerically (the statistical analyses are reported a bit haphazardly, at least by modern standards). Meanwhile in New York, Sudhalter and Braine (1985) were conducting essentially the same "Whodunnit?" study as Maratsos' Minnesota-based team, with essentially the same results.

In both cases – although, of course, I'm speaking with the benefit of hindsight – these studies' authors don't quite join the dots. On the one hand, both groups are clearly aware that certain passives (e.g., **John was fit by the new suit*) are unacceptable in the adult grammar, and that there's "considerable semantic predictability" with regard to which (Maratsos et al, 1985: 168). On the other hand, both can't quite shake off the idea that the semantic constraint they find in their studies is a quirk – a deficit even – of *children's* grammars; one that is long-gone by adulthood. For example, Sudhalter and Braine (1985) suggest that "*by* is understood as signaling Actor before it is understood as signaling Experiencer" (p.469), while Maratsos et al (1985: 189) discuss how the passive "extends its semantic range after the period of initial restriction".

So let me join the dots for them: The *same* – as near as dammit – semantic constraint on the passive is present in the heads of adults and 4-year-olds. The constraint *seems* to go away, because adults and older children are very good at "Whodunnit?" and picture-matching tasks,

and so show ceiling performance *even if the sentences are slightly weird and unnatural* (e.g., *Superman was missed by Batman*). But, as we will see in a minute, if you use a more sensitive task, even adults show a certain disfluency with – and dislike of – these mental-state/EXPERIENCER-THEME type passives, which are inconsistent with the *affectedness* constraint on the passive construction.

Act II: Passive semantics in peril (eventually)

Although they didn't know it yet, Maratsos and his Minnesota colleagues had strong evidence for a semantic constraint on adult passives. But on the other side of Lake Michigan, trouble was brewing for this hypothesis. Bock and Loebell (1990) presented Michigan State University students with Passive "prime" sentences (e.g., *The construction worker was hit by the bulldozer*), along with suitable pictures, and then asked them to describe pictures of their own. They found that participants produced significantly more passives (e.g., *The 747 was alerted by the airport's control tower*) in this condition than in a control condition with active "prime" sentences (e.g., *The construction worker drove the bulldozer*).

OK, but how is this evidence against a semantic constraint on the passive? It isn't – That evidence comes from a third condition, which used locative "prime" sentences (e.g., *The construction worker was digging by the bulldozer*). The key here is that locative sentences are not, by any stretch of the imagination, passive sentences (they're active intransitives [*The construction worker was digging…*] with a prepositional phrase expressing the location [*…by the*

bulldozer]). All that they share with passives is a very abstract grammatical structure. Exactly what grammatical structure that is depends on your theory of grammar, but essentially it's along the lines of: [S[NP *The something*] [AUXP[AUX *was*] [VP[V *Verbing*] [PP [P *by*] [NP *The something*]]]]]. Now, what Bock and Loebell showed was that these locative primes were *just as effective* at priming participants to produce passive sentences as were genuine bona-fide passive primes. This is very bad news for the idea of semantic affectedness constraint on the passive. Why? Well, suppose this constraint takes the form of a semantic prototype construction; something, very roughly speaking, like *[HIGHLY AFFECTED PATIENT] was [ACTIONed] by [HIGHLY AFFECTING AGENT]*). If that were the case, then presumably it wouldn't be possible to prime this structure with something that is semantically *nothing like this at all*. Presumably it wouldn't be possible to prime this structure with a locative, whose first NP slot is filled not by a HIGHLY AFFECTED PATIENT, but with the exact opposite: An AFFECTING AGENT or ACTOR (e.g., *The construction worker was digging…*).

Now, before we move on, I think it's important to point an often-overlooked fact about Bock and Lobell's findings. My argument doesn't hinge on this, so disregard it if you like, but it seems to me that there's a gaping hole right in the middle of Bock and Lobell (1990) that everyone has somehow missed. Textbook and summary descriptions of these findings (like mine above, and like Bock and Loebell's own summary) focus on the *differences* between conditions. But let's take a moment to look at the actual numbers, or at least the percentages: Following passive and locative primes, participants produced 79% and 80% passives (21% and 20% actives) respectively. Following active primes, participants produced 74% passives (26% actives). OK, so 74% is significantly less than both 79% and 80%. But 74% is a *staggeringly* high rate of passive

production. Do you know how many passives crop up in spontaneous speech? Neither do I exactly, but I can tell you from some corpus counts that we did for Ambridge, Bidgood, Pine, Rowland and Freudenthal (2016) that the rate is *well* under 1%. Can you see where I'm going with this? *Participants were consciously aware that the study was about passives, and that they were supposed to produce lots of passives.* So much so, that they produced passives at around *100 times* the usual rate, even when "primed" with actives. Given this context, it is hardly surprising that sentences which share some superficial overlap with the passive – such as the presence of *by* – boost passive production further (and, even then, only by 5 percentage points).

And here is the smoking gun: Ziegler, Bencini, Goldberg and Snedeker (2019) replicated Bock and Loebell's main finding, but found that the locative→passive priming disappeared when the locative prime didn't include the word *by* (e.g., *The construction worker has dug near* [c.f. *was digging by*] *the bulldozer*). This is not to say that passive priming always requires *by*; it doesn't. For example, Messenger, Branigan & McLean (2011) found that, for both adults and children, *by*-less short passives (e.g., *The girls are being shocked*) primed full *by* passives (e.g., *The king was scratched by the tiger*). But that's passive→passive priming with quite a bit of semantic overlap in terms of SUBJECT-affectedness. Purely structural *locative*→passive priming, with syntactic overlap but no semantic overlap (à la Bock & Loebell, 1990), is a myth; driven by explicit awareness that producing passives is the aim of the game, and/or by the presence of *by*.

Let's set aside, for now, my misgivings about Bock and Loebell (1990), because it's important for the dramatic arc of this chapter that Act II ends with the semantic-affectedness hypothesis for adult passives in peril. But if Bock and Loebell won't provide that peril, then who will?

Not Gordon and Chafetz (1990), that's for sure: Their evidence against the semantic affectedness hypothesis is weak, due to a quite bizarre methodological decision. This is a shame, because Gordon and Chafetz's (1990) passives task is actually rather ingenious. Rather than asking "Whodunnit?" (which is pretty odd for EXPERIENCER-THEME passives, where no one really *does* anything) or using picture pointing (which requires children to keep in mind who is the AGENT/EXPERIENCER and who the PATIENT/THEME) they used non-reversible passives with *yes/no* questions. For example, a child might be told that John hated peas and then asked (a) *Were the peas hated by John?* and (b) *Was John hated by the peas?* This is a clever task, since only children who truly understand passives can answer both *Yes* to (a) and *No* to (b). Using this method, Gordon and Chafetz (1990) again found that children did much better with actional (AGENT-PATIENT) than non-actional (EXPERIENCER-THEME) passives (63% vs 35%), but showed no such difference for actives (85% vs 82%).

Next, in order to see whether this pattern was due to differences in semantic affectedness, Gordon and Chafetz asked six adults to rate the relevant 18 verbs for affectedness ("How affected is something that is VERBed?") using a 7-point scale. They then ran a correlational analysis to investigate whether these affectedness ratings predicted children's performance with passives on the *yes/no* question task. So far, so good: This is a method I thoroughly approve of – I've done something very similar myself (Ambridge et al, 2016). But here's the bit that's totally bizarre: Because all the non-actional verbs received uniformly low affectedness ratings (a mean of 1.1/7), they were excluded from the correlation analysis. *Wait, what?* So, Gordon and Chafetz (1990) have demonstrated that there is no relationship between affectedness and passivizability,

as long as you exclude all the verbs that score very low on both affectedness and passivizability. *Wait, what?!*

No, the real evidence against the semantic-affectedness hypothesis comes from the other side of the Atlantic – Scotland, to be specific – in the form of a series of studies by Messenger, Branigan, McLean and Sorace (2012). First, Messenger et al gave children and adults a picture-choice comprehension task, very similar to that used by Maratsos et al (1985). One difference was that as well as the familiar AGENT-PATIENT (actional) verbs (*bite, carry, hit, pat, pull, squash*) and EXPERIENCER-THEME (mental-state) verbs (*hear, ignore, like, love, remember, see*), Messenger et al introduced a third type: THEME-EXPERIENCER verbs (*annoy, frighten, scare, shock, surprise, upset*). Although, in one sense, these verbs are similar to "mental-state" verbs (in that they are "psychological"), in a more important sense – i.e., in terms of semantic affectedness – they are similar to AGENT-PATIENT actional verbs. If anything, they are even more affecting. If *Bob was hit by Wendy* (AGENT-PATIENT), it's possible that he barely even noticed (e.g., if the contact was very light, and he was already being jostled in a crowd, say). But if *Bob was frightened by Wendy* (THEME-EXPERIENCER), he is by-definition affected: If he didn't notice Wendy's Halloween costume, or wasn't bothered by it, we can't even say *Bob was frightened*. Anyway, like the studies before them, Messenger et al found that children (and even adults, though the age groups weren't analysed separately) did much worse with EXPERIENCER-THEME than AGENT-PATIENT (and also THEME-EXPERIENCER) passives (e.g., *A soldier is being seen < pulled/scared by a horse*). But *unlike* all the studies before them, Messenger et al found that children and adults did much worse with EXPERIENCER-THEME than AGENT-PATIENT (and also THEME-EXPERIENCER) *actives*.

(e.g., *A horse is seeing* < *scaring/pulling a solider*). But why?

> Not only is it easier to depict action verbs such as *hit* or *kiss* than to depict experiential verbs such as *love* or *hate* or even perception verbs such as *see*, but it is also easier to distinguish the verb's underlying subject – the causer of the event – for pictures involving verbs like *hit* than for verbs like *see* (and therefore easier to distinguish correctly the target picture from the distractor picture containing the same entities but with the roles swapped). (Messenger et al, 2012: 584)

So, after all that, there was never any difficulty with EXPERIENCER-THEME *passives* after all, it was just EXPERIENCER-THEME *verbs*; or maybe just EXPERIENCER-THEME *pictures*. That can't be right, can it?

It certainly looks that way from the other studies in Messenger et al (2012). Across two priming studies (again with both children and adults, who performed similarly), EXPERIENCER-THEME passives (e.g., *see*) were just as good as AGENT-PATIENT passives (e.g., *pull*) and THEME-EXPERIENCER passives (e.g., *scare*) at priming AGENT-PATIENT passives.

A slight wrinkle is that it's not exactly clear quite what a semantic-affectedness constraint account would predict here. On the one hand, there are quite a few studies (e.g., Bernolet & Hartsuiker, 2010; Jaeger & Snider, 2013) showing that more surprising prime sentences show a bigger priming effect. This means that EXPERIENCER-THEME passives (e.g., *see*), which

don't fit the semantic prototype would show a *bigger* passive priming effect than the other two types (e.g., *pull, scare*). On the other hand, it feels intuitively – to me at least – that more prototypical passives should activate the stored semantic prototype for the passive *more*. This means that EXPERIENCER-THEME passives (e.g., *see*), which don't fit the semantic prototype would show a *smaller* passive priming effect than the other two types (e.g., *pull, scare*). As I said, it's not clear. What is clear is that – whichever direction it goes in – we'd certainly expect a *different* degree of passive priming from more -versus less-semantically-prototypical passive prime sentences. If, that is, the semantic-affectedness account is right. But that's not what Messenger et al found. Which suggests that it isn't.

Now, if you're up to date with all your methods and stats and the like, you should be feeling a bit queasy at the *argument from absence of evidence* here. Messenger et al (2012) – and also Branigan and Pickering (2017) – take the *absence of evidence* for any priming differences between EXPERIENCER-THEME versus AGENT-PATIENT/THEME EXPERIENCER passives as *evidence of absence* of any priming differences between the three types. But, hey, we were all doing that way back in 2012, before the replication crisis had really hit: That power-posing study was in 2010, but the you-know-what didn't really hit the fan until that precognition paper in 2011 (Carney, Cuddy & Yap, 2010; Bem, 2011). In fact, the sample size (24 children and 24 adults) was much too small to rule out the possibility of a genuine, but small and hard-to-detect, difference in priming propensity between the three verb types. Again, this isn't to single out Messenger and colleagues, we were all running underpowered studies back in 2011 (yes, even me! e.g., Ambridge, Pine & Rowland, 2011). It just means that we need a suitably-powered replication.

So that's what we did (that's me, and a PhD student, Sena Darmasetiyawan). And, in order to ensure fair play, and that the replication was as close as possible, we asked Kate Messenger to come on board as a co-author, which she did. It turns out that the means in the original study were in the direction predicted by the semantic affectedness hypothesis ("intuitive" version): i.e., less priming by EXPERIENCER-THEME passives than the other two types. So, we assumed, for the purposes of a simulation-based power analysis, that an effect of this (very small) size was underlyingly there in Messenger's original data, and worked out how many (adult) participants we'd need to have a 95% chance of detecting it. It turned out to be 240. Undaunted, we (Darmasetiyawan, Messenger & Ambridge, 2022) – well, Sena – recruited the participants and ran an online version of the study (which was just as well when the pandemic came along).

Did we find a passive priming difference between the different semantic verb types? It depends. You could certainly *p*-hack (or *B*-hack) your way to one if you were determined to, with the right combination of framework (frequentist/Bayesian), random-effects structure (maximal models or not?) and so on. The (pretty conservative) pre-registered analysis finds a Bayes Factor (with priors based on Messenger's original data) of around 1.5. The presence of an effect is about one and a half times more likely than its absence, but the evidence is "Weak" (Raftery, 1995) or "Anecdotal" (Jeffreys, 1961). So, did we find a semantic effect? Put it this way: If, despite having 240 participants, you are scrabbling around to find an effect that just about meets some criteria for significance, your effect is too small to be worth worrying about; particularly when compared to the absolutely HUGE overall priming effect found in both Messenger et al (2021) and our replication. Dark days, then, for the semantic-affectedness account of passives.

Act III: Passive resistance

Or so it would seem, until our hero – ahem, your humble narrator – arrives to save the day. Well before I got the idea of trying to replicate Messenger et (2012) directly, it seemed to me that this study wasn't quite the final word on the issue, and that things might look rather different with a couple of methodological tweaks.

Thinking about the picture-choice study, it seemed quite right to me that, as noted by Messenger and colleagues, EXPERIENCER-THEME verbs (e.g., *ignore, remember, see, love, hear, like*) are just more difficult to illustrate in pictures. So, I wondered if using animations might make a difference. How exactly do you illustrate – to take one of our sentence stimuli – *Marge was ignored by Homer*? Using still pictures, with great difficulty. But what if Homer approaches Marge, she does a funny dance, and he turns his back and frowns? It's not perfect, but it's pretty clear who is ignoring whom here. And how to address the problem of ceiling effects? Despite the difficulty of illustrating particular actions in still pictures, Messenger et al's 24 adults made just 16 incorrect picture choices between them in the entire study (vs 848 correct choices). This hardly leaves a lot of room for passive-specific by-verb differences. So, what if we asked adults to choose the matching picture *as quickly as possible*, and timed them? And what of trichotomizing a continuous variable – verbs' semantic affectedness – into discrete categories (AGENT-PAITENT / THEME-EXPERIENCER / EXPERIENCER-THEME verbs)? Aren't statisticians always telling us that this is "a bad idea" (Royston, Altman & Sauerbrei, 2006)?

How about we take a leaf out of Gordon and Chafetz's playbook and use instead a continuous measure of semantic affectedness obtained from adult raters. but – bear with me here– retaining all the verbs, rather than throwing away half of them (Ambridge et al, 2016; Study 1)?

When we did all of this (Ambridge et al, 2016, Study 4; see also Meints, 1999), we found a main effect of semantic affectedness; such that the more "affecting" the verb, the quicker participants chose the matching picture, for active and passive sentences alike: Messenger et al were quite right that some verbs are just difficult to process, and/or to illustrate in pictures. Crucially, though, we also found a significant interaction: The reaction time "cost" for verbs scoring low on affectedness was greater for passive than active sentences. Verbs scoring low on semantic affectedness really do display "passive resistance". In Bidgood, Pine, Rowland and Ambridge (2020) we replicated this finding with adults (though using a Messenger-et-al style three way split of verb types, rather than a continuous measure of affectedness) and – using correct/incorrect points rather than reaction time – children.

Wait, what happened to Ambridge et al (2016) Studies 2 and 3? Thanks for asking! These were grammaticality judgment tasks (or, as some insist I call them, "acceptability judgment" tasks), in which adults simply rated the grammatical acceptability of active and passive sentences with a bunch of verbs. In fact, in Study 2, we went all out and included every verb we could think of from Maratsos et al (1985), Sudhalter and Braine (1985), Gordon and Chafetz (1990) and Messenger et al. (2012), as well as Pinker et al. (1987: 250–6) and Levin (1993); for a total of 475 (Study 3 was just a replication with a subset of 72 reversible and – in binary sense – passivizable verbs). Again, we found an interaction between semantic affectedness (as rated by a

separate group of adults) and sentence type: People don't *love* sentences with low-affectedness verbs whether those sentences are active or passive. But they dislike them *a lot more* when they're passive, as you can see in Figure 1 (the colours are just for convenience – the analysis used a continuous rather than class-based measure of verbs' semantic affectedness)

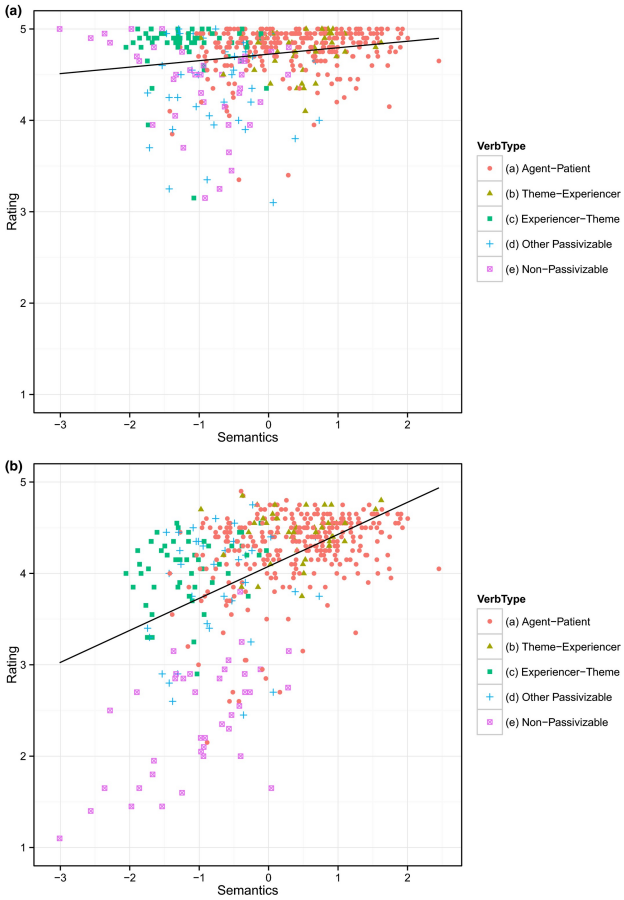

Figure 1: Mean grammaticality judgment score for (a) actives and (b) passives as a function of the semantic affectedness predictor (475 verbs; Study 2). Reproduced from Ambridge et al (2016) under a Creative Commons licencse.

Personally, I thought these findings were pretty convincing. Branigan and Pickering (2017: 50) did not:

> Ambridge's results suggest that acceptability judgments are affected by semantic factors, a point that reinforces our conclusion that acceptability judgments do not straightforwardly reflect syntactic representation.

This seems to me to be a rather circular – even unfalsifiable – argument: "We know that 'syntactic representations do not contain semantic information' (Branigan & Pickering, 2017: 8). So, any study which suggests that they do *must not actually be measuring syntactic representations*". Presumably following this logic, we should also dismiss the findings from my timed animated-picture-choice studies for the same reason?

No worries: If evidence from syntactic priming studies is what Branigan and Pickering demand, then that's what they shall have. In Bidgood et al (2020) Study 2, we ran a modified version of Messenger et al's priming study, varying the verb type (AGENT-PATIENT, THEME-EXPERIENCER, EXPERIENCER-THEME) of the *target* verb rather than, as in Messenger et al (2012), the *prime* verb. Why the switch? Well in Messenger et al's original study, the target verb was always an AGENT-PATIENT verb. And as we already know from Bock and Loebell (1990) and Ziegler et al, 2019) priming passives with AGENT-PATIENT verbs is almost trivially easy: You can do it using things that *aren't even passives at all* – but that just have a passing resemblance to them (e.g., *The construction worker was digging by the bulldozer*). Particularly when participants can tell fairly easily that the study is about passives, and that they are supposed to produce lots of them (74% remember, in the *active* [!] prime condition of Bock and Loebell).

So, it's hardly surprising that – in Messenger et al (2012) – ever-just-so-slightly-awkward EXPERIENCER-THEME passives (e.g., *A soldier is being seen by a horse*) prime passives very effectively; and only – at most – a *shade* less effectively (Darmasetiyawan et al, 2022) than AGENT-PATIENT / THEME-EXPERIENCER passives (e.g., *A soldier is being pulled/scared by a horse*). Flipping around the design of Messenger et al (2012) – varying the type of the target verb, not the prime verb – makes between-verb differences much easier to see, because it requires participants to *actually say* the slightly-awkward EXPERIENCER-THEME passives, which they might be a bit more reluctant to do than the more natural AGENT-PATIENT and THEME-EXPERIENCER passives.

And that's how it turned out: Both adults and children (a) produced fewer passives with EXPERIENCER-THEME verbs than with the other two verb types; and (b) showed a bigger priming effect with EXPERIENCER-THEME verbs than with the other two verb types; exactly as predicted by a semantic-affectedness account of the passive. Game, set and match.

Denouement: A few more studies and some theory

As students of film and theatre will know, the third act doesn't end with the climax, but with the denouement: We go back to normal everyday life, but reflect on how things have changed as a result of the upheaval we've just witnessed: What are the wider implications? What have we learned?

In terms of wider implications, it's worth mentioning just briefly that we have since replicated the production-priming experiment of Bidgood et al (2020, Study 2) with children with and without autism spectrum condition (Ambridge, Bidgood & Thomas, 2020; Jones, Dooley & Ambridge, 2021); and the grammaticality judgment experiment of Ambridge et al (2016, Study 3) in Indonesian (Aryawibawa & Ambridge, 2019), Mandarin (Liu & Ambridge, 2021), Balinese (Darmasetiyawan & Ambridge, submitted) and Hebrew (you'll have to take my word for it on this one: I've got the data but haven't yet got around to writing it up). Interestingly, for Indonesian, Mandarin and Balinese, actives show the semantic affectedness effect to pretty much the same extent as passives; or, in the case of the dedicated affectedness ("BA-") construction in Mandarin, even more so. But this doesn't mean that affectedness just predicts something nonspecific like general sentence acceptability. In each case, a semantic affectedness effect is observed for "genuine" passives, but not topicalization constructions, which mimic passive word order, but lack passive morphology. This is not to say that passive affectedness is a semantic universal (e.g., Maratsos et al, 1985, suggest that it may not apply to Gaelic); but it certainly seems to be something that many languages find useful.

This raises the question of just *why* the passive construction in English and many other languages displays a semantic affectedness constraint in the first place. The answer seems to lie with discourse structure. Gordon and Chafetz (1990: 247) – surprisingly, given their overall scepticism regarding a passive affectedness constraint itself – put it best:

> The reason that adults fail to use passives for non-action verbs [i.e., EXPERIENCER-THEME verbs- BA] is that there is usually nothing happening to the logical object of the non-action verb, and hence there is no reason to focus on that argument by passivizing the verb. The reason that action verbs [i.e., AGENT-PATIENT and THEME-EXPERIENCER verbs – BA] are good candidates for passivization…is that PATIENTS of action verbs [including EXPERIENCERs of THEME-EXPERIENCER verbs – BA] are normally affected by the action and hence are things to focus on in discourse.

In order to dig into this idea a bit, I ran just one more study (sorry, I promise I'll stop after this one). If Gordon and Chafetz are right, then it should be possible to increase the rated acceptability of a slightly awkward EXPERIENCER-THEME passive by placing it in a context in which the THEME is quite highly affected. For example:

THEME-affected context: Jack was trying very hard to avoid Emily, because he owed her a huge amount of money he couldn't pay back. Then Jack was seen by Emily.

Neutral context: Jack was looking for his friend in the park. Then Jack was seen by Emily.

My intuition is that the target sentence *Jack was seen by Emily* is slightly more acceptable in the first context than the second, because only here is he really affected by having been seen (e.g., he'll probably feel embarrassed, worried etc.). We've only just started running this study, but the pilot data are broadly supportive. You can see them, and try out the study for yourself, here:

https://osf.io/4cdga/.

If the final judgment data match my intuitions, then the implications are really quite radical. These findings, if confirmed, would suggest that grammaticality doesn't depend on the fit between the semantics of the construction (here, affectedness) and the semantics of the verb in some fixed dictionary-definition sense. Rather, grammaticality depends on the fit between the semantics of the construction (here, affectedness) and *the semantics of the whole scene* – i.e., the extent to which the first entity is affected. Even more radically, on this view, grammaticality is no longer a property of particular sentences (remember, the rated sentence is identical in both contexts), but of *pairs* of sentences and real-world meanings+contexts.

Now, let's return to the topic of this volume, and to a potential objection I raised back in the Introduction: Why does a sematic constraint rule out an algebraic rule? Why can't we have both: an algebraic passive rule with a semantic constraint sitting on top; perhaps somewhere outside of syntax, such as in the discourse-pragmatics system, or as a phenomenon of performance rather than competence? Well, as I acknowledged in the Introduction, if this is your preferred option, you're welcome to it; the data I've summarized don't rule it out. But it's *clunky* – look at all the components we need to make it work.

To our basic algebraic rule, we first need to add a mechanism that blocks the rule entirely for non-passivizable verbs like *sleep, resemble, dribble* and *have* (remember the examples from earlier: **Three people are slept by this tent; *Grandpa was resembled by Homer; *Water was*

*dribbled by the leak. *Four windows are had by the house*). Presumably, and standardly, this would be done via the verb's lexical entry (or valance frame etc.). If so, this solution is already starting to look a little clunky, as it treats passivizability – in this binary sense – as an arbitrary lexical property of the verb, rather than as semantically motivated. If, on the other hand, we don't rule out these passives using the verb's lexical entry – postponing everything until a later non-syntactic stage – we would be treating ungrammatical passives as somehow different to ungrammatical transitives, ungrammatical intransitives and all the other ungrammatical utterances that are assumed to result from using a verb in a sentence structure that is not licensed by its lexical entry.

Next, we need to add – in the discourse-pragmatics system, or in performance, a probabilistic constraint that explains the fine-grained semantic differences observed between – broadly-speaking – "grammatical" passives in the studies summarized here. We'll also need an explanation of how this constraint interacts with the syntax system to yield the production, judgment and comprehension findings summarized here. But where did this pragmatic constraint come from? Since it varies from language to language, it can only have been learned probabilistically from the input. But if children can learn this subtle fine-grained pragmatic constraint on the passive probabilistically from the input, why can't they learn the (relatively-straightforward) *form* of the passive probabilistically from the input too? And if they can, then why do we need an algebraic rule?

Finally, assuming the preliminary findings of my context study hold up, we need to add another mechanism which explains how and why *the same sentence* becomes more or less grammatical

in different real-world contexts; that is, how real-world knowledge interacts with the underlying system of pure syntax. And, again, we have to assume that children are skilled and voracious probabilistic input-based learners when it comes to learning which real-world scenarios improve or worsen passives sentences, but hopeless input-based learners when it comes to the apparently-much-simpler task of learning the form of the passive construction itself.

You can, of course build such a system if you want to. But what I don't understand is why anyone *would* want to. Why is it *better* – either for the learner, or for a linguistic analysis – to split the phenomenon of semantic effects on the passive across syntax, the lexicon, discourse-pragmatics and real-world context than to bundle it all into a single probabilistic input-based learning mechanism?

My answer, of course, is that it isn't. So how would this latter possibility work? Let me try to sketch it out. I fully acknowledge, of course, that what follows *is* just a sketch; the proposal would need to be implemented formally – for example as a computational model – before it can be properly tested (as of course is also the case for rival proposals). So, what kind of model of language representation and acquisition could explain all the findings I've set out above (including the very preliminary context findings)? On the one hand, it would need to represent its inputs at a quite *staggering* level of detail: Not just what words were said (e.g., *Jack was seen by Emily*), but the context in which they were said (e.g., that Jack owed Emily a large sum of money and was trying to avoid her). And it would need to store all of this information for every single utterance that it hears. I know this sounds ridiculous, but what's the alternative? How could an affectedness constraint on the passive emerge – on *any* model – if speakers were NOT storing

real-world event information that includes (amongst many other linguistically-relevant notions) affectedness?

So, on the one hand, the model needs to represent passives at a ludicrously highly-specified level. But on the other hand, the model also needs to represent passives at a ludicrously underspecified level: Remember, utterances that are not passives at all but just have a vague resemblance to a passive (e.g., *The construction worker was digging by the bulldozer*) are sufficient to cause passive priming. They do so, most theorists would agree, by activating some kind of stored abstract representation of the passive construction.

The solution, as I argued in a response to commentators called *Abstractions made of exemplars or 'You're all right, and I've changed my mind'* (Ambridge, 2020), is to assume that (p.641)

> …we store all the exemplars that we hear (subject to attention, decay, interference, etc.), but that – in the service of language use – these exemplars are re-represented in such a way as to constitute abstractions … A useful metaphor for this account is a multiple-level connectionist neural network that stores every exemplar, re-representing it in increasingly abstract ways as we move up the hidden layers…[But] this is not just a metaphor… The brain *really does* contain multiple layers of units (i.e. neurons), each of which aggregates input signals using a nonlinear function and outputs signals to other units.

In a sense, this is nothing more than just a traditional constructivist account (e.g., Goldberg,

1995) which posits that speakers store constructions at every level of generality, from fully lexicalized whole utterances (e.g., *Jack was seen by Emily*) to fully-abstract sentence-level schemas (e.g., [THING] BE/GET [VERB] by [THING]), and everything in between (e.g., partly-lexically-specified schemas like *He got VERBed*).

But there is a small difference; at least I think there is. At least on my reading, constructivist accounts assume that the most abstract, top-level constructions are traditional, "human-readable" linguistic constructions like "The English passive" (e.g., [THING] BE/GET [VERB] by [THING]; or even [SUBJECT] BE/GET [VERB] [PP by [NP]]. On my version "it is naïve to expect an explanation couched in terms of 'human-readable' concepts like [NOUN] or [DATIVE CONSTRUCTION] to be anything more than a broad-brush sketch that should not be taken literally" (p.642):

> To see why, let's use an analogy from a domain that is much closer than language to being 'solved': image classification (McClelland, 2020). Show a multi-level neural network model a picture, and it will tell you whether it's a cat, a dog or a house. How does it work? Well, if you insist on an explanation in terms of 'human- readable' concepts like 'nose', 'tail' and 'window' you can have one. But you know full well that this explanation is just a dumbed-down approximation generated to give humans some vague sense of how the system works. How does it actually work? The point is, nobody really knows; at least, not if you define 'knows' as 'able to give an explanation in terms of human-readable concepts'. The bottom-level, least-abstract layer represents the pixels of the image. As we move up through the layers, the representations become increasingly

more abstract. If we plot the activation patterns of these more abstract layers and squint a bit, maybe we can just about make out something that looks sort of like a 'nose detector'. But we know full well that it isn't really one, and that any explanation couched in such simplistic terms is doomed to failure. Sorry, my fellow (psycho-/developmental-)linguists, but language is exactly the same.

And that's where I'll leave it: Far from being a simple algebraic rule, the passive is nothing more than a rough-and-ready human-readable approximation of a certain ever-changing pattern of hidden-unit weightings in a system that maps from real world meaning+contexts to speech sounds, and vice versa. Fade to black. Roll credits:

References

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 https://doi.org/10.1017/S0305000920000215.

Ambridge, B. (2020). Abstractions made of exemplars or 'You're all right and I've changed my mind' Response to commentators. *First Language, 40*(5-6), 640-659. https://doi.org/10.1177/0142723720949723.

Ambridge, B., Bidgood, A., Pine, J.M., Rowland, C.F. & Freudenthal, D. (2016). Is passive syntax semantically constrained? Evidence from adult grammaticality judgment and comprehension studies. *Cognitive Science, 40*(6), 1435-1459.

Ambridge, B., Pine, J.M. & Rowland, C.F., (2011). Children use verb semantics to retreat from overgeneralization errors: A novel verb grammaticality judgment study. *Cognitive Linguistics,* 22(2), 303–323.

Aryawibawa, I.N & Ambridge, B. (2019). Is syntax semantically constrained? Evidence from a grammaticality judgment study of Indonesian. *Cognitive Science, 42*(8), 3135-3148.

Bem, D. J. (2011). Feeling the future: experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, *100*(3), 407-425.

Bencini, G. M. L., & Valian, V. V. (2008). Abstract sentence representations in 3-year-olds: Evidence from language production and comprehension. *Journal of Memory and Language, 59*, 97 - 113.

Bernolet, S., & Hartsuiker, R. J. (2010). Does verb bias modulate syntactic priming?. *Cognition, 114(3), 455-461.*

Bever, T.G. (1970) The cognitive basis for linguistic structures. In J.R. Hayes (ed.), Cognitive and the development of language. New York, Wiley.

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. https://doi.org/10.1111/cogs.12892.

Bock, K., & Loebell, H. (1990). Framing sentences. *Cognition*, *35*(1), 1-39.

Boeckx, C. (1998). *A Minimalist view on the passive*. Harvard: MIT Press.

Branigan, H. P., & Pickering, M. J. (2017). An experimental approach to linguistic representation. *Behavioral and Brain Sciences*, *40*, 1-61.

Carney, D. R., Cuddy, A. J., & Yap, A. J. (2010). Power posing: Brief nonverbal displays affect neuroendocrine levels and risk tolerance. *Psychological Science*, *21*(10), 1363-1368.

Carnie, Andrew (2007) *Syntax: A Generative Introduction* (3rd edition) Oxford: Wiley Blackwell.

Chomsky, N. (1993). A minimalist program for linguistic theory. In *The View from Building 20*, by K. Hale & S.J. Keyser (Eds), 1–52. Cambridge, MA: MIT Press.

Collins, C. (2005). A Smuggling approach to the passive in English. *Syntax, 8*(2), 81-120.

Darmasetiyawan, S. & Ambridge, B. (submitted). Syntactic representations contain semantic information: Evidence from Balinese passives. *Collabra: Psychology.*

Darmasetiyawan, S. Messenger, K. & Ambridge, B. (submitted). Is passive priming really impervious to verb semantics? A high-powered replication of Messenger et al (2012). *Collabra: Psychology.*

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Gordon, P., & Chafetz, J. (1990). Verb-based versus class-based accounts of actionality effects in children's comprehension of passives. *Cognition, 36*(3), 227-254.

Jackendoff. R. (1972) Semantic Interpretation in Generative Crammer. Cambridge, MA: MIT Press.

Jaeger, T. F., & Snider, N. (2013). Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition, 127*, 57–83.

Jeffreys, H. (1961). *Theory of probability* (3rd Ed.). Oxford: Oxford University Press.

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,* doi.org/10.1515/cog-2019-0100.

Maratsos, M. P., Fox, D. E., Becker, J. A., & Chalkley, M. A. (1985). Semantic restrictions on children's passives. *Cognition, 19*(2), 167-191.

McClelland, J. (2020). Exemplar models are useful and deep neural networks overcome their limitations: A commentary on Ambridge (2020). *First Language, 40*(5-6): 612–615.

Meints, K. (1999) Protoypes and the acquisition of passives. In B. Kokinov (ed.) *Perspectives on Cognitive Science, 4*, 67-77. Sofia: NBU Press

Messenger, K., Branigan, H. P., & McLean, J. F. (2011). Evidence for (shared) abstract structure underlying children's short and full passives. *Cognition*, *121*(2), 268-274.

Messenger, K., Branigan, H.P., McLean, J.F. & Sorace, A. (2012). Is young children's passive syntax semantically constrained? Evidence from syntactic priming. *Journal of Memory and Language, 66*(4), 568–587.

Pinker, S. (1989). *Learnability and cognition : the acquisition of argument structure*. Cambridge, MA; London: MIT.

Pinker, S., Lebeaux, D. S., & Frost, L. A. (1987). Productivity and Constraints in the Acquisition of the Passive. *Cognition, 26*(3), 195-267.

Raftery, A. E. (1995). Bayesian model selection in social re- search. In P. V. Marsden (Ed.), *Sociological methodology 1995* (pp. 111–196). Cambridge, MA: Blackwell.

Royston, P., Altman, D. G., & Sauerbrei, W. (2006). Dichotomizing continuous predictors in multiple regression: a bad idea. *Statistics in Medicine, 25*(1), 127-141.

Sudhalter, V., & Braine, M. D. S. (1985). How does comprehension of passives develop? A

Comparison of actional and experiential Verbs. *Journal of Child Language, 12*(2), 455-470.

Ziegler, J., Bencini, G., Goldberg, A., & Snedeker, J. (2019). How abstract is syntax? Evidence from structural priming. *Cognition*, *193*, 104045.