Structural Alignment in Dialogue and Monologue (And What Attention May Have to Do with It)

Iva Ivanova\textsuperscript{a*}, William S. Horton\textsuperscript{b}, Benjamin Swets\textsuperscript{c}, Daniel Kleinman\textsuperscript{d}, and Victor S. Ferreira\textsuperscript{e}

\textsuperscript{a}Department of Psychology, University of Texas at El Paso
\textsuperscript{b}Department of Psychology, Northwestern University
\textsuperscript{c}Department of Psychology, Grand Valley State University
\textsuperscript{d}Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign
\textsuperscript{e}Department of Psychology, University of California, San Diego

*Corresponding author

Mailing address:
500 W. University Ave.
Department of Psychology
University of Texas at El Paso
El Paso, TX 79968, USA
Email: imivanova@utep.edu
Abstract

In the Interactive Alignment Theory, alignment is promoted by dialogic features on a dialogue-monologue continuum. More alignment in prototypical dialogue (a chat among friends) than in prototypical monologue (a lecture) seems plausible, but the role of other dialogic features for alignment is less clear. The current study tests the joint influence of two such features: communicative intent and a live interlocutor. Four structural priming experiments contrasted the magnitude of structural alignment between a dialogue situation that involved a joint task and an otherwise identical monologue situation that lacked a communicative component. None of the four experiments showed any statistical differences in alignment between dialogue and monologue. Post-hoc analyses on the pooled data of Experiments 2-4 further suggested that individual differences in attention (operationalized through reaction time variability in the picture verification component of the task we used to elicit alignment) might modulate overall structural alignment, and that dialogue (presumably through sustaining attention) might be able to sustain alignment over time to a greater extent than monologue. These results suggest a role for attention in supporting structural alignment, but the lack of clear differences between alignment magnitudes in monologue and dialogue (absent in individual experiments and weak even in the pooled analyses) suggests that communicative intent and a physically-present interlocutor have at best a minor role for promoting alignment.

Keywords: communication; joint task; attention; structural priming; dative alternation
Structural Alignment in Dialogue and Monologue (And What Attention May Have to Do with It)

Speaking seems effortless for most, yet is extremely complex. Even the simplest utterance involves compiling a message to communicate, computing a structure for it that is grammatical, selecting the appropriate words and retrieving the sounds that make them up. In addition, speakers in a conversation need to begin planning their contribution at the same time as they are comprehending the speech of their conversational partner, dealing with distractions from the environment and timing their breath intake to their turn to speak. But speaking is also remarkably efficient: Speakers produce around two to three words per second but only one in a thousand might be a mistake (Levelt, 1989). Among the factors contributing to the efficiency of language production might be alignment – the tendency of speakers to reuse aspects of the language they encounter (Pickering & Garrod, 2004). In this study, we investigate whether communicative situations enhance such efficiency by comparing the magnitude of structural alignment in dialogue versus monologue.

Interactive alignment during dialogue has been widely attested at different levels of linguistic representation. For example, conversational partners tend to reuse each other’s words (lexical alignment), helping them to converge on a shared conceptualization of a situation (Garrod & Anderson, 1987). Conversational partners can also reuse the syntactic structure of each other’s utterances (structural alignment: e.g., Branigan, Pickering, & Cleland, 2000). Interactive alignment in patterns of use of function words (prepositions, pronouns and auxiliary verbs) has been observed in oral arguments of Supreme Court Justices and lawyers (Danescu-Niculescu-Mizil, Lee, Pang, & Kleinberg, 2012). Speakers in dialogue also align in acoustic-
phonetic features of speech (Pardo, 2006), vocal fry (Borrie & Delfino, 2017), speech rate and accent (Giles & Powesland, 1975; Webb, 1972) and pauses (Cappella & Planalp, 1981).

Interactive alignment seems to bring multiple benefits in conversational interactions. Apart from being associated with improved mutual understanding (e.g., Ferreira, Kleinman, Kraljic, & Siu, 2012) and greater efficiency and better performance in problem solving through verbal communication (Nenkova, Gravano, & Hirschberg, 2008, for lexical alignment; Borrie, Lubold, & Pon-Barry, 2015, for acoustic-prosodic alignment), alignment seems to benefit conversations through its sensitivity to social factors and interpersonal dynamics (see Beňuš, 2014, for a review). That is, changes to one’s speech in dialogue may be undertaken to help ensure the smoothness of an interaction (Communication Accommodation Theory, Giles & Powesland, 1975). For example, people with a higher need for social approval aligned more to their conversational partner’s vocal intensity and pause length than people with a lower need for social approval (Giles & Coupland, 1991; for related evidence, see Pittam, 1994). Speakers might also diverge from their conversational partners to indicate disaffiliation (Bourhis, Giles, Leyens, & Tajfel, 1979; Doise, Sinclair, & Bourhis, 1976). (We note, however, that interactive alignment does not necessarily or always confer communicative benefits: Complementarity of behavior, in essence misalignment, is more beneficial than alignment in tasks such as joint visual detection or visual search; Coco, Dale, & Keller, 2017; Fusaroli, Bahrami, Olsen, Roepstorff, Rees, Frith, & Tylén, 2012.)

Notably, though, there is ample evidence that speakers also align to the language they hear or read in the absence of an interactional partner (e.g., Bock, 1986; Borrie & Liss, 2014; Pickering & Branigan, 1998; Potter & Lombardi, 1998; Weatherholtz, Campbell-Kibler, & Jaeger, 2014; Wynn, Borrie, & Sellers, 2018). Such effects suggest that the alignment
phenomenon is not limited to dialogic situations, and raise the question of whether alignment has similar or different functions in dialogue and monologue. (We distinguish here between “interactive alignment”, alignment in a conversation, and simply “alignment”, reusing aspects of language without a conversational interaction, different from a use of the term “alignment” as limited to a dialogue situation, for example as in Pickering & Garrod, 2004.)

According to the Interactive Alignment theory (Pickering & Garrod, 2004), alignment has a fundamental function in dialogue that is absent in monologue. The claim, as we understand it, is that prototypical dialogue (“face-to-face spontaneous dyadic conversation between equals with short contributions”, p. 187) is characterized by the spontaneous automatic coupling of representations across the two interlocutors without a mediating “decision box” (p. 177). This coupling facilitates language use because it reduces both the need to construct linguistic representations from scratch, and the need to engage in cognitively costly partner modeling. The coupling is enabled or promoted by the characteristic features of prototypical dialogue, such as backchannel responses, interruptions, repairs, joint construction of utterances and gaze coordination. On the other hand, prototypical monologue with little direct listener engagement (such as giving a lecture) lacks such alignment-promoting features. While it can give rise to alignment, such alignment is “epiphenomenal” (p. 177) because it does not lead to a cycle by which two interlocutors become more similar to each other. According to Pickering and Garrod (2004), prototypical dialogue and monologue are not categorically different but lie on two ends of a continuum defined by the degree of coupling between communicators (p. 187), which varies with the presence or absence of the tightly-coupled dialogue features listed above, as well as with parameters such as number of participants (dyadic versus multi-party conversations) and communication channel (video versus face-to-face conversations).
It seems straightforward, then, that there will be less alignment in the monologue than in the dialogue endpoint of this continuum due to the absence of features that directly promote alignment, such as turn-taking. But it is less clear whether other prototypical dialogue features, such as communicative intention and the interlocutor’s physical presence, enhance alignment relative to a comparable situation that lacks these features. Although it has seemed to be an implicit assumption in the field that this is so, to our knowledge only one other study to date has addressed this question, and when our study was conceptualized this direct comparison was missing. In the present work, we focus on structural alignment, and compare the magnitude of alignment in closely-matched situations of communicative versus non-communicative comprehension-production alternation. We refer to these as “dialogue” and “monologue,” respectively, for ease of exposition and to connect with prior literature.

**Structural alignment with and without interaction**

There have been numerous demonstrations of structural alignment per se: Speakers tend to reuse the structure of utterances they have been previously exposed to, by themselves (e.g., Bock, 1986) or others (e.g., Branigan et al., 2000). For example, in an interactive structural alignment study, Branigan et al. (2000) showed that participants were more likely to describe a picture with a prepositional dative structure (e.g., *The cowboy is offering the banana to the burglar*) after hearing a confederate’s description with the same structure (e.g., *The nun is showing a book to the chef*) than after hearing a confederate’s description with a double object structure (e.g., *The nun is showing the chef a book*).

Structural alignment has been attested in a variety of constructions and languages in both interactive and non-interactive settings. Interactive structural alignment has been demonstrated in English for the dative alternation (Branigan et al., 2000; see also Hartsuiker, Pickering, &
Veltkamp, 2004, for structural alignment to datives between Spanish and English), the use or omission of the complementizer *that* (e.g., *Put the penguin (that’s) in the cup onto the star*; Haywood, Pickering, & Branigan, 2005), passives (by children with autism: Allen, Haywood, Rajendran, & Branigan, 2011) and word order and code-switching patterns in Dutch-English code-switched sentences (e.g., *Een grappig plaatje, waarop het meijsje the horse kicks* [a funny picture, on which the girl..:] Kootstra, van Hell, & Dijkstra, 2010). Studies of non-interactional structural alignment (structural priming) have also demonstrated effects from datives (Cai, Pickering, Yan, Branigan, 2012, Mandarin and Cantonese; Salamoura & Williams, 2007, Greek and English; Shin & Christianson, 2009, Korean), noun-phrase structure (Cleland & Pickering, 2003, English), presence or absence of “*that*” (Ferreira, 2003, English), thematic-role order (Griffin & Weinstein-Tull, 2003, English), attachment preferences (Scheepers, 2003, German), and order of auxiliary and past participle (Hartsuiker & Westenberg, 2000, Dutch).

In sum, structural alignment has often been attested in both dialogue and monologue, but are there reasons why greater alignment might be observed specifically in contexts involving a genuine communicative intention and a live interlocutor than in the absence of these factors? We can imagine at least two such reasons. First, the goal of most dialogue is mutual understanding. In the Interactive Alignment account, mutual understanding is achieved through interlocutors’ convergence on similar situation models (Zwaan & Radvansky, 1998), which is a direct function of convergence on linguistic representations. That is, linguistic (including structural) alignment may be critical in helping interlocutors achieve their overarching goal in dialogue, but this goal is missing in non-communicative settings, predicting less alignment.

Second, theory and evidence converge in detailing how structural alignment is frequently sensitive to factors pertaining to the social situation (of which, we argue, a physically-present
interlocutor is part). Balcetis and Dale (2005) found that participants structurally aligned to a greater extent (in three of the four tested structures) with a confederate they perceived as nice than with a confederate they perceived as mean (but aligned more with a confederate who acted annoyed than with a confederate who acted patiently). Heyselaar, Hagoort, and Segaert (2015) found that participants aligned their choice of syntactic structure to a similar extent when interacting with human confederates and human-like avatars, but aligned less with non-human-like avatars (avatars that did not have any facial expressions, did not look at the participant, and had a computerized voice). In the non-interactive study of Weatherholtz et al. (2014), the magnitude of structural alignment was modulated by the perceived distance of a speaker’s accent to participants’ own, perceived similarity of the speaker (as manifested in presumed political ideology), and participants’ preference for compromise as a conflict management style.\footnote{Note that the results of Weatherholtz et al. (2014) converge with those of Balcetis and Dale (2005) conceptually but diverge empirically: Weatherholtz et al. found alignment for double object sentences but not prepositional dative sentences, whereas Balcetis and Dale found alignment for prepositional dative sentences but antialignment for double object sentences.} These findings are consistent with the notion from Interactive Alignment Theory that alignment can be “post-conscious” (Garrod & Pickering, 2004), or sensitive to social and partner-specific factors, similar to non-verbal behavioral mimicry (Lakin & Chartrand, 2003). They also seem to predict that socially engaging situations would elicit more structural alignment than non-social situations.

But is there evidence for more structural alignment in dialogue than in monologue? Indirect evidence can be obtained by carrying out numerical comparisons of structural alignment magnitudes in published studies.\footnote{In studies involving the dative alternation, structural alignment is often calculated as the proportion of prepositional dative or double object picture descriptions after exposure to a prime with the same structure minus the proportion of such descriptions after exposure to a prime with...} Performing such a comparison on studies by Ivanova and...
colleagues suggests an equal magnitude of structural alignment in monologue and dialogue. Ivanova, Pickering, Branigan, McLean, and Costa (2012) and Ivanova, Branigan, McLean, Costa, and Pickering (2017) studied structural alignment from anomalous sentences in a “fake dialogue” situation, which differed from real dialogue in that it did not involve genuine interaction (even if participants believed the contrary) but did involve alternating between reading text and describing pictures. That is, participants were told that they would interact via an automatic speech transcriber with another participant in another room, but in fact the “other participant’s” responses were researcher-created sentences; the vast majority of participants believed this cover story. Across 5 experiments, the magnitude of structural alignment in this paradigm with different verbs between primes and targets (10-14%), was comparable to that of a study otherwise similar in design but involving a standard monologue situation (e.g., 10% in Experiment 2 of Ivanova, Pickering, McLean, Costa, & Branigan, 2012). While this indirect comparison suggests that structural alignment might not differ between monologue and dialogue, the decisive factor could be the physical presence versus absence of an interlocutor and not a dialogue situation per se (cf. the lexical alignment study of Bergmann, Branigan, & Kopp, 2015).

On the other hand, an indirect comparison of studies by Branigan, Pickering and colleagues suggests greater structural alignment in dialogue than in monologue. A dialogue task with a real, physically-present confederate produced a numerically larger effect (26% with different verbs between prime and target; Branigan et al., 2000) than both the studies of Ivanova

the alternative structure (where the proportions of prepositional dative and double object descriptions are complementary, i.e., sum to 100%, so only one of them is reported). As the only exception from the studies below, the proportions of prepositional dative and double object descriptions in Pickering and Branigan (1998) were not fully complementary because they were calculated over all produced responses, including those that did not involve the standard variants of the dative alternation. In the rest of the studies above (Branigan et al., 2000; Ivanova et al., 2012a; 2012b; 2017), proportions were calculated after excluding such “other” responses.
et al. (10-14%) and other studies involving a written sentence completion task (e.g., 4% for double object descriptions and 5% for prepositional dative descriptions; Pickering & Branigan, 1998). However, task differences across studies render such comparisons suggestive at best.

The only study to date directly comparing alignment magnitudes in dialogue and monologue was conducted by Schoot, Hagoort, and Segaert (2019). In this study, one group of participants completed an interactive structural alignment task in which they heard active or passive prime sentences from a confederate, and another group completed a monologue version of this task in which they heard the recorded confederate producing the prime sentences. Participants produced more passive descriptions after passive primes relative to baseline (intransitive) primes overall (while there was no priming for actives), but this effect was significantly larger in the dialogue than in the monologue condition (a difference of about 4%). The authors attributed these differences to the presence of communicative intent, or, more broadly, social goals. However, the dialogue and monologue conditions in this study also differed in the presence or absence of an explicit cooperation incentive. In the dialogue condition, participant and confederate were instructed to detect mismatches between their partner’s description and a concurrently presented verification picture, and a joint score was kept; thus, to be successful, both participant and confederate had to pay careful attention to what their partner was saying. In the monologue condition, the “mismatch” score was based on the participant’s responses alone. Schoot et al.’s (2019) results thus suggest that having the goal to cooperate influences the magnitude of alignment. But it remains unclear whether alignment would differ between dialogue and monologue without a difference between conditions in terms

---

3 For half of the participants in each group, the partner or recording mimicked participants’ non-standard structure choices 90% of the time (adaptive) or 10% of the time (non-adaptive), but this manipulation did not produce any effects and is not relevant here.
of an explicit collaborative task (such as during many conversations which involve only the implicit goal to reach mutual understanding).

**Mechanisms of structural alignment**

There are two main mechanistic accounts of the automatic priming mechanism that drives structural alignment. According to the Implicit Learning model (Chang, Dell, & Bock, 2006; Chang, Janciauskas, & Fitz, 2012), every time comprehenders experience a message expressed with a given structure (e.g., hear it from their conversational partners), the mappings between the message and the structure get strengthened, making them more likely to express similar messages with this structure in the future. According to the Residual Activation model (Pickering & Branigan, 1998), abstract representations of lexical items are linked to nodes representing the structures they can occur with. When a structure is experienced, the respective structural node retains activation. This residual activation makes it more likely that this same structure enters the production process when one subsequently takes the turn to speak.

Neither of these accounts, however, incorporates a mechanism that could account for a differential degree of coupling of representations in dialogue versus monologue. One candidate mechanism could be attention. Indeed, there is evidence that structural alignment is influenced by attention to linguistic input. In a study exploring how the magnitude of interactive structural alignment differs with participant role in dialogue, Branigan, Pickering, McLean, and Cleland (2007) found that speakers who heard prime sentences as addressees showed more structural alignment than speakers who heard prime sentences as side participants. Branigan et al. interpreted their findings in terms of depth of processing: Because only addressees, but not side participants, are expected to actively collaborate with speakers in dialogue to achieve mutual understanding, addressees might pay more attention in an attempt to achieve a deeper
understanding of the speakers’ contributions (although side participants seemingly attend enough to achieve comprehension).

In view of this evidence, one reason for greater structural alignment in dialogue relative to monologue (as defined here and in Schoot et al., 2019) could be enhanced attention to a physically-present partner’s contributions in a communicative situation. On this account, enhanced sustained attention in a particular situation throughout a period of time would translate into enhanced attention to individual utterances, leading to deeper processing of each utterance, including its structure. This in turn would result in greater changes in weights for connections between particular syntactic constituents in the Implicit Learning model, or higher activation of structural nodes in the Residual Activation Model. More attention in dialogue could be externally driven by a live interlocutor, for example, because of humans’ tendency to attend to motion onset (even if the motion is not task-relevant; Abrams & Christ, 2003). A physically present person would naturally initiate perceptible movements such as co-speech gestures, other bodily movements, or speech articulation, while text on a computer or phone screen would not. More attention in dialogue could also be driven internally by high-level goals. In Schoot et al.’s (2019) study, participants in the dialogue group (who aligned more) were explicitly instructed to pay attention to what their partner was saying (to reach the maximum joint score of detected mismatches between prime sentences and verification pictures), while participants in the monologue condition were not. We note that, to account for any differences between dialogue and monologue, it is necessary to assume that global sustained attention in a situation leads to more attention to linguistic input. However, differences in focused attention to particular aspects of utterances may also bring about differential structural alignment effects (see General Discussion).
Another mechanism that could account for differential structural alignment between dialogue and monologue is top-down executive control (alternatively or in addition to attention). Such a mechanism would regulate alignment by further strengthening the mappings between message and structure and reinforcing connection weights between syntactic constituents in the Implicit Learning model (Chang et al., 2006), or by boosting the activation of structural nodes in the Residual Activation Model (Pickering & Branigan, 1998). Similar systems, involving a regulatory mechanism acting on domain-specific representations, have been proposed for action control (Norman & Shallice, 1986) and language control in bilinguals (Green, 1998; although the mechanism proposed by Green is inhibitory). Such a mechanism could account for greater structural alignment in dialogue than in monologue by assuming that alignment is (implicitly or explicitly) more desirable in communicative than in non-communicative situations to promote rapport, liking, or the overall smoothness of the interaction. Indeed, this is the mechanism to which Schoot et al. (2019) attributed their findings (p. 7). We return to attentional and top-down influences on alignment in the General Discussion.

The Present Study

In four experiments, we investigated whether a genuine communicative intention and the physical presence of an interlocutor increase the magnitude of structural alignment. We did this by directly comparing structural alignment between dialogue and monologue. To be clear, we use “monologue” here to mean “absence of interaction and communicative intention,” not “absence of comprehension-production alternation,” as found in other monologic situations like giving a lecture.

The four structural alignment experiments employed a picture-matching task. In the monologue condition of each experiment, participants listened to recorded dative prime
sentences, performed a matching task, and then described pictures of transfer events to no one in particular. In the dialogue condition, participants did the same, except they heard prime sentences from another participant (in Experiments 1, 3 and 4) or an experimenter (in Experiment 2), and the participant or experimenter also matched (or pretended to match) pictures after participants’ descriptions. In Experiment 1, the monologue-dialogue comparison was between subjects, and in Experiments 2-4 this comparison was within-subjects.

In all experiments, we looked at the number of double object descriptions (e.g., *The ballerina gave the soldier the apple*) produced following double object primes versus prepositional dative primes (the alignment effect), and examined how this effect differed between the monologue and dialogue conditions. Note that, as in much previous work (e.g., Branigan et al., 2000), the prepositional dative and double object responses in our study are complementary (the proportions sum up to 100% because we exclude from analyses the small number of responses that did not have any of these two structures); as such, it is not meaningful to report separate analyses of prepositional dative and double object responses. We chose number of double objects as the dependent variable to keep in line with our previous work (e.g., Ivanova et al., 2012a; 2012b; 2017a; 2017b). Our current experiments, however, are not informative about the differential ability of prepositional datives and double objects to induce priming (because we lack a baseline to compare them against), since such a comparison falls outside of the goals of our study.

An additional measure in Experiment 1 were picture description latencies. Latencies can be faster when people produce picture descriptions or sentence completions with the same (versus different) structure as those they produced previously, showing that structural priming brings processing facilitation (Corley & Scheepers, 2002; Segaert, Weber, Cladder-Micus, &
In Experiment 1, we measured description latencies with two goals in mind. First, we needed to determine whether such processing facilitation would occur from (auditory) comprehension to production. If so, we further wanted to determine if such facilitation would be similar in dialogue and monologue, or whether it would be larger in dialogue than in monologue.  

To explore attentional influences on structural alignment generally, and context differences specifically, we additionally analyzed alignment in the pooled data of Experiments 2-4 as a function of (1) individual differences in participants’ attention levels, and (2) the time course of alignment effects within and between experimental blocks. In these analyses, individual differences in attention were operationalized through reaction time variability in the button-press responses to verification pictures in our task. Greater reaction time variability has been associated with less attention in a variety of experimental tasks, including studies of mind-wandering (Cheyne, Solman, Carriere, & Smilek, 2009) and distraction during simulated driving tasks (Wester, Böcker, Volkerts, Verster, & Kenemans, 2008). Accordingly, a global influence of attention on structural alignment thus predicts more overall alignment for less variable participants (by assumption, high-attenders) than for more variable participants (low-attenders). An influence of attention on context differences in alignment predicts that high-attenders would show a bigger alignment difference between monologue and dialogue (driven by dialogue) than low attenders. Time course was operationalized through trial order and block order. In the time-course analyses, an influence of attention on context differences in alignment predicts that alignment would be differentially affected in dialogue and monologue throughout the course of Experiment 1 and Experiments 2-4. These experiments were planned and conducted independently of each other by two different research teams; this is the reason for some methodological differences between the two sets of experiments. Upon discovering that we were addressing the same research questions and that our respective studies fit together well, we decided to combine our separate efforts into a single, more comprehensive project.
the experiment. Specifically, to the extent that both attention and alignment tend to decline throughout an experiment, due to fatigue and adaptation to the distribution of alternatives, respectively, alignment should decline less in dialogue if features of dialogue help sustain attention to a greater extent than features of monologue. These analyses are reported after Experiment 4. We note that our experiments were not designed to address top-down influences on alignment, beyond the basic comparison of context type.

Trial-level data for all experiments are publicly available at https://osf.io/4j6dn/.

Experiment 1

This experiment contrasted the magnitude of structural alignment as well as the production latencies of a group of speakers describing pictures in a monologue situation (after listening to recorded sentences and matching them to pictures) to that of a group of speakers describing pictures to a conversational partner (who produced the sentences to be matched to pictures).

Method

Participants. Ninety-six Northwestern undergraduate students who were native speakers of English participated in this task for partial course credit. In the dialogue condition, 32 individuals participated in the role of test participants (from whom data were collected), and 32 further participants took part in the role of “naïve confederates” (from whom data were not collected). Naïve confederates were randomly paired with test participants, and the individuals in a pair were unacquainted with each other prior to the study. Thirty-two further individuals (tested alone) took part in the monologue condition. In total, data were collected from 64 individuals (32 in the dialogue condition, and 32 in the monologue condition).
Materials. **Dialogue condition.** In the dialogue condition, each individual described one set of items (description set) and, based on their partner’s descriptions, selected images from a separate set of items (selection set). The description set assigned to the test participant consisted of 48 cartoon-like images, originally used in Branigan et al. (2000), that depicted recognizable stock characters (e.g., nun, robber) carrying out simple actions involving an object (see Figure 1). Characters and objects could appear across multiple items, and each picture involved a unique combination of entities. Each picture also included a printed verb that participants were requested to use in their descriptions. The verbs were *give, hand, offer, sell, show,* and *throw.* As in Branigan et al. (2000), twelve images in the test participant’s description set were critical target items. Each target depicted a transfer event involving an agent, a patient, and a recipient (e.g., a nun offering a globe to a monk); there were two different target pictures for each verb.

*Figure 1. Examples of experimental pictures (left: Experiment 1; right: Experiments 2-4).*

The test participant’s description set also contained 36 filler items, which depicted either intransitive actions (e.g., a policeman laughing) or simple transitive actions involving an agent and patient only (e.g., a robber kicking a ball). The entire set of images (experimental and filler) involved 24 unique actions.
The description set assigned to the naïve confederate consisted of 48 scripted sentences that involved the same actors and actions as the test participant’s description set (in different combinations). This included 12 prime sentences and 36 unique filler sentences. The critical primes described transfer events and contained the same six ditransitive verbs as the target items, with two items for each verb. There were two versions of each prime: a prepositional dative sentence (the X is verb-ing the Y to the Z) and a double object sentence (the X is verb-ing the Z the Y), describing the same event (i.e., involving the same characters and object).

Two different item lists were created by rotating the structure of the prime sentences across items, such that primes involving the same entities had the prepositional dative structure in one list and the double object structure in the other list. In each of the two lists, six of the 12 prime sentences involved prepositional dative descriptions, while the other six primes involved double object descriptions. Targets were the same for the two lists and, across the two versions, each target item was preceded by a prepositional dative prime or a double object prime. In Experiment 1, verbs were always different between primes and targets, and the two instances of a given target verb per list was preceded by two different verbs (e.g., GIVE in the target would be preceded once by offer in the prime and once by show in the prime). Experimental prime-target pairings were separated by two to four filler items (involving a sentence produced by the naïve confederate and a picture for description by the test participant). There were no more than two prime sentences involving a particular structure (either prepositional dative or double object) in the critical item order before switching to the other structure on the next prime sentence.

A PowerPoint presentation presented the picture description set for test participants, and two further PowerPoint presentations contained the two lists of scripted sentences for naïve confederates (only one of which was used per experimental session). In these presentations, each
picture or sentence was presented on its own slide. Additional slides presenting the word “LISTEN” informed both roles at the appropriate times when they should listen to their partner’s description in order to locate and remove the matching card from their selection set.

There were two complementary selection sets, one for each participant role. Each set consisted of 48 images that matched the set of events and actions described by the other partner, plus one additional distractor item for each of the 24 verbs in the set. These distractors depicted the same range of entities and actions as the other items. The selection sets for both partners were presented on 3” x 5” index cards, with a single image plus verb per card.

**Monologue condition.** The monologue condition was identical to the dialogue condition, with the following exceptions. Most importantly, there were no naïve confederates. Instead, the PowerPoint presentation for test participants auditorily presented pre-recorded sentences on alternating slides. These sentences were recorded separately by a male speaker in a soundproof booth, following the same script as those used for the naïve confederates from the dialogue condition. This included the two lists, each containing the prepositional dative or double object version of each critical prime sentence, plus the 36 filler sentences. The recordings were saved to separate digital sound files and then embedded in the PowerPoint files at the appropriate places.

**Procedure. Dialogue condition.** Upon arrival at the lab, participants were randomly assigned to roles (identified as “Partner A” and “Partner B”) and told that they would be taking turns helping each other find images within sets of picture cards. Both participants were given a written version of the instructions, which they read separately but sitting together in the same room. These instructions were, unbeknownst to the two participants, slightly different for each role. After participants had read the instructions, the experimenter gave a verbal summary of the
procedures to both of them together that mostly focused on the logistics of the alternation of turns across individuals.

The participants sat on opposite sides of a large table with a barrier between them (so they could not see each other’s cards). Each side of the table contained an open laptop computer plus 24 stacks of cards comprising each participant’s selection set. These stacks were laid out in a 4 x 6 grid in front of each position, with four cards at each location, grouped alphabetically by verb to speed the process of finding the correct cards.

To start the session, participants received instructions explaining that they would be alternating between “DESCRIBE” trials and “FIND” trials. The instructions for FIND trials explained that they should listen carefully to their partner’s description, locate the card in front of them that best matched that description, and remove it from the table. The instructions for DESCRIBE trials varied across role. For test participants, these instructions indicated that the laptop in front of them would present a series of pictures of simple events, each accompanied by a brief tone (included to establish an audible starting point for each trial for the purpose of measuring description latencies). The test participants were told that they were to briefly describe each event so that their partner could find the matching card, using a single sentence containing the present participle form of the verb shown (as an example, they were told that the verb POP would result in a description like “The waitress is popping the balloon.”). For participants in the naïve confederate role, the instructions for DESCRIBE trials informed them that the laptop in front of them would present a series of written sentences that described cards in their partner’s selection set, each accompanied by a brief auditory tone. The confederate’s task was to clearly read this sentence out loud so that their partner could find the corresponding card, and it was noted that these sentences would use the present participle form of verbs printed on their
partner’s matching cards (using the same example of POP and “The waitress is popping the balloon.”). After identifying the picture on each DESCRIBE trial, participants were told to wait for their partner to locate and remove the described card, at which point the roles switched again and it became their turn to find the next card. For FIND trials, they were told to simply listen carefully to their partner’s description, locate the appropriate card in the stacks before them, and remove that card from the tabletop. Participants were informed that their selection sets included more cards than they would need.

For both roles, at no point was any indication given that each participant’s own task on DESCRIBE trials (describing a picture or reading a sentence) differed in any way from their partner’s task. The experimenter simply informed participants that many of the actors and objects would repeat across cards, and that not all of the described actions would be perfectly realistic.

As part of the instructions, both participants were also given a separate “cheat sheet” that presented labelled images of the entire set of characters used in the study materials. The experimenter explained that this would help them efficiently recognize the relevant characters, and that they could consult this sheet as needed for the task. Since some of the characters might not be immediately recognizable and “naïve confederate” participants still had the task of listening to their partner’s descriptions and searching the tabletop for the matching event, it was reasonable for both participants to have access to this “cheat sheet.” As observed by the experimenter, few participants made use of this sheet beyond the first several trials.

When both participants were ready, the experimenter led them through four practice trials before starting the main task. The naïve confederate (as Partner A) always gave the first description, and the description and selection actions alternated from there until all 96 cards were described (48 by each individual). On DESCRIBE trials, the participant responsible for
describing the next item pressed a key on his/her laptop to advance to the next picture to be
described (or sentence to be read). This remained visible for 5 seconds before the PowerPoint
presentation automatically advanced to the next slide introducing a FIND trial with the word
“LISTEN”, to remind participants to wait for their partner’s next description. The experimenter
recorded each session using a digital voice recorder with a stereo microphone placed on the table
equidistant from each participant.

After all of the trials were complete, the experimenter debriefed the participants about the
purpose of the study. This included revealing that a different task was carried out on DESCRIBE
trials by each participant. While a few participants playing the role of naïve confederate
suspected that their partner might have been doing something other than reading sentences
(typically because of the occasional stumble or speech error), none of the test participants
suspected that the partner was not also engaged in describing pictures (although a few did
comment that their partner seemed good at the task). Given that the data of interest come from
utterances produced by the test participants, it is important that they were not suspicious about
the scripted nature of the naïve confederate’s task.

**Monologue condition.** The procedure for the monologue condition was as similar as
possible to the procedure for test participants in the dialogue condition. The most important
difference was that monologue participants were told that this was simply a study of language
comprehension and production. They were informed that they would alternate between locating
cards based on auditory descriptions on FIND trials and describing pictures of simple actions on
DESCRIBE trials. For FIND trials, they were instructed to carefully listen to the pre-recorded
sentences presented via the laptop in order to locate and remove the matching picture cards from
the stacks on the table in front of them. For DESCRIBE trials, they were told that they would be
shown pictures of simple events accompanied by a tone, and that their task was to generate a simple description for each picture using the verb given. The other details in the instructions were the same as for dialogue participants.

Following four practice items, the experimenter answered any questions and then allowed participants to begin the task, starting with the first FIND trial. On FIND trials, each sound file was set to play automatically at the onset of a slide presenting the word “LISTEN.” After 5 seconds, the presentation advanced automatically to the next slide, which asked participants to find and remove the card matching the description they just heard, after which they pressed the spacebar to continue to the next DESCRIBE trial. A digital voice recorder with stereo microphone recorded participants’ responses. When all trials were complete, the experimenter debriefed the participants, explaining that their utterances would be compared to utterances produced by participants in a communicative version of the same task.

**Coding.** To examine the magnitude of structural alignment, participants’ descriptions for the 12 critical targets (768 in total) were coded for whether they used a prepositional dative (493, or 64.2%) or double object construction (271, or 35.3%). Four descriptions could not be coded as either construction, and 36 cases included a different verb from the one intended; these 40 cases (5.2%) were removed from analysis.

To examine the timing of utterances that exhibit alignment or not, we calculated speech onset latencies from the audio-recordings of the critical target descriptions. This was done using Praat to manually identify the onset of the auditory tone that accompanied the start of each critical trial, as well as the onsets of three distinct locations in each utterance: a) the initial speech onset; b) the onset of the noun phrase immediately following the verb (e.g., *the globe* in the prepositional dative sentence *The nun is offering the globe to the monk* or the monk in the double
object sentence *The nun is offering the monk the globe*); and c) the onset of the first postverbal noun (e.g., *globe* in the prepositional dative sentence *The nun is offering the globe to the monk* or *monk* in the double object sentence *The nun is offering the monk the globe*). The second measurement was collected to assess speakers’ readiness to commit to either a prepositional dative or a double object structure at that point in the description, and the third measurement was collected to ensure that this decision was not taken while uttering the definite determiner “the”. Specifically, while Segaert et al. (2014) did find faster utterance-onset latencies for ditransitive sentences in German, there is evidence that utterance planning prior to speech onset extends only to the first phrase (e.g., Smith & Wheeldon, 1999). Since prepositional datives and double objects are identical up to the post-verbal noun phrases, we measured latencies for these phrases as well as the post-verbal nouns to address the possibility that they were planned after speech onset. (In practice, analyses of the post-verbal noun phrases and post-verbal nouns produced very similar results, with one exception noted in the Results section. We therefore only report the latter.)

For the analysis of speech latencies, we removed the 40 descriptions mentioned previously, in which speakers failed to produce a prepositional dative or double object construction or misidentified one or more of the entities involved. We also removed an additional 37 cases involving speech disfluencies (which included both filled pauses, typically occurring at speech onsets, and self-corrections, usually involving a mistaken labeling of a particular actor, e.g., *the painter, uh, monk*). This ensured, as much as possible, that the examination of speech timing would reflect relatively “clean” utterances, and the accurate measurement of speech timing was not interrupted. Altogether, this affected 9.8% of the total data. Then, for each remaining target description, we computed the difference between the onset
of the auditory tone at the start of each trial and the measured speech onset or the measured onset
of the post-verbal noun phrase or post-verbal noun.

**Design and Data Analysis.** Context Type (monologue, dialogue) was manipulated
between-participants and within-items, while Prime Structure (prepositional dative, double
object) was manipulated within both participants and items. For the analysis of speech onset
latencies, we also included an Alignment factor (aligned, unaligned) based on the coding of
target descriptions.

To examine structural alignment, we recoded whether each target description involved a
double object structure or not (1 = double object response; 0 = prepositional dative response).
These data were analyzed using a logistic mixed-effects model, with Context Type (dialogue,
coded as 0.5, and monologue, coded as -0.5), Prime Structure (double object, coded as 0.5, and
prepositional dative, coded as -0.5), and their interaction as fixed predictors. To examine speech
timing, we applied a log transformation to the calculated (initial, noun-phrase or first postverbal
noun) onset latencies, which we then analyzed with a linear mixed-effects model, with Context
Type, Prime Structure and Alignment (aligned, coded as 0.5, and unaligned, coded as -0.5) and
their interactions as fixed predictors. Additional models included Trial order as a continuous
predictor, centered around the mean, to check if adaptation or potential strategies adopted
throughout the course of the experiment affected differently the monologue and dialogue
contexts.

After being set as described above, fixed predictors were mean-centered (though the
balanced design ensured that the manipulated predictors – Context Type and Prime Structure –
did not differ much from +/- 0.5). All models were implemented with the maximal random-
effects structure (Barr, Levy, Scheepers, & Tily, 2013) using the *glmer* or *lmer* functions in the
lmerTest package (version 2.0-33, lme4 version 1.1-13) in R (version 3.4.1), using the Satterthwaite method to approximate denominator degrees of freedom for models with a continuous dependent variable. If the full random-effects model did not converge, the model was simplified by removal of random-effects correlations as a first step, and then step-wise removal of the random effect accounting for least variance, with the restriction that random slopes were removed before random intercepts. The *bobyqa* optimizer was used to facilitate convergence.

**Results and Discussion**

**Structural alignment.** The by-subject means for all experiments are reported in Table 1, and results of the statistical models for Experiment 1 are reported in Table 2. There was an alignment effect: Participants produced more double object descriptions after double object primes (44%) than after prepositional dative primes (27%; Prime Structure was a significant predictor). However, the magnitude of the alignment effect in the dialogue condition was larger than in the monologue condition numerically (11%), but not statistically (the interaction between Context Type and Prime Structure was not significant). Separate models on the data from each context type further indicated that there was a significant alignment effect in each context type [monologue: \( \text{Estimate} = -0.85, SE = 0.43, z = 1.98, p = 0.048 \); dialogue: \( \text{Estimate} = -2.28, SE = 0.72, z = -3.18, p = 0.002 \)].
Table 1

**Mean proportions of double object target descriptions out of all prepositional dative and double object target descriptions in each condition in Experiments 1-4**

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Context Type</th>
<th>Prime Structure</th>
<th>Priming Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Double object</td>
<td>Prepositional dative</td>
</tr>
<tr>
<td>Experiment 1</td>
<td>Monologue</td>
<td>.40</td>
<td>.28</td>
</tr>
<tr>
<td></td>
<td>Dialogue</td>
<td>.48</td>
<td>.25</td>
</tr>
<tr>
<td></td>
<td><strong>Difference</strong></td>
<td><strong>.11</strong></td>
<td></td>
</tr>
<tr>
<td>Experiment 2</td>
<td>Monologue</td>
<td>.49</td>
<td>.15</td>
</tr>
<tr>
<td></td>
<td>Dialogue</td>
<td>.50</td>
<td>.13</td>
</tr>
<tr>
<td></td>
<td><strong>Difference</strong></td>
<td><strong>.04</strong></td>
<td></td>
</tr>
<tr>
<td>Experiment 3</td>
<td>Monologue</td>
<td>.52</td>
<td>.08</td>
</tr>
<tr>
<td></td>
<td>Dialogue</td>
<td>.64</td>
<td>.09</td>
</tr>
<tr>
<td></td>
<td><strong>Difference</strong></td>
<td><strong>.11</strong></td>
<td></td>
</tr>
<tr>
<td>Experiment 4</td>
<td>Monologue</td>
<td>.55</td>
<td>.09</td>
</tr>
<tr>
<td></td>
<td>Dialogue</td>
<td>.58</td>
<td>.07</td>
</tr>
<tr>
<td></td>
<td><strong>Difference</strong></td>
<td><strong>.05</strong></td>
<td></td>
</tr>
</tbody>
</table>

**Speech onset latencies.** Separate models were carried out for initial onset latencies, post-verbal noun-phrase onset latencies and onset latencies for the post-verbal noun after production of the definite determiner. We report here the model results for the initial and post-verbal noun latencies (see Table 2); the model on the post-verbal noun phrase latencies (including the determiner the) produced equivalent results to the model of the noun latencies, except the effect of Prime structure was not significant ($p = .13$). In general, the models revealed effects of the communicative context: Speakers started uttering their descriptions significantly more quickly in the dialogue condition (M = 1.594) than in the monologue condition (M = 1.854), and also
started uttering post-verbal nouns more quickly in dialogue (M = 3.270) than in monologue (M = 3.583). There was no evidence of any differences in onset latencies for descriptions that exhibited alignment (initial onset: M = 1.729; post-verbal noun onset: M = 3.454) compared to descriptions that did not exhibit alignment (initial onset: M = 1.710; post-verbal noun-phrase onset: M = 3.382). Also, post-verbal nouns for double object descriptions were initiated slightly more slowly (M = 3.474) than post-verbal nouns for prepositional dative descriptions (M = 3.372; a main effect of Prime structure for post-verbal noun onset). No other effects were significant.

The analyses including Trial order as predictor did not show any interactions with the rest of the predictors for any of the dependent measures. There was only a significant effect of Trial order for post-verbal noun phrase onset \([-7.647e-03, SE = 3.290e-03, t = -2.32, p = .05]\) and post-verbal noun onset \([Estimate = -7.542e-03, SE = 3.102e-03, t = -2.43, p = .04]\) such that participants sped up the initiation of the post-verbal noun phrases and nouns throughout the course of the experiment, indicating that they got more used to the task.
### Table 2

**LMER results for Experiment 1**

<table>
<thead>
<tr>
<th>Model</th>
<th>Predictors</th>
<th>Estimate</th>
<th>SE</th>
<th>z or t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Syntactic choice</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Context Type</td>
<td>-0.001</td>
<td>.54</td>
<td>-0.03</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Prime Structure</td>
<td>1.57</td>
<td>.49</td>
<td>3.21</td>
<td>.001</td>
<td></td>
</tr>
<tr>
<td>Context Type x Prime Structure</td>
<td>.95</td>
<td>.68</td>
<td>1.40</td>
<td>.16</td>
<td></td>
</tr>
<tr>
<td><strong>Latencies (log RT)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial speech onset</td>
<td>Context Type</td>
<td>-0.065</td>
<td>.021</td>
<td>-3.09</td>
<td>.003</td>
</tr>
<tr>
<td>Prime Structure</td>
<td>.002</td>
<td>.008</td>
<td>.23</td>
<td>.82</td>
<td></td>
</tr>
<tr>
<td>Alignment</td>
<td>.005</td>
<td>.009</td>
<td>.51</td>
<td>.61</td>
<td></td>
</tr>
<tr>
<td>Context Type x Prime Structure</td>
<td>-.005</td>
<td>.017</td>
<td>-.27</td>
<td>.79</td>
<td></td>
</tr>
<tr>
<td>Context Type x Alignment</td>
<td>.006</td>
<td>.020</td>
<td>.29</td>
<td>.78</td>
<td></td>
</tr>
<tr>
<td>Prime Structure x Alignment</td>
<td>-.021</td>
<td>.022</td>
<td>-.98</td>
<td>.33</td>
<td></td>
</tr>
<tr>
<td>Context Type x Prime Structure x Alignment</td>
<td>-.077</td>
<td>.042</td>
<td>-1.84</td>
<td>.07</td>
<td></td>
</tr>
<tr>
<td>Post-verbal noun onset</td>
<td>Context Type</td>
<td>-0.042</td>
<td>.008</td>
<td>-5.36</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Prime Structure</td>
<td>-.013</td>
<td>.007</td>
<td>-1.97</td>
<td>.06</td>
<td></td>
</tr>
<tr>
<td>Alignment</td>
<td>.014</td>
<td>.009</td>
<td>1.51</td>
<td>.16</td>
<td></td>
</tr>
<tr>
<td>Context Type x Prime Structure</td>
<td>.002</td>
<td>.013</td>
<td>.17</td>
<td>.87</td>
<td></td>
</tr>
<tr>
<td>Context Type x Alignment</td>
<td>.012</td>
<td>.017</td>
<td>1.13</td>
<td>.28</td>
<td></td>
</tr>
<tr>
<td>Prime Structure x Alignment</td>
<td>.002</td>
<td>.015</td>
<td>.16</td>
<td>.87</td>
<td></td>
</tr>
<tr>
<td>Context Type x Prime Structure x Alignment</td>
<td>-.011</td>
<td>.027</td>
<td>-.41</td>
<td>.68</td>
<td></td>
</tr>
</tbody>
</table>

*Note: In this and all subsequent tables reporting statistical analyses, grey rows indicate significant effects.*
Experiment 2

This experiment contrasted the magnitude of structural alignment of speakers describing pictures in a non-communicative situation (after listening to sentences recorded by one of two experimenters and matching them to pictures) to that of the same speakers describing pictures to the other of the two experimenters (who produced the sentences to be matched to pictures).

Method

Participants. Thirty-two undergraduate students from the University of California, San Diego (UCSD), took part in exchange for course credit. All were native speakers of English.

Materials. The task and materials in this experiment were overall similar to that of Experiment 1, but this (and all subsequent) experiments were conducted in a different lab, and used a similar but distinct set of materials. As in Experiment 1, the materials had a tripartite structure, consisting in a prime utterance, a verification picture (only one in this experiment, see below) and a target picture. There were 48 critical items and 120 filler items (thus, 168 items in total). Of the critical items, 24 occurred in the dialogue condition and 24 occurred in the monologue condition (which was a between-subject factor in this experiment). In addition, the prime utterances for 24 critical items had a prepositional dative structure and 24 had a double object structure. Thus, there were 12 critical prepositional dative primes and 12 double object primes in the dialogue condition, and 12 critical prepositional dative primes and 12 double object primes in the monologue condition.

Critical target pictures were 48 black-and-white drawings selected from the UCSD Language Production Lab picture database, which depicted transfer events with the same range of cartoon-like characters and objects as in Experiment 1 but drawn by a different artist (e.g., a waitress, a boxer, a pirate, a ballerina; a gun, a shoe, a cake, an apple; see Figure 1). Characters
and objects could appear across multiple items, and each picture involved a unique combination of entities. On half of the target pictures, the agent was on the right and the beneficiary was on the left; on the other half, the positions of the agent and beneficiary were reversed. A verb written in capital letters appeared under the scene depicted on each picture, and participants were requested to use it in their descriptions. The verbs were *fling*, *give*, *hand*, *loan*, *offer*, *sell*, *show* and *throw*, and there were six target pictures for each verb.

Critical prime utterances described the same range of entities as depicted in the target pictures, and contained the same eight verbs. There were two versions of each prime: a prepositional dative sentence (*the X is verb-ing the Y to the Z*) and a double object sentence (*the X is verb-ing the Z the Y*), describing the same event (i.e., involving the same characters and object). Each critical prime utterance was paired with a target picture containing the same verb (e.g., a prime utterance such as *The ballerina is giving the ball to the clown* was paired with a target picture depicting a boxer, a pirate and a book, with the verb *GIVE* presented below the image). Of the six target pictures with the same verb, three were paired with a prime sentence in the double object structure, and three with a prime sentence with a prepositional dative structure.

In the dialogue condition, the prime sentences were read by one of the two female research-assistant experimenters who were native speakers of American English; in the monologue condition, they were recorded by the same experimenters and presented to participants through loudspeakers situated on both sides of the computer screen. If a participant performed the dialogue condition with Experimenter 1, they listened to the recordings made by Experimenter 2 in the monologue condition; for another participant, the experimenter assignment was reversed.

In Experiment 2, the matching task consisted of viewing a single image (the *verification picture*) immediately after each prime utterance, and pressing a button to indicate whether or not
this image depicted the content of the prime sentence. The critical verification pictures contained the same range of characters, objects and actions as in the target pictures. Half of these pictures depicted the same scene as the preceding prime sentence (and thus required a “yes” response), and half differed by one entity from the scene described in the preceding prime sentence (and thus required a “no” response).

The 120 filler items also consisted of a sentence, a verification picture and a picture for description. The fillers employed 18 intransitive verbs and their corresponding actions (e.g., fall, jump, sleep, wink). Forty filler sentences contained a single agent and action (e.g., The doctor is sleeping), 40 sentences contained two agents performing the same action (e.g., The chef and the sailor are jumping), and 40 sentences contained two agents performing two distinct actions (e.g., The boxer is winking and the cowboy is smiling). The same was true for the agents and actions depicted on the pictures for description. Across all items, 84 items (48 critical and 36 fillers) involved the same event type in sentence and picture (i.e., the same number of actions and entities); the remaining 84 fillers items involved different event-type combinations across primes and targets (e.g., a picture of a waitress winking followed a sentence such as The chef and the sailor are jumping). Across all items, 56 (48 critical and 8 filler) repeated verbs between prime sentence and description picture, 56 fillers repeated a noun, and 56 fillers involved different entities. As for critical items, half of the filler verification pictures required a “yes” response, and half required a “no” response.

Sixteen different versions of the materials were created in the following way. There were two context types, dialogue and monologue. Each prime sentence could occur either in the prepositional dative structure or the double object structure across versions (within each version, each prime sentence occurred only once, and each version contained an equal number of
prepositional dative and double object primes). Prime sentences could be read by one of two experimenters in the dialogue block (whereby the primes recorded by the other experimenter would be played to the same participant in the monologue block). Lastly, the 48 critical items were divided into two blocks of 24 items each. The combination of these four factors produced sixteen lists. Two or three filler items occurred at the beginning of each list, and separated each two critical items. The order of fillers was constant across the 16 lists.

To summarize the differences between this experiment and Experiment 1: First, in Experiment 2, each participant completed both the monologue and the dialogue condition, in counterbalanced order. Second, the partner in the dialogue condition was one of two experimenters instead of a naïve confederate; the recordings for the monologue condition were done by the same two experimenters. Third, in all experimental items, primes and targets used the same verbs, to enhance overall alignment (the lexical boost effect: Pickering & Branigan, 1998) and thus maximize the potential to find significant differences between context types. Fourth, the matching task consisted of a single verification picture. Fifth, response latencies for the picture descriptions were not measured, but response times were recorded for the picture verification task.

Procedure. Half of the participants began with the dialogue condition, and half with the monologue condition. The instructions for both conditions began by stating that the purpose of the experiment was to study how people describe pictures. In the dialogue condition, participants were informed that they would take turns with the experimenter to describe pictures to each other, and match them. In the monologue condition, participants were informed that they would take turns to describe pictures and match pictures after hearing recorded descriptions. In both conditions, additional instructions then stepped participants through the procedure. As an
example of a picture description, participants were told that they could see a picture of a waitress and a burglar, with the word “sneeze” written under them, to which their description should be something like, *The waitress and the burglar are sneezing.* Both conditions began with four practice trials.

**Dialogue condition.** An experimental trial began with the experimenter reading a scripted sentence off her screen (which participants believed to be a picture description). During this time, the participant saw a blank screen. The experimenter then pressed a key, which triggered the appearance of a verification picture on the participant’s screen. The participant pressed the M key to indicate a match with the preceding sentence, and the N key to indicate a mismatch. This caused a target picture to appear on her or his screen, which the participant proceeded to describe. Under the pretense that she was matching pictures, the experimenter pressed a key on the keyboard which coded the participant’s response (coding options are explained in detail in the next section). This caused another sentence to appear on her screen, thus initiating the following trial.

**Monologue condition.** An experimental trial began with a recorded sentence playing through the loudspeakers. During this time, the participant saw a blank screen. When the recording was over, a verification picture appeared on the participant’s screen. The participant pressed the M key to indicate a match with the preceding sentence, and the N key to indicate a mismatch. This caused a target picture to appear on her or his screen, which the participant proceeded to describe, then pressed the space bar to initiate the following trial. The experimenter was present in the testing room during the time participants completed the monologue condition and pretended to be reading from a textbook but was actually coding participants’ responses on printed-out coding sheets.
The experiment was programmed in PsyScope X (Bonatti, n.d.; Cohen, MacWhinney, Flatt, & Provost, 1993). Button-press response times on the verification task (included in analyses reported after Experiment 4) were registered by PsyScope. In addition, the experimental session was recorded with a digital voice recorder for coding verification. After the experiment, participants filled out a debriefing questionnaire, asking for background information and language history, and probing their beliefs about the nature of the experiment. After completing the questionnaire, participants were informed about the true purposes of the study.

**Coding.** The experimenters coded participants’ descriptions of the 48 critical targets during each experimental session in real time. For six of the eight target verbs (*give*, *hand*, *loan*, *offer*, *sell*, and *show*), experimenters had the following coding options:

1 - The nun gives the ball to the boxer
2 – The nun gives the boxer the ball
4 – Something else
5 – Not sure

For the verbs *throw* and *fling*, there were five coding options, to ensure that only descriptions involving the traditional version of the dative alternation (containing the preposition *to*) were counted as valid responses:

1 - The nun throws the ball to the boxer
2 – The nun throws the boxer the ball
3 – The nun throws the ball AT the boxer
4 – Something else
5 – Not sure
The experimenters (research assistants) were informed about the experimental hypotheses (so that they could maximally benefit from their research assistantships) but were emphatically instructed to never try to guess what participants had said if they were not sure, but to choose the “Not sure” option instead. Responses coded as “Not sure” were recoded after the experiment from the recordings of participants’ descriptions.

After participants’ descriptions (1,536 in total) were coded in this way by experimenters and verified offline where necessary, they were recoded as prepositional dative sentences (1,039 or 67.6%), double object sentences (479 or 31.2%), or “other” (18 or 1.2%). Descriptions classified as “other” were excluded from analyses. Such responses included cases in which participants produced the wrong verb, including when they initially produced the wrong verb and subsequently self-corrected (to make sure that the verb repetition between prime and target was instantiated properly), as well as prepositional dative descriptions using the preposition at instead of to (to stick to the traditional version of the dative alternation).

**Design and Analysis.** In this and the following experiments, context type (monologue, dialogue) was a within-subject variable. All other aspects of the data analyses were as in Experiment 1.

**Results and Discussion**

The mean by-subject proportions of double object target descriptions out of all prepositional dative and double object target descriptions in each condition are reported in Table 1 and the results of the statistical model for Experiment 2 are reported in Table 3. Participants produced more double object descriptions after hearing double object utterances (49%) than after hearing prepositional dative utterances (14%); that is, there was a robust alignment effect (Prime Structure was a significant predictor). The magnitude of the alignment effect in dialogue was 4%
larger than, but not statistically different from, the alignment effect in the monologue condition (the interaction between Context Type and Prime Structure was not significant).

Table 3

*LMER results for double object descriptions in Experiment 2*

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Estimate</th>
<th>SE</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Context Type</td>
<td>-.27</td>
<td>.43</td>
<td>-.62</td>
<td>.54</td>
</tr>
<tr>
<td>Prime Structure</td>
<td>3.40</td>
<td>.48</td>
<td>7.04</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Context Type x Prime Structure</td>
<td>.79</td>
<td>.86</td>
<td>.92</td>
<td>.36</td>
</tr>
</tbody>
</table>

Experiment 2 replicated the patterns of alignment in Experiment 1: There was a robust alignment effect, but the magnitude of this effect was not statistically different between monologue and dialogue.

There are several aspects of the designs of both Experiments 1 and 2 that may have been suboptimal for finding a differential effect of the monologue and dialogue contexts on alignment. First, the partners in both contexts were reading scripted sentences instead of describing pictures. Reading might have involved flatter intonation, greater speed and fewer disfluencies, rendering the prime utterances different from the way people would naturally describe pictures. Further, when making the monologue-condition recordings in Experiment 2, the experimenters read these sentences for the very first time (so they were least practiced) and also followed instructions to “act like they were describing pictures and hesitate from time to time.” This might have made the recordings in the monologue block more like real language, and the “partner” participation in the dialogue block less like real language. These factors might have made the two context types more similar to each other and obscured any differential alignment effects between them (for some relevant discussion on how such factors might affect participant linguistic behavior, see...
Kuhlen & Brennan, 2012, but note that neither of our two experiments involved deceit about the naïve confederate or experimenter role).

To improve on the design of Experiments 1 and 2, Experiment 3 had (1) a within-subjects design (unlike Experiment 1); (2) dialogue partners who were naïve participants (unlike Experiment 2); and, most importantly, (3) utterances in both monologue and dialogue that were picture descriptions instead of scripted sentences (unlike both Experiments 1 and 2).

**Experiment 3**

This experiment contrasted the magnitude of structural alignment of speakers describing pictures in a non-communicative situation (after listening to recorded picture descriptions and matching them to pictures) to that of the same speakers describing pictures to other naïve participants (who produced picture descriptions instead of reading scripted sentences).

**Method**

**Participants.** Forty-seven UCSD undergraduates participated as test participants, and a further 47 UCSD undergraduates as naïve confederates, in exchange for course credit. Each naïve confederate was randomly paired with a test participant, and the individuals in a pair, with one exception, were unacquainted prior to participating in the experiment. All were native speakers of English. The data for one additional pair was lost due to experimenter error.

**Materials.** These were identical to those in Experiment 2, with the following exceptions. Most importantly, the prime utterances in both the dialogue and monologue conditions were picture descriptions instead of scripted sentences.

To implement this change, 168 new pictures (48 critical and 120 filler) were created to depict the content of the prime sentences used in Experiment 2. These pictures contained red numbers above each entity. This manipulation ensured production of the structure (prepositional
dative or double object) required by the experimental design. For prepositional dative primes, the number 1 was placed above the agent, the number 2 above the object, and the number 3 above the beneficiary; for double object primes, the number 1 was placed above the agent, the number 2 above the beneficiary, and the number 3 above the object.

To create the prime sentences for the monologue condition, a female research assistant described these pictures and her descriptions were digitally recorded.

Procedure. This was the same as in Experiment 2, with the following exceptions. Test participants were randomly assigned to complete the dialogue condition first or the monologue condition first. In the dialogue condition, test participants were led into the testing room and read the written instructions on the computer screen. Naïve confederates read printed instructions in a different room, and, unbeknownst to the test participants, received training to follow the order of the red numbers printed on the pictures when giving their descriptions. When this training was completed, naïve confederates were led into the testing room and the practice trials were initiated. Before the start of the experiment proper, both participants were shown the cartoon-like characters to appear in the experiment, and informed of the characters’ names. The test participant and naïve confederate remained unaware until the end of the experiment that they were performing slightly different tasks (crucially, the test participant did not know that the naïve confederate’s utterances were constrained regarding order of mention). In the monologue condition, test participants read written instructions on the computer screen and then completed the practice trials for this condition. Experimental details were the same as in Experiment 2 except that in the dialogue condition, naïve confederates described pictures instead of reading sentences, and actually performed a matching task in response to the test participants’ descriptions. During both conditions, an experimenter was present in the room, transcribing in
real time test participants’ responses as well as noting whether or not naïve confederates produced the prime descriptions as intended. The whole experimental session was digitally recorded for subsequent verification.

At the end of the experiment, participants completed the post-experimental questionnaire (for test participants, this was the same as in Experiment 2; for naïve confederates, it was a slightly shortened version). After this, both participants were informed about the purposes of the experiment and the naïve confederate manipulation.

**Coding and data analyses.** These were the same as in Experiment 2. In this experiment, there were 2,256 total responses. Of these, 71 trials (3.1%) were excluded because the naïve confederates did not produce the prime description as intended. Of the remainder, 1,432 descriptions (65.5%) were coded as prepositional datives, 711 (32.5%) as double objects, and 42 (1.9%) as others (the latter were excluded from analysis).

**Results and Discussion**

By-subject means are reported in Table 1 and the statistical model results are reported in Table 4. As in Experiments 1 and 2, participants produced more double object descriptions after hearing double object utterances (58%) than after hearing prepositional dative utterances (9%) (Prime Structure was a significant predictor). Most relevant to our aims, the alignment effect was 11% larger in the dialogue than in the monologue condition, but this difference was not significant (i.e., the interaction between Context Type and Prime Structure was not significant).
Table 4

LMER results for double object descriptions in Experiment 3

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Estimate</th>
<th>SE</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Context Type</td>
<td>.58</td>
<td>.40</td>
<td>1.46</td>
<td>.15</td>
</tr>
<tr>
<td>Prime Structure</td>
<td>4.35</td>
<td>.51</td>
<td>8.59</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Context Type x Prime Structure</td>
<td>.78</td>
<td>.77</td>
<td>1.01</td>
<td>.31</td>
</tr>
</tbody>
</table>

In Experiment 4, we attempted to eliminate any remaining differences between the dialogue and monologue contexts. In Experiment 3 (as well as Experiment 1), the monologue block consisted of recordings made by a research assistant (and so were the same for all participants); the partners in the dialogue block were all different individuals. Also, the recorded research assistant was indeed performing the same task as were the naïve confederates, but it could still be that making a recording made her intonation more unnatural, or made her speak more carefully. To make the two context types maximally comparable, in Experiment 4 the live partners from the dialogue condition were recorded and served as the recorded speech in the monologue condition. In this set-up, across participants, the prime utterances in dialogue and monologue were produced by the same speakers.

**Experiment 4**

This experiment contrasted the magnitude of structural alignment in monologue and dialogue (as defined here) by having a naïve participant describe pictures in the dialogue condition for one participant, and then playing their recorded speech in the monologue condition for the following participant. This design ensured that, aside from disruptions noted below, the exact same utterances acted as prime descriptions in the dialogue and monologue conditions, spoken live by a naïve confederate in the dialogue condition and played as a recording (for the subsequent participant) in the monologue condition.
Method

Participants. Forty-eight UCSD undergraduates participated as test participants, and a further 48 UCSD undergraduates as naïve confederates, in exchange for course credit. Each naïve confederate was randomly paired with a test participant, and the individuals in a pair were unacquainted prior to participating in the experiment. All were native speakers of English. The data of four additional test participants were excluded from analyses because the naïve confederates paired with those participants did not correctly produce any double object descriptions (instead converting them to prepositional dative descriptions or producing descriptions such as *The nun gave to the boxer the ball*). We also excluded data from two further test participants because they heard primes recorded by a research assistant instead of a naïve confederate in the monologue condition (see below).

Materials and procedure. These were identical to those in Experiment 3, with the following exceptions. Most importantly, the descriptions of a naïve confederate in the dialogue condition for one test participant served as the recorded prime utterances in the monologue condition for the following test participant.

To implement this change, each naïve confederate’s description in the dialogue condition, including pauses on both ends, was recorded as a separate file during the experimental session. The experiment was programmed to play these recordings in the monologue condition for the following test participant, who performed the dialogue condition with a new naïve confederate. This created testing chains as illustrated in Table 5. There were two such testing chains, corresponding to the two lists each containing a prepositional dative or double object version of each prime utterance. The first test participant in each chain heard recordings from a pre-recorded research assistant (from the materials used in Experiment 3), and their data were
excluded from analyses. The chain had to be broken twice due to the experimenters being initially unaware that a naïve confederate had not produced usable prime sentences in the double object condition. As a result, four test participants had to be replaced after the completion of the experiment. This meant that two test participants performed the monologue condition with recorded descriptions of a naïve confederate from an excluded pair; thus, in two cases, the prime utterances in the monologue condition were made by different individuals than the prime utterances in the dialogue condition for the preceding participant. The order of the dialogue and monologue conditions was counterbalanced, as in our previous within-subject experiments (Experiments 2 and 3). Both naïve confederates and test participants wore a headset microphone for the length of the experimental session.

Table 5

An example of the beginning of a testing chain in Experiment 4

<table>
<thead>
<tr>
<th>Test participant #</th>
<th>Monologue condition (recordings made by)</th>
<th>Dialogue condition partner</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant 1 (excluded)</td>
<td>Research assistant</td>
<td>Naïve confederate 1</td>
</tr>
<tr>
<td>Participant 2</td>
<td>Naïve confederate 1</td>
<td>Naïve confederate 2</td>
</tr>
<tr>
<td>Participant 3</td>
<td>Naïve confederate 2</td>
<td>Naïve confederate 3</td>
</tr>
<tr>
<td>Participant 4</td>
<td>Naïve confederate 3</td>
<td>Naïve confederate 4</td>
</tr>
</tbody>
</table>

*Note. Order of monologue and dialogue conditions was counterbalanced (not reflected here).*

**Coding and data analysis.** These were the same as in Experiment 2. In this experiment, there were 2,304 total responses. 180 trials (7.8%) were excluded because the naïve confederates did not produce prime descriptions with the intended structure. Of the remainder, 1,394 (65.6%)
were coded as prepositional datives, 654 (30.8%) as double objects, and 76 (3.6%) as others (the latter were excluded from analysis).

Results and Discussion

By-subject means are reported in Table 1, and the statistical model results are plotted in Table 6. As in Experiments 1-3, participants produced more double object descriptions after hearing double object utterances (57%) than after hearing prepositional dative utterances (8%) (Prime Structure was a significant predictor). The alignment effect was 5% larger in the dialogue than in the monologue condition but this difference was not significant (i.e., the interaction between Context Type and Prime Structure was not significant).

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Estimate</th>
<th>SE</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Context Type</td>
<td>-.45</td>
<td>.62</td>
<td>-.73</td>
<td>.48</td>
</tr>
<tr>
<td>Prime Structure</td>
<td>5.16</td>
<td>.61</td>
<td>8.51</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Context Type x Prime Structure</td>
<td>1.25</td>
<td>1.19</td>
<td>1.05</td>
<td>.29</td>
</tr>
</tbody>
</table>

Interim Summary

We conducted four structural alignment experiments to determine how the magnitude of structural alignment in dialogue (here, communicative alternation with a live partner) compares to that in monologue (here, non-communicative alternation). Each experiment was designed to eliminate possible confounds in the previous experiments, and Experiment 4 had dialogue and monologue blocks that were maximally comparable. Although numerically there was more alignment in dialogue than in monologue, there were no statistical differences between the two contexts in any experiment. These results seem to suggest that two characteristic features of
prototypical dialogue, communicative intent and an interlocutor’s physical presence, do not
ostensibly facilitate the automatic coupling of linguistic representations that leads to alignment.

Given that we observed consistent numerical differences in the magnitude of alignment
between the two contexts, one might speculate that the effect of context is real, but very small.
We next address the possibility that small context differences were obscured by additional
factors, and we explore the attentional account outlined in the Introduction. In a series of post-
hoc analyses, we look at whether individual differences in attention predicted overall alignment
magnitudes and alignment differences between monologue and dialogue specifically, and
whether alignment followed a different time course in monologue and dialogue.

Effects of Individual Differences in Attention

Our experiments included a picture verification task to ensure that participants processed
the prime utterances. In Experiments 2–4, we can operationalize individual differences in
attention by examining button-press reaction time variability to critical items in this task.5

Many factors have been found to affect reaction time variability, which is larger in older
populations, patient populations who have sustained various kinds of brain injuries, and
individuals who have attention deficit/hyperactivity disorder (for a review, see Flehmig,
Steinborn, Langner, Scholz, & Westhoff, 2007). Reaction time variability in experimental tasks
has also been linked to mind-wandering (Cheyne et al., 2009; Mrazek, Smallwood, & Schooler,
2012) and distraction during simulated driving tasks (Benedetto, Pedrotti, Minin, Baccino, Re, &
Montanari, 2011; Wester et al., 2008). Importantly for our purposes, reaction time variability

5 We note that this task involved an unintended confound between prime structure and
verification response: For half of the participants, all double object primes had verification
pictures that required a “yes” response and all prepositional dative primes had verification
pictures that required a “no” response; for the other half, this assignment was reversed. As
reaction time variability was computed across both “yes” and “no” responses, both prime types
contributed to the variability calculation for every participant.
also appears “to better predict an individual’s capacity to retain an attentional state optimal for task demands than measures of central tendency” such as mean reaction time (Flehmig et al., 2007, p. 136). Thus, although reaction time variability is not purely an attentional measure, it does capture stable individual differences and situational factors (both task-related and task-unrelated) that contribute to attentional focus.

If attentional differences influence the magnitude of alignment, we should see more overall alignment for less variable participants (whom we assume were high-attenders) than for more variable participants (whom we assume were low-attenders). Further, if more attention leads to an enhancement of alignment in dialogue, less variable participants should show a bigger difference between monologue and dialogue in alignment than more variable participants, and that difference should be driven by the magnitude of alignment in dialogue.

We performed these individual differences analyses on the pooled data of Experiments 2-4 for adequate power (excluding Experiment 1, which involved a different verification task; for power considerations for interactions in structural alignment experiments, see Mahowald, James, Futrell, & Gibson, 2016). Since we did not plan to pool the data a priori, we first confirmed that context differences did not statistically differ across experiments [Experiment x Context type x Prime structure interaction: \( p = .90 \)]; details are reported in the Appendix. After discarding trials with RTs less than 300 ms or greater than 10000 ms (0.3% of the data), we log-transformed reaction times to reduce the strength of the correlation between each participant’s reaction time mean and standard deviation across critical trials (cf. Wagenmakers, Grasman, & Molenaar, 2005) [untransformed: adjusted \( r(125) = .83, p < .01 \); log-transformed: adjusted \( r(125) = .15, p = .048 \)]. We then computed the standard deviation for each participant’s reaction times, weighting data equally from the monologue and dialogue blocks. We represented RT variability in this
way, instead of with a coefficient of variability (the standard deviation of RT divided by mean RT; Flehmig et al., 2007; Segalowitz & Segalowitz, 1993), because we believe it is more directly interpretable, and because the two measures were strongly correlated [adjusted \( r(125) = .89, p < .01 \)] (as in Wagenmakers et al., 2005). We also note that the coefficient of variability produced identical results in the analyses below.

We fitted a model on the response type data (1 = double object response; 0 = prepositional dative response) with the fixed predictors RT variability (a continuous, between-subjects variable; mean = 0.35, median = 0.34, SD = 0.06, range = [0.21, 0.60]), Context Type (monologue, dialogue), and Prime Structure, as well as all two- and three-way interactions between those factors. The final model included maximal random effects by subjects and items, but no correlations between random effects. The model results are reported in Table 7, and the alignment percentages for the two context types are plotted in Figure 2 as a function of reaction time variability across context types.

These analyses revealed that participants with lower RT variability in the picture verification task showed greater overall alignment (there was a significant interaction between RT Variability and Prime Structure). In other words, following our operationalization of individual differences in attention in our experiments, participants who paid more attention to the task aligned more than participants who paid less attention to the task. This result suggests that structural alignment can be modulated by the attention paid in a given situation. However, there was no evidence that individual differences in attention brought about differential alignment in monologue and dialogue (the three-way interaction was not significant).\(^6\)

\(^6\) We also conducted analyses with participants’ verification accuracy instead of reaction time variability as a proxy for attention. We assumed that more accurate participants paid more attention to the task than less accurate participants. The only significant effect involving verification accuracy was the interaction with Context Type [\( Estimate = 6.18, \ SE = 2.64, z = \)
Alignment in dialogue and monologue as a function of individual differences in attention (reaction time variability in the picture verification task across context types). For ease of visualization, the proportion of aligned responses is plotted on the y-axis (instead of response type, which was used in analyses).

Table 7

*LMER analyses of the pooled data of Experiments 2-4 with attention represented by RT*

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Estimate</th>
<th>SE</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Context Type</td>
<td>.14</td>
<td>.17</td>
<td>.82</td>
<td>.41</td>
</tr>
<tr>
<td>Prime Structure</td>
<td>4.29</td>
<td>.26</td>
<td>16.24</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>RT Variability</td>
<td>-2.42</td>
<td>2.86</td>
<td>-.85</td>
<td>.40</td>
</tr>
<tr>
<td>Context Type x Prime Structure</td>
<td>.57</td>
<td>.31</td>
<td>1.83</td>
<td>.07</td>
</tr>
<tr>
<td>Context Type x RT Variability</td>
<td>1.38</td>
<td>2.65</td>
<td>.52</td>
<td>.60</td>
</tr>
<tr>
<td>Prime Structure x RT Variability</td>
<td>-7.88</td>
<td>3.60</td>
<td>-2.19</td>
<td>.03</td>
</tr>
<tr>
<td>Context Type x Prime Structure x RT Variability</td>
<td>1.94</td>
<td>5.24</td>
<td>.37</td>
<td>.71</td>
</tr>
</tbody>
</table>

*Note:* RT Variability was calculated as the standard deviation of each participant’s log RT in the picture verification task.

2.34, *p* = .02], such that more accurate participants produced more double object descriptions in dialogue than in monologue. We do not report these analyses in full because verification accuracy is correlated with reaction time variability [adjusted *r*(125) = .23, *p* < .01].
Effects of Time Course

We next considered whether the time course of structural alignment throughout our experiments was differentially modulated in dialogue and monologue. Typically, the magnitude of structural alignment gradually declines throughout an experiment because of reduction of alignment to a particular structure with an increase of the frequency of that structure in the linguistic environment (Jaeger & Snider, 2013). Such an effect is readily explained by error-driven learning models (Chang et al., 2006) by assuming that a structure becomes more expected with increased exposure to it, causing smaller prediction error and thus smaller adjustments of the linguistic system. If more attention leads to more alignment (as our previous analyses suggest) and greater structural alignment in dialogue is related to sustaining attention to prime utterances, alignment should decline more in monologue than in dialogue. Further, since attention can be expected to wane later in the experiment due to fatigue (our experiments lasted up to 1.5 hours), any dialogue-monologue differences might become evident later on.

To address time-course effects, we analyzed the effects of block position on context type in the pooled data of Experiments 2-4 (whether a monologue or a dialogue block was first or second in an experimental session). The model results are reported in Table 8, and alignment effects by context type and block position are shown in Figure 3, Panel A. The statistical model included Block Position (first, second) and its interactions, in addition to Context Type, Prime Structure and their interaction. This analysis revealed that alignment in dialogue and monologue was differentially modulated by block position (the three-way interaction was significant). Separate analyses on the data of the two context types including all predictors except context type and its interactions revealed that alignment significantly declined between blocks in the monologue context [a significant interaction between Block Position and Prime Structure,
Estimate = -0.97, SE = 0.31, z = -3.17, p = .002] but not in the dialogue context [no significant interaction between Block Position and Prime Structure, Estimate = 0.07, SE = 0.40, z = 0.17, p = .866]. Concurrently, separate analyses on the data of the two blocks including all predictors except Block position and its interactions showed significantly greater rates of alignment in dialogue than in monologue in the second block [a significant interaction between Context and Prime Structure, Estimate = 1.17, SE = 0.378, z = 3.07, p = .002], though no difference was evident in the first block [no significant interaction between Context and Prime Structure, Estimate = 0.13, SE = 0.35, z = 0.38, p = .703]. In other words, those participants who did the dialogue task second aligned more than those participants who did the monologue task second (while for the first block there was no difference). A model with Trial order instead of Block position did not produce a significant 3-way interaction, although a visual inspection of Figure 3, Panel B suggests a sharper decline of alignment towards the end of the experiment in monologue than in dialogue.

In addition, the production of double object descriptions and overall alignment decreased (in the latter case, marginally so) from the first block to the second (there was a main effect of Block position, and a marginal Prime structure x Block position interaction). This result is consistent with adaptation to structure frequency in the linguistic environment (Jaeger & Snider, 2013).²

² We also examined the dataset of Schoot et al. (2019) for trial order effects on context (analyses of block order were not possible because context type was a between-subjects variable in their study). In their dataset, alignment (on passives) increased throughout the experiment, unlike in ours (although overall passive production decreased). However, alignment increased to a greater extent in dialogue than in monologue, consistent with our finding that the difference between contexts was greater in the second block than in the first (Figure 3, Panel A). We report how these analyses were conducted in the Appendix. We thank the authors for generously providing their data and analysis code.
Figure 3. Time-course of structural alignment in dialogue and monologue. Panel A: Mean by-subject proportion alignment effects in Experiments 2-4 as a function of serial order of the dialogue and monologue blocks in an experimental session. Panel B: By-subject proportions of double object descriptions after double object and prepositional object primes as a function of trial order in the pooled data of Experiments 2-4. For visualization, trial-level data are LOESS-smoothed. Values were computed from the relevant raw data and do not represent F1 means. Ribbons represent 95% confidence intervals.

Panel A

Panel B
Table 8

**LMER analyses of the pooled data of Experiments 2-4 of the effects of block position**

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Estimate</th>
<th>SE</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Context Type</td>
<td>-.16</td>
<td>.28</td>
<td>-.57</td>
<td>.57</td>
</tr>
<tr>
<td>Prime Structure</td>
<td>4.37</td>
<td>.32</td>
<td>13.67</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Block Position</td>
<td>-.56</td>
<td>.27</td>
<td>-2.02</td>
<td>.04</td>
</tr>
<tr>
<td>Context Type x Prime Structure</td>
<td>1.30</td>
<td>.52</td>
<td>2.53</td>
<td>.01</td>
</tr>
<tr>
<td>Context Type x Block Position</td>
<td>.43</td>
<td>.76</td>
<td>-.56</td>
<td>.57</td>
</tr>
<tr>
<td>Prime Structure x Block Position</td>
<td>-.92</td>
<td>.49</td>
<td>-1.89</td>
<td>.058</td>
</tr>
<tr>
<td>Context Type x Prime Structure x Block Position</td>
<td>2.08</td>
<td>1.03</td>
<td>2.02</td>
<td>.04</td>
</tr>
</tbody>
</table>

*Note:* This model did not have by-subject random slopes for the Context type x Block position and the three-way interactions, because context assignment to block positions varied between subjects.

**Summary of Additional Analyses**

Additional analyses of the pooled data of Experiments 2-4 present some evidence for a role of attention for structural priming in showing that participants who seemed to pay more attention to the task aligned more than participants who seemed to pay less attention to the task. The influence of attention for context effects on alignment, however, was supported by only one of the two analyses: Context effects were not modulated by individual differences in attention, but did emerge as a function of block position.

Specifically, there was less structural alignment in the monologue block when it was second than in the dialogue block when it was second. Such an effect does not find a ready explanation by assuming that it originates within the architecture of the language system alone. It is instead consistent with an attention account. Attention during the second experimental block would be expected to naturally decline due to fatigue. What we found, however, was that
alignment declined in the second block only in the monologue but not in the dialogue context, suggesting that the latter was able to sustain attention to a greater extent.

We also note that two of the pooled-data analyses presented here showed significant context differences in alignment (a significant interaction between Context Type and Prime Structure). We base our conclusions about context effects on the results of the individual experiments because we did not plan to pool the data a priori, all pooled analyses included additional, unplanned predictors, and only some (not all) of these analyses showed a significant interaction.

General Discussion

In this study, we asked if communicative intent and the physical presence of a conversational partner would facilitate the coupling of linguistic representations assumed to take place during alignment. To address this question, we compared structural alignment in dialogue (a communicative turn-taking situation with a live interlocutor) and monologue (a non-communicative situation involving alternation between comprehension and production). Each of four structural alignment experiments manipulating context type between-participants (Experiment 1) and within-participants (Experiments 2-4) failed to show statistically significant differences in the magnitude of structural alignment between monologue and dialogue, nor did Experiment 1 reveal (differential) alignment effects on initiation, post-verbal-noun-phrase and post-verbal-noun latencies in either monologue or dialogue. These results constrain the Interactive Alignment Theory (Pickering & Garrod, 2004), which posits that alignment is modulated by a continuum of dialogic features. Our theoretical contribution is in determining the joint influence of two such features, the presence or absence of a communicative situation and a physically-present conversation partner. Our data suggest that these features do not play a major
role in driving alignment, and that, in both our monologue and dialogue contexts, structural alignment was overwhelmingly driven by linguistic influences on underlying mental representations.

We note that our results contrast with those of Schoot et al. (2019), who found significantly stronger evidence of syntactic alignment in dialogue than in monologue. We are unsure of the reason underlying this discrepancy, beyond methodological differences. The most relevant such differences seem to be in the type of instruction that dialogue participants received: An explicit incentive to collaborate in the study of Schoot et al., versus no explicit mention that collaboration is necessary in ours. Another relevant difference could be in the tested structure types: Alignment in Schoot et al. came from passives, which are rarer than the double object structures used in the present experiments. In implicit learning accounts, the rarer a structure is, the larger the error between predicted and encountered input, resulting in larger adjustments to the language system, and thus more robust alignment. Finally, given that the context effect in their data did increase over the course of the experimental block, it could be significant that their participants named 100 critical pictures, compared with 48 critical pictures named by participants in our Experiments 2-4. Indeed, fitting the model they reported only on data from the first 48 critical trials yields a non-significant interaction between passive priming and context \([Estimate = 0.06, SE = 0.08, z = 0.76, p = .45]\), whereas fitting the same model on data from the last 48 critical trials yields a significant interaction \([Estimate = 0.28, SE = 0.10, z = 2.64, p < .01]\). We think that the two set of results, taken together, are not inconsistent with our conclusions, especially given that the magnitude of the dialogue-monologue difference in Schoot et al.’s study (~4%) was comparable to the context differences in ours (4-11%), and both were small.
Role of attention for structural alignment

In additional analyses, we investigated whether overall alignment effects and context effects in our experiments were modulated by attention. First, we operationalized attentional differences between participants as differences in reaction time variability on the picture verification task they performed after hearing each prime utterance, and assumed that participants with lower variability were paying more attention throughout the experiment than participants with higher variability. This measure was indeed associated with overall differences in alignment (the lower the participant’s reaction time variability in the pooled data of Experiments 2-4, the more they aligned), but not specifically with context differences. Second, we found that context differences in alignment were modulated by experimental time course: There was more alignment in dialogue than in monologue only when each was the second experimental block, but not when each was the first block. If more attention leads to more alignment (consistent with our findings above) and attention tends to decline throughout an experiment, this result may suggest that dialogue is able to sustain attention to a greater extent than monologue. We note, however, that this evidence about the role of attention for (context differences in) structural priming is suggestive but not conclusive, and further experiments designed with this issue in mind would be necessary (see Slevc & Ivanova, 2017).

But what is the precise role of attention for structural alignment? Our post-hoc analyses suggest that more attention might lead to more structural alignment. Consistent with this idea, there is evidence that dividing or re-allocating attention decreases structural alignment in certain situations. Weatherholtz et al. (2014) found that participants in a non-communicative situation showed less structural alignment to recorded speakers with an accent less similar to theirs, and suggested as one of two possible explanations that less attention might be allocated to utterance
form in such cases because attention would be required for decoding the pronunciation differences. In a non-interactive alignment study aimed at teasing apart the effects of conceptual and syntactic influences on structural priming, Bock, Loebell, and Morey (1992) found that participants instructed to attend to the form of prime sentences showed a structural persistence effect in their own picture descriptions, whereas participants instructed to attend to the meaning of prime sentences (and thus presumably paying less attention to form) did not.

These results, together with ours and Branigan et al.’s (2007) finding that addressees show larger structural alignment than side participants, suggest that allocation of focal attention to different aspects of linguistic input might differentially impact structural alignment. More sustained attention to an utterance’s structural form that is non-degraded (i.e., produced by a native speaker without noise) seems to enhance alignment. Such enhanced attention might come from situational aspects that attract and sustain attention (for example, body motion or facial expressions of conversational partners), or speech content (such as an interesting story). On the other hand, attention to other utterance aspects such as uniquely focusing on meaning (at the expense of form) or focusing on the decoding of degraded form (such as for an utterance produced in noise or with an accent different from the speaker’s) might reduce alignment. It is also possible that a degraded form influences the time course of processing, such that deep processing sufficient to give rise to structural alignment is ultimately achieved but takes too long to exert an immediate influence on a speaker’s own production. Our present findings contribute to this issue by suggesting a possible role of overall attention for enhancing structural alignment to non-degraded input, and by showing that (possibly small) differences in processing depth could influence structural alignment magnitudes. These findings are also consistent with the notion that alignment is “post-conscious”, in the sense that attention to the input is necessary for
it to occur (Garrod & Pickering, 2004, p. 10). However, our results suggest that alignment magnitudes are sensitive to differences in depth of processing in a continuous, rather than binary manner. Attention to the input is unlikely to be wholly absent or wholly engaged, but to vary along a continuum.

A role of attention for structural alignment may have implications for the origin of its sensitivity to social and interpersonal factors. A number of studies (Balcetis & Dale, 2005; Heyselaar et al., 2015; Weatherholtz et al., 2014) suggest that structural alignment is sensitive to factors such as the perceived identity or personality characteristics of the other speaker, but the underlying mechanism has been less clear. We suggest that at least some of these findings might stem from differential attention paid to conversational partners in different situations. Attentional differences might depend on the conversational partner’s behavior (for example, an annoyed person might attract more attention than a patient one; Balcetis & Dale, 2005), ideology (one might pay more attention to an ideologically-agreeable diatribe but get caught up in angry thoughts while listening to an ideologically-disagreeable one; Weatherholtz et al., 2014), or facial expressions (a human-like avatar exhibiting facial movement such as eyebrow raising, blinking and smiling might attract more attention than a non-human-like avatar with a blank expression; Heyselaar et al., 2015).

However, other aspects of these studies’ results seem inconsistent with a simple attentional account. For example, in the study of Balcetis and Dale (2005), it remains unclear why an annoyed person would attract more attention than a patient one but a person acting mean would attract less attention than a person acting nice. In the study of Heyselaar et al. (2015), it remains unclear why the non-human-like avatars would not attract more attention than human
conversational partners, to the extent that some of these avatars might have exhibited awkward behaviors (such as eyebrow raising over a longer period than is typical for humans).

In the Introduction, we also entertained the possibility that top-down executive control – another language-external mechanism – may modulate alignment. Such a mechanism could account for situational sensitivity effects on structural alignment by reinforcing alignment when it is desirable (e.g., with a likable conversational partner or one that the speaker wishes to appease) and downregulating alignment when it is undesirable (e.g., with a disagreeable conversational partner or one from whom the speaker wishes to distance herself). Some form of controlled processing could also be the mechanism underlying socially-mediated accounts of alignment (discussed in Branigan, Pickering, Pearson, McLean, & Brown, 2011) such as audience design (Clark & Schaefer, 1987). In such accounts, alignment comes about from the desire to aid communication by using the same utterance features as one’s conversational partner, under the assumption that they will be more likely to understand such features relative to ones they have never produced in a given conversation. Audience design thus involves making judgments about what conversational partners are more versus less likely to understand, and such judgments are influenced by beliefs about the conversational partners’ knowledge and background. A regulatory executive control system would provide an explanation of some forms of alignment in service of audience design (although we note that this might be more common for lexical than for structural alignment, given that structural choices are usually implicit, do not exhibit partner specificity (Ostrand & Ferreira, 2019), and may impact comprehension success to a lesser extent than lexical choices).

But does executive control and attentional regulation of alignment then mean that it is always strategically deployed? We think that it might be so in many cases, but not necessarily. In
our view, executive control could regulate the goal to pay more attention, or manage directly a goal to align. An attentional goal would be determined by specifics of the situation (such as the need to understand instructions well enough to complete a task or to navigate to a location) and enhanced by factors such as time pressure or reward. A direct alignment goal may be driven by a speaker’s liking of or respect for their interlocutor (because they are a friend, an in-group member or a person of higher status), or the desire to enhance rapport and prosociality. This possibility is based on evidence from non-verbal mimicry (alignment being a form of verbal imitation; Chartrand & Van Baaren, 2009) that interpersonal imitation smooths the communication flow and enhances an interlocutor’s liking for the speaker, and is also consistent with Communication Accommodation Theory (Giles & Powesland, 1975). Such motives can lead to a strategic (although perhaps typically non-conscious) deployment of alignment. In addition, enhanced attention could reinforce the goal to align by providing richer input for speakers’ inferences about the benefits of alignment in the current situation (whereby, if these inferences are supported, alignment might be further enhanced). However, alignment need not be strategic, given that it readily occurs (in our findings, to almost the same extent) in non-communicative and non-social situations, where its differential desirability is less clear. Even so, we think that attention could explain alignment modulations in non-communicative situations insofar as attention to relevant input can be externally driven by stimuli rather than internally driven by goals.

Conclusions

In conclusion, we aimed to determine the role of two prototypical dialogue features for structural alignment, thus constraining the Interactive Alignment Theory (Pickering & Garrod, 2004). We found that being in a communicative situation and having a physically present
conversational partner did not strongly enhance the likelihood that speakers would exhibit alignment: In four structural alignment experiments, there were no statistical differences between alignment in monologue (defined here as non-communicative comprehension-production alternation) and dialogue (here, communicative turn-taking with a live interlocutor). Perhaps somewhat surprisingly, our results suggest that these features play at best a minor role in alignment, and that stronger alignment differences between prototypical dialogue and monologue would likely be due to other factors.

Further, post-hoc pooled analyses suggested a relationship between overall structural alignment and attention: Participants who seemed to pay more attention to the experimental task aligned to a greater extent. These analyses also tentatively indicated that dialogue might sustain attention to a greater extent than monologue: Alignment (globally declining throughout the experiment) was greater in dialogue than in monologue only in the second half of the experiment, when attention would be more likely to wane. These results are consistent with a role of attention in structural alignment and suggest a mechanism to explain part of its sensitivity to particular production contexts.
Acknowledgements

This research was supported by an R01 from NICHD (HD051030) awarded to Victor S. Ferreira and by an R01 from NICHD (HD050287) and an R01 from NIDCD (DC011492) awarded to Tamar H. Gollan. Heartfelt thanks go to Chris Schmader for recording the prime sentences in Experiment 1, to Patrick Chambers at Northwestern and Alexander Simmons, Kaylie VanGalder, and Cameron Bunker at Grand Valley, for collecting and coding the data in Experiment 1, to Emily Barker, Cristina Lloyd, Yashna Bowen, Monica Jeong, Lillian Hwang, Katherine Ong, Kristi Cheng and Grace Rasmusson at UCSD, for testing the participants and coding the responses in Experiments 2, 3 and 4, and to Kristi Cheng, Grace Rasmusson and Cristina Lloyd for recording the prime sentences in Experiments 2 and 3. The authors declare no conflict of interest.
References


Appendix

Details of Additional Analyses

**Pooled analyses of Experiments 2-4 with Experiment as Predictor**

For this analysis, Experiment was coded as two numeric variables in the model: For each variable, Experiment 4 was coded as 0.5, Experiment 2 or Experiment 3 was coded as -0.5, and the remaining experiment was coded as 0. To test for the main effect of Experiment (or any interaction that included it), significance testing was performed by comparing the model to a subset model with all levels of that variable removed, yielding a $\chi^2$ test statistic with two degrees of freedom. The first converging model did not include random correlations. The only effect involving Experiment that approached significance was an interaction between Experiment and Prime structure [$\chi^2(2) = 5.77$, $p = .056$], indicating that the overall alignment effect was marginally lower in Experiment 2 (.36) than in Experiments 3 (.50) and 4 (.49). All other $ps > .19$.

**Trial-order Analysis of Schoot et al. (2019)**

We analyzed the data of Schoot et al. (2019) with a logistic mixed-effects model using the same data and model specifications the authors reported in their paper, with three changes. (1) We added a new fixed effect of Trial order (1-100 for critical production trials, mean-centered and divided by 100 so that the effect represents the difference between the beginning and end of the block), as well as accompanying random slopes for both subjects and items, and interactions and correlations with all other random slopes. (2) To reduce unnecessary model complexity, we omitted all fixed and random effects of Partner type (adaptive vs. non-adaptive, a variable that yielded no significant main effects or interactions in their model). (3) We used the
*bobyqa* optimizer to facilitate model convergence. In keeping with their analytic approach, passive production was coded as a ‘success’.

The analysis yielded three significant effects involving trial order. There was a main effect of trial order, indicating that overall passive production decreased over the block \(\text{Estimate} = -2.73, \ SE = 0.23, \ z = -12.11, \ p < .001\]. There was an interaction between Trial order and Context type, indicating that this decrease in passive use was greater in the dialogue condition than in the monologue condition \(\text{Estimate} = -0.63, \ SE = 0.14, \ z = -4.46, \ p < .01\], and an interaction between Passive priming and Trial order, indicating that passive priming increased over the block \(\text{Estimate} = 0.68, \ SE = 0.22, \ z = 3.11, \ p < .01\]. Finally, a three-way interaction between Passive priming, Context, and Trial order indicated that this increase in passive priming over the block was larger in the dialogue condition than in the monologue condition \(\text{Estimate} = 0.56, \ SE = 0.22, \ z = 2.55, \ p = .01\]. Separate analyses conducted for each context revealed that passive priming increased over the block for participants in the dialogue condition \(\text{Estimate} = 1.23, \ SE = 0.32, \ z = 3.85, \ p < .001\] but not for participants in the monologue condition \(\text{Estimate} = 0.08, \ SE = 0.30, \ z = 0.25, \ p = .80\].