1. Background

In this paper we consider the do so construction, exemplified in (1):

(1) As an imperial statute the British North America Act could be amended only by the British Parliament, which did so on several occasions. [= amended an imperial statute] (Groliers Encyclopedia)

Previous accounts of this construction have noted its seemingly idiosyncratic syntactic and anaphoric properties (Lakoff and Ross 1966, Anderson 1968, Bouton 1970, Halliday and Hasan 1976, Hankamer and Sag 1976, Sag and Hankamer 1984, Miller 1990, Ward, Sproat, and McKoon 1991, Cornish 1992, Fu and Roeper 1993, Dechaine 1994, Kehler and Ward 1995, Kehler and Ward 1999, Fu, Roeper, and Borer 2001, Ward and Kehler 2005, inter alia). Our focus in this paper is on the anaphoric properties of do so, which have confounded previous attempts to answer even the most basic questions regarding its interpretation, such as the level of representation at which its meaning is resolved. On the one hand, for instance, do so requires linguistic evocation, per Hankamer and Sag’s (1976) treatment of it as a SURFACE ANAPHOR, and thus it resists pragmatic control:

(2) [Andy is holding a newborn baby with one hand behind his head, and shows Gregory]
   a. Andy: By doing this, you add support to his developing neck muscles.
   b. Andy: # By doing solso doing, you add support to his developing neck muscles.

On the other hand, we have previously argued (Kehler and Ward 1995, Kehler and Ward 1999, Ward and Kehler 2005) that do so does not satisfy the other criterion of surface anaphora, in that it imposes no requirement for a syntactically-matching antecedent (note the syntactic mismatch in (1), for instance). In that work we argued for an anaphoric theory that captured the requirement for linguistic evocation (see these papers for further details).
In contrast to our analysis, Fu, Roeper, and Borer (2001, henceforth FRB) assume that do so categorically requires a syntactic VP antecedent. Further, in light of felicitous examples of do so with nominalized antecedents (see Section 2.3), they use this assumption to argue that process nominals must contain a VP in their underlying syntactic representations.

In the next section, we review some of the data that support an anaphoric theory of do so, comparing it to Fu et al.’s proposal and their arguments against the conclusions drawn in Kehler and Ward (1995). In Section 3 we argue that do so patterns with other anaphors in being sensitive to the degree of semantic transparency between the antecedent expression and the referent. Finally, in Section 4 we report on a corpus study of cases in which role nominalizations serve as antecedents for do so that supports this conclusion.

2. Anaphora versus Syntax

2.1 Syntactically-mismatched antecedents

A variety of data has previously been offered to show that do so does not require a syntactically-matched antecedent (see Ward and Kehler (2005) and references therein). For instance, examples (3–4) involve voice mismatches: Whereas the clause containing do so is in the active voice, the antecedent clause is in the passive, and hence a matching antecedent VP as required by syntactic treatments does not exist:

(3) As an imperial statute the British North America Act could be amended only by the British Parliament, which did so on several occasions [= amended an imperial statute]. (=1)

(4) Section 1 provides the examples to be derived by Gapping, and a formulation of Gapping capable of doing so. [= deriving the examples] (text of Neijt 1981)

Although problematic for a syntactic analysis, these data are predicted on an anaphoric one.

Fu et al. briefly discuss example (3), and state:

it is not clear that it seriously jeopardizes the claim that do so requires a VP/V’ antecedent. That in the first conjunct the direct object is occupied by a trace, rather than a full NP may very well turn out to be immaterial for the licensing of the anaphor do so. (pp. 572–573)

However, they do not pursue this idea, and thus do not provide the details necessary to evaluate it. Indeed, they never state their assumptions about how do so is
interpreted nor why it would require a syntactic antecedent. On many syntactic theories of VP-ellipsis, the need for syntactic parallelism stems from the fact that a VP needs to be reconstructed at the ellipsis site. This logic does not extend to do so, since it is not associated with an ellipsis site.

Indeed, such details would be required to address a range of other types of attested syntactic mismatch discussed by Kehler and Ward (1995) but not addressed by Fu et al.:

(5) There was a lot more negativity to dwell on, if anyone wished to do so. [= dwell on more negativity]

(6) With or without the celebration, Belcourt is well worth seeing, and you can do so year round. [= see Belcourt]

(7) ... He went on to claim that the allegedly high-spending Labour authorities had, by so doing, damaged industry and lost jobs. [= spent highly] (Cornish 1992)

However, FRB do contrast cases like (3) with cases involving adjectivals, which are presumably worse because they do not involve a trace:

(8) ?? This act turned out to be amendable, and the British Parliament did so in its last session. [= amend the act]

However, a considerably more acceptable variant can be readily constructed, seemingly ruling out a purely syntactic explanation:

(9) After the British Parliament found out that the act was amendable, they elected to do so at their very first opportunity. [= amend the act]

All of these data are thus compatible with an anaphoric theory, but need to be explained on a surface-anaphoric approach.

2.2 Split antecedents

Another type of example discussed by Kehler and Ward (1995) but not addressed by FRB involves split antecedents. It is well-known that pronominal reference is compatible with referents that have antecedents that are ‘split’ across the discourse:

(10) The first person to die each year is usually listed in the newspaper, as is the first couple to file for divorce. In a rare show of respect, this year their names were kept private.
As pointed out by Dalrymple et al. (1991), *do so* is felicitous with split antecedents as well, as illustrated by example (11):

(11) Fortunately, the first person to die in 1990 and the first couple to file for divorce in 1990 were allowed to *do so* anonymously. [= die / file for divorce] (text of Roeper (1990), cited by James McCawley’s “1990 Linguistic Flea Circus”)

Again, this is exactly what we would expect on an anaphoric theory, but is problematic for a syntactic theory since no suitable VP antecedent exists in the discourse.

### 2.3 Process nominalizations

One area of agreement for the two analyses is the potential felicity of *do so* with nominalized antecedents:

(12) The defection of the seven moderates, who knew they were incurring the wrath of many colleagues in *doing so*, signaled that it may be harder to sell the GOP message on the crime bill than it was on the stimulus package. [= defecting] (Washington Post)

(13) Even though an Israeli response is justified, I don’t think it was in their best interests to *do so* right now. [= respond] (token provided by Dan Hardt)

The two analyses draw very different conclusions, however. For our anaphoric analysis, these simply constitute further support. For FRB, on the other hand, these show that process nominalizations must have a VP in their syntactic representations, in light of their assumed requirement that *do so* have a syntactic VP antecedent.

For their conclusion, FRB cite the contrast between (14–15) as evidence:

(14) His removal of the garbage in the morning and Sam’s *doing so* in the afternoon were surprising. (= their 42b)

(15) * Kim’s accident in the morning and Sue’s *doing so* in the evening were not coincidences. (= their 43b)

That is, whereas the nominalization *removal* can antecede *do so* in (14), the non-nominalized event-denoting noun *accident* cannot in (15). The question then is why event-denoting nouns cannot antecede *do so* on an anaphoric theory.

Whereas FRB are correct in stating that this contrast ultimately needs to be explained, it does not provide evidence for their particular explanation. The problem is that the judgments do not significantly change when *do so* is replaced by an anaphor such as *do it*:
(16) His removal of the garbage in the morning and Sam’s *doing it* in the afternoon were surprising.

(17) # Kim’s accident in the morning and Sue’s *doing it* in the evening were not coincidences.

Since *do it* is widely agreed to place no formal restriction on its antecedent (i.e., it is uncontroversially a form of *deep anaphora* per Hankamer and Sag (1976)), the contrast between (14–15) cannot be attributed to a syntactic requirement on *do so*, and in turn does not provide any evidence that nominalizations have a VP within their syntactic representations.

### 2.4 Non-process nominalizations

Finally, FRB’s analysis as described in the last section applies only to process nominalizations; they make no claims to the effect that other types of nominalizations similarly incorporate a VP in their syntax. However, felicitous, naturally-occurring examples of *do so* with other types of nominalizations as antecedents, e.g. role nominalizations, are readily found (the following examples were collected from various internet webpages):

(18) One study suggests that almost half of young female smokers *do so* in order to lose weight. [= smoke]

(19) The majority of horse riders *do so* purely for leisure and pleasure. [= ride horses]

(20) AmericaNet.Com, its officers, directors or employees are not responsible for the content or integrity of any ad. Sellers/buyers/subscribers/investors *do so* at their own risk. [= sell.buy.subscribe.invest]

(21) Data from the Retirement Survey reveals that 5% of early retirees *do so* because of the ill health of others. [= retire early]

An obvious way to maintain FRB’s analysis would be to claim that role nominalizations also incorporate VPs in their syntax. The problem is that this move would overgenerate, allowing many cases that remain unacceptable:

(22) # Most professors will *do so* for hours even when no one is listening. [= profess]

(23) # In my opinion, our governor *does so* better than the last one did. [= govern]
We know of no independent evidence that some role nominals incorporate VP syntax and others do not, nor do we anticipate that such evidence exists.2 What is instead apparent from the contrast between (18-21) and (22-23) is that role nominalizations display gradience with respect to compositionality: Whereas a smoker is someone who smokes and horse rider is someone who rides horses, a professor is not merely one who professes, nor is a governor merely one who governs. We argue that it is the degree of semantic transparency between the nominalization and the verb that it nominalizes that determines the extent to which the nominalization makes the event denoted by the verb accessible for subsequent reference.

3. Anaphoric Islands

Indeed, the distinction between (18–21) and (22–23) mirrors an analogous distinction at the nominal level with respect to outbound anaphora, a type of reference that Postal (1969) characterizes as involving so-called ‘anaphoric islands’. Standard examples that demonstrate the infelicity of such anaphora involve semantically non-transparent relationships (examples from Ward et al. (1991)):

(24) Fritz is a cowboy. # He says they can be difficult to look after. [= cow]

(25) Dom’s clothes are absolutely elephantine. # Indeed, you could almost lose one in them. [= an elephant]

On Postal’s analysis, reference fails in these cases because it involves ‘word internal’ reference: they cannot have cow in cowboys as its antecedent in (24), and similarly for elephant and elephantine in (25).

However, Ward et al. offer an alternative interpretation:

...we shall argue that the degree to which outbound anaphora is felicitous is determined by the relative accessibility of the discourse entities evoked by word-internal lexical elements, and not by any principle of syntax or morphology. (p. 449)

This characterization predicts that felicitous examples that are similar to (24–25) should occur, as long as a sufficient degree of semantic transparency holds between the antecedent expression and its word-internal counterpart. This is in fact the case (examples again from Ward et al. (1991)):

(26) Do parental reactions affect their children? [=parents]

(27) I think if I were a Peruvian I wouldn’t want to live there for the next couple of years. [= Peru]
(28) It’s awfully foggy tonight so you people out there driving better watch out for it. [= fog]

(29) Very well. But I warn you that if you continue in such foolishness you’ll be the last paleontologist alive by the time you retire. There’s no future in it. [= paleontology]

Such examples should be as bad as (24–25) on a purely syntactic account.

Therefore, the facts regarding pronominal anaphora appear to mirror those for do so anaphora that we discussed at the end of the last section. Ironically, however, Ward et al. drew the opposite conclusion about do so, citing example (30) in support of their idea that it is a surface anaphor:

(30) Mary is a heavy smoker – even though her doctor told her not to

   a. # do so. (judgment theirs)
   b. do it.

However, we have already seen a felicitous case of do so with smoker as its antecedent in (18). Example (31) is another:

(31) In some cases removing triggers from your home is as simple as asking smokers to do so outside, removing pets from the house, and vacuuming rugs and washing ...

We therefore conclude that the key difference between (18–21) and (22–23) lies in precisely the sort of semantic transparency and activation factors that Ward et al. cited in arguing against the existence of grammatical anaphoric island constraints.

4. Corpus Study

This conclusion predicts that felicitous uses of do so should occur only with those role nominalizations that have a highly transparent semantic relationship with the verb they nominalize. To test this prediction, we conducted a simple corpus study. We first collected the -er/-or agent nominalizations that occur at least 2000 times in the British National Corpus. This resulted in a list of 42 nominalizations. For each nominalization (N), a search of the form N do so (e.g., drivers do so) was performed using Google, and the results carefully filtered and analyzed. Cases in which there was any ambiguity about the identity of the antecedent were set aside. For 29 of the 42 nominalizations, felicitous examples of do so in which the nominalization unambiguously serves as antecedent were found.

An approximate measure of the degree of semantic transparency between each role nominalization and the verb it nominalizes was then computed. For this purpose we used the Latent Semantic Analysis (LSA) vector cosine value$^3$, as such
values have been claimed to correlate well with human semantic similarity judgments (Landauer, Foltz, and Laham 1998). The idea is that the more transparent the relationship between a role nominalization and the verb it nominalizes, the more likely they will be found in similar contexts within a large corpus. For instance, *smoke* and *smoker* receive a 0.82 score, which is indicative of a high degree of relatedness (1 represents perfect contextual overlap). On the other hand, *profess* and *professor* receive a 0.06 score, which indicates a degree of relatedness close to chance (0 represents chance).

The average LSA value for the 29 nominalizations for which our searches revealed felicitous naturally-occurring examples of *do so* was 0.491. In contrast, the average LSA value for the remaining 13 cases not found in the corpus search was only 0.264. By a one-tailed t-test, these two means are significantly different (p=.004).

Whereas this informal study cannot be considered definitive – LSA values are no doubt a fairly crude approximation of semantic transparency and not all possible felicitous cases will be necessarily found on the web – the results nonetheless strongly point to the idea that degree of semantic transparency between a role nominalization and the verb it nominalizes is a significant factor in determining when felicitous reference with *do so* is possible. Indeed, a post-hoc examination suggests that the effect might be even stronger than the results summarized above suggest. It turns out that no examples were found for two of the highest-scoring nominalizations of the 42: *farmer* at 0.81, and *developer* at 0.77. The reasons for this appear to be idiosyncratic given that, unlike many of the other unattested cases, felicitous examples can be readily constructed. Example (32) is a constructed case with *marijuana farmers* as the antecedent of *do so*.

(32) Many marijuana farmers *do so* with a sincere sense that they are doing nothing wrong, and not out of greed without regard for the law.

Furthermore, example (33) is an attested case with *developers of free software* as the antecedent:

(33) Whatever approach you use, it helps to have determination and adopt an ethical perspective, as we do in the Free Software Movement. To treat the public ethically, the software should be free – as in freedom – for the whole public.

Many developers of free software profess narrowly practical reasons for *doing so*: they advocate allowing others to share and change software as an expedient for making software powerful and reliable. (http://www.gnu.org/philosophy/university.html)

This example was not found during our corpus search because the searches we used – of the form *Ns do so* – only find examples in which *do so* is adjacent to
the nominalization. Example (33) shows nonetheless that developers can antecede do so. Setting these aside, the other 11 verb-nominalization pairs had an average 0.168 transparency score.

5. Conclusions

Our investigation of do so has revealed that it is governed by the same pragmatic principles that apply to other forms of anaphoric reference. In concert with other well-known factors (e.g., topicality, recency of mention), semantic transparency influences accessibility: The more transparent the semantic relationship between a nominalization and the verb it nominalizes, the more accessible the event evoked by the nominalization will be.

We therefore find that do so does not directly impose purely syntactic restrictions on its antecedent, and hence it provides no evidence that process nominals incorporate a syntactic VP. On the other hand, there is an indirect connection between syntactic form and felicity of do so, in the sense that choice of syntactic form influences discourse accessibility, and discourse accessibility in turn influences the felicity of reference with do so.

Acknowledgments

We thank Ivano Caponigro, Laura Kertz, and Hannah Rohde for comments and helpful discussions.

Notes

1 Our corpus contains felicitous naturally-occurring cases of do so with role nominalizations as antecedents for over sixty different verbs.
2 FRB offer a second type of evidence that process nominalization incorporate VP syntax, based on a purported co-occurrence with adverbs:

(34) A presentation of the awards separately (was attended by parents).
(35) His removal of the evidence deliberately (resulted in obscuring the case).

Here we simply disagree with the judgments: We find such data to be ungrammatical on the relevant syntactic modification relationships. We are thus not persuaded by this evidence. Note, however, that even if we were to defer to their judgments on such data, a move to extend FRB’s analysis to role nominalizations would generate the prediction that adverbs should be able to modify those as well. In that case, the subject of (36) should be a grammatically-acceptable way of referring to the person who presented the awards separately, and (33) should be a grammatically-acceptable way of referring to the people who developed the software quickly:

(36) * The presenter of the awards separately (was pleasing to the group of parents).
(37) * The developers of the software quickly (are the ones that I would recommend).
We find such examples to be strongly ungrammatical on these interpretations.

3 Computed at lsa.colorado.edu using Matrix Comparison over the General Reading up to 1st year college corpus.

4 LSA failed to produce a value for one of the otherwise qualifying nominalizations, vendor, when paired with its corresponding verb vend. This case was therefore not included in the aforementioned list of 42.

Appendix

The following list includes an example of *do so* anaphora for each of the qualifying role nominalizations in the BNC for which felicitous cases were found. They were all collected from various webpages in June, 2007. All were carefully analyzed to ensure that the role nominalization was the intended antecedent. For many of the nominalizations, multiple examples were found.

(38) Gulden ignored a race official and jumped the tape marking the finish area to shake hands with his runners. He was the only coach in the shutes, a coach whose instincts have always told him this moment is important. After 25 years, he is remarkably more tenacious than ever. Other coaches show up at meets in jacket and tie, assigning their assistants to points on the course. Gulden shows up in sweats and puts in a few thousand meters himself, running from point to point. The greatest teachers do so by example.

(39) Most evening and night shift workers do so not because they choose to, but because they have to. [= work evening and night shifts]

(40) Successful leaders can be made as well as born. They must be themselves so that others trust them and devote their energy to their work. They are grounded, comfortable with their own strengths and weaknesses and can be appropriately open. The finest leaders do so by example, with integrity, and by knowing every person well in order to encourage other leaders.

(41) Injuries incurred during lunchtime volleyball are NOT covered by W.C.B. Players do so at their own risk.

(42) This site is provided AS-IS and users of this site do so at their own risk.

(43) We have a 100 Club at the school with a guest speaker once a month or so. We have had Joost van der Westhuizen, Bruce Grobbelaar, Dick Muir etc. All speakers do so for free gratis and money raised goes to the school. Usually a large pissup follows. Bruce Grobbelaar left at 2 o’clock.

(44) Avid readers do so as a hobby because they find pleasure in the activity!

(45) DUI is a problem that goes way beyond accidents or simple bad choices, although many who get arrested for DWI fall into one or both of the above categories. Many drunk drivers do so not by choice or stupidity, but because they have a drinking problem.

(46) ...harm to computer systems. Many of these Virus writers do so simply because they can, or because they are bored. Un-wiling, or just...
In France, 22% of consumers think their tap water is too hard and another 22% fear health or sanitary hazards or contamination by toxic substances: this rate is rather low, but it has doubled between 1989 and 2000 (IFEN, 2000). Concern for sanitary water situation is much higher in the United States, where nearly half of bottled water consumers do so out of health and safety reasons.

Every time you raise a site’s profile and rankings, you are pushing others down. SEO is a zero sum game - winners do so at the expense of losers.

Why is the traditional MTS approach often considered less desirable?
Most traditional Make-to-Stock manufacturers do so according to some kind of forecast. In our experience, there are only two kinds of forecasts – wrong and really wrong.

The latest public opinion poll suggest that about 64% of supporters of capital punishment do not believe it is a deterrent. Most supporters do so because they believe in revenge or retribution (not the same thing).

I heard an interesting statistic recently. 80% of Corvette buyers do so for looks, not performance.

The issue of what constitutes small quantities of primary products is particularly important to this sector as many artisan poultry and egg producers do so on a very small scale as a supplement to farm income.

Every time you raise a site take a lot of pride in their genealogy. There are many Luo whose works have roots in Sudan. The researchers do so with great pride. Mr Paul Mboya and Tom Ojienda, both my friends, have completed an exhausting genealogy covering hundreds of years.

How does a scholar develop a publishing strategy?
Although some people including even some very successful publishers do so without a strategy, for many people it is helpful to have an overall plan for publications.

Bollocks. I have found that most Linux lovers do so because they actually enjoy using software that works (I have had less than a dozen major Linux crashes, etc. - they were all either massive H/W failure or my fault).

I am a big fan of the double count method. I have been told that it isn’t done much here in the U.S. Most active observers do so singly. Our group did some simultaneous observing and encoded our sighting by generating a sound tone on audio tape.

The survey also discovered that 50% of “uber affluent” travellers do so with their families in tow.

Organized real estate has evolved to meet that goal. 95% of homesellers come to that conclusion. Most private sellers do so to save the commission and net more money but in most cases a buyer will not let the homeowner save the commission.

A recent survey conducted for Microsoft reveals that 85% of new home computer purchasers do so for access to the World Wide Web.
(60) slimepuppy’s mostly right though there are short-film creators that work exclusively in that market and don’t want to move to features. Many short makers do so with their own money (and own equipment which means less money) and many also do it with grants from government and cultural entities.

(61) Lanny, you are 100% correct. There are 300 plus million people in this country counting children. Usually less than half of the “Could-be” voters do so.

(62) I was just trying to point out that in almost all cases afk miners do so naked with minimal skills beyond the focus of the script. and manual miners are often more balanced chars.

(63) It is important that all visitors to our sites understand that they do so at their own risk.

(64) I would ask the writers of Movie and TV show articles if being a fan of something is a valid reason to want to improve it. I am sure 95% of article editors do so because they are a fan or have an interest in the topic they wrote about.

(65) The majority of mutual fund investors do so through employer-sponsored plans, such as 401(k)s. So, what’s the allure of mutual funds and what has made them so popular?

(66) The industries represented by Noosa Home-based businesses (HBBs) tend to be more diverse than in other locations (e.g. the Sunshine Coast or in NSW/ACT). Noosa HBB operators do so primarily because they do not need a shopfront, want flexible hours, lower overheads and for lifestyle reasons.

References


Miller, Philip. 1990. “Pseudogapping and *do so* substitution”, *Papers from the 26th Regional Meeting of the Chicago Linguistic Society*, 205–239.


Andrew Kehler
University of California San Diego
Department of Linguistics
9500 Gilman Drive, #0108
La Jolla, CA, 92093-0108
kehler@ling.ucsd.edu

Gregory Ward
Northwestern University
Department of Linguistics
2016 Sheridan Road
Evanston, IL 60208-4090
gw@northwestern.edu