Psychological Methods

ALIGN: Analyzing Linguistic Interactions With Generalizable techNiques—A Python Library

Nicholas D. Duran, Alexandra Paxton, and Riccardo Fusaroli

Online First Publication, February 28, 2019. http://dx.doi.org/10.1037/met0000206

American Psychological Association

Psychological Methods

1082-989X/19/$12.00

http://dx.doi.org/10.1037/met0000206

ALIGN: Analyzing Linguistic Interactions With Generalizable techNiques—A Python Library

Nicholas D. Duran
Arizona State University

Alexandra Paxton
University of California, Berkeley

Riccardo Fusaroli
Aarhus University

Abstract

Linguistic alignment (LA) is the tendency during a conversation to reuse each other's linguistic expressions, including lexical, conceptual, or syntactic structures. LA is often argued to be a crucial driver in reciprocal understanding and interpersonal rapport, though its precise dynamics and effects are still controversial. One barrier to more systematic investigation of these effects lies in the diversity in the methods employed to analyze LA, which makes it difficult to integrate and compare results of individual studies. To overcome this issue, we have developed ALIGN (Analyzing Linguistic Interactions with Generalizable techNiques), an open-source Python package to measure LA in conversation (https://pypi.python.org/pypi/align) along with in-depth open-source tutorials hosted on ALIGN's GitHub repository (https://github.com/nickduran/align-linguistic-alignment). Here, we first describe the challenges in the study of LA and outline how ALIGN can address them. We then demonstrate how our analytical protocol can be applied to theory-driven questions using a complex corpus of dialogue (the Devil's Advocate corpus; Duran & Fusaroli, 2017). We close by identifying further challenges and point to future developments of the field.

Translational Abstract

An important field of study involves how to bring about mutual understanding and close interpersonal relationships. The concept of linguistic alignment (LA), which is the tendency to mirror each other's linguistic expressions, is believed to increase such interpersonal rapport. A challenge in this area of research is that there are numerous methods available to study LA, making it difficult to systematically analyze and compare results across studies. The current study presents a new open-source Python package, ALIGN (Analyzing Linguistic Interactions with Generalizable techNiques), to help in analyzing conversation to assess the presence of LA (https://pypi.python.org/pypi/align). We also present in-depth open-source tutorials that are provided on ALIGN's GitHub repository (https://github.com/nickduran/align-linguistic-alignment). We then outline some of the concerns when conducting research on LA, and discuss how the ALIGN package can help deal with these issues. We also show how ALIGN can be used to address theory-driven questions using a

Nicholas D. Duran, School of Social and Behavioral Sciences, Arizona State University; Alexandra Paxton, Institute of Cognitive and Brain Sciences, Berkeley Institute for Data Science, University of California, Berkeley; Riccardo Fusaroli, School of Communication and Culture & School of Culture and Society, Aarhus University.

Alexandra Paxton is now at the Department of Psychological Sciences & Center for the Ecological Study for Perception and Action, University of Connecticut.

The collection of the DA data was supported by the National Science Foundation under a Minority Postdoctoral Research Fellowship [SBE-1103356] to author Nicholas D. Duran. Funding for this project also came from an Aarhus University Interacting Minds Centre seed funding grant in 2013 to authors Riccardo Fusaroli (PI), Nicholas D. Duran, and Alexandra Paxton ("The Linguistic Dynamics of Conflict And Deception"). This project was also funded in part by a National Science Foundation grant [DUE-1660894] to Nicholas D. Duran and a Moore-Sloan Data Science Environments Fellowship to Alexandra Paxton (thanks to grants from the Gordon and Betty Moore Foundation [Grant GBMF3834] and the Alfred P. Sloan Foundation [Grant 2013-10-27] to the University of California, Berkeley). Please note that an early presentation on this work was given in 2015 at the Annual Meeting of the Society for Computers in Psychology. Our thanks go to Rick Dale (University of California, Los Angeles) for his crucial feedback in the development and design of the DA study analyzed in this paper and for his thoughtful conversations about the nature and quantification of alignment. We thank J. P. Gonzales and Josh Espano for helping to collect and transcribe the DA study data while serving as research assistants at the University of California, Merced, and Grace Petersen while a research assistant at Arizona State University. Finally, we would also like to thank Nelle Varoquaux (University of California, Berkeley) for her assistance and advice on Python packaging and Zoe Hopkins (University of Edinburgh) for sharing her early work on automated analyses of syntactic alignment.

Correspondence concerning this article should be addressed to Nicholas D. Duran, School of Social and Behavioral Sciences, Arizona State University, 4701 West Thunderbird Road, Glendale, AZ 85306. E-mail: nicholas.duran@asu.edu

complex collection of texts (e.g., the Devil's Advocate corpus; Duran & Fusaroli, 2017). Finally, we offer a summary of other considerations and future directions in this area of research.

Keywords: linguistic alignment, interpersonal coordination, automated text analysis, deception, conflict

During conversation, interlocutors are much like skilled dancers or improvisational musicians: Each person's actions depend upon what their partner has done, is doing, and is anticipated to do. When speakers choose a particular set of lexical forms to express an idea, sequence these words into some syntactic structure, and shape the sounds to communicate, these are actions—whether the speakers realize it or not—that are being performed in coordination with one another.

One way that this coordination has been revealed is in interlocutors' tendency to reuse each other's linguistic behaviors, including low-level lexical, syntactic, and phonetic forms (Babel, 2012; Branigan, Pickering, McLean, & Cleland, 2007; Brennan & Clark, 1996). This phenomenon—often known as linguistic alignment (LA)[1]—has provided new insights into language use, revealing the extent to which communication partners' comprehension and production systems are interconnected. Although there are a number of competing accounts to explain LA, this language behavior is generally theorized as a key mechanism in explaining the ease and speed with which people create, express, and maintain a sense of shared understanding in conversation (Branigan & Pickering, 2017; Pickering & Garrod, 2004, 2006).

However popular, the variety of methods used to study LA has resulted in a number of debates and unresolved issues. In the sections that follow, we touch upon some of the open questions and what would be required to resolve them. Of most relevance for the current research, we focus on three issues as they pertain to lexical/conceptual and syntactic LA: (a) how different conversational goals and social contexts (including comparison between experimental and naturalistic contexts) can shape the emergence of alignment; (b) how alignment can vary over time and linguistic levels; and (c) how the functional role of alignment on reciprocal understanding and interaction quality can change across contexts (for broader discussion of these issues, see Dale, Fusaroli, Duran, & Richardson, 2013; Duran & Fusaroli, 2017; Fusaroli, Rączaszek-Leonardi, & Tylén, 2014; Healey, Purver, & Howes, 2014; Paxton, Dale, & Richardson, 2016).

We argue that a productive way forward is to draw upon advances in natural language processing to create standardized methods that would allow for a more uniform comparison of results across studies. Ideally, these methods should be easy to use, allow for parameterization that can capture wide-ranging alignment characteristics, and—perhaps most importantly—have the potential to bridge analyses of laboratory data and collections of large-scale naturalistic or naturally occurring texts (i.e., corpora).

Understanding Context-Sensitivity in Alignment

A major focus for LA research is in understanding the degree to which the social environment and varying communicative conditions might mediate alignment. According to one influential account, these factors do not necessarily play an immediate role, as LA is thought to automatically occur via a direct perception-action link that is largely insulated from conscious or strategic social goals (Pickering & Garrod, 2004).

Recently, researchers have begun to challenge this claim by turning to more contextually rich paradigms. Rather than completely eschewing traditional paradigms that present scripted linguistic stimuli to participants in passive listening or simple turn-taking tasks, researchers have, for example, begun extending these tasks by manipulating participants' perceptions about the needs or mental states of the person communicating the critical linguistic information (Schoot, Heyselaar, Hagoort, & Segaert, 2016; Weatherholtz, Campbell-Kibler, & Jaeger, 2014). For example, in Balcetis and Dale (2005), participants heard a partner describe a series of images in a way that was meant to prime a particular syntactic structure (i.e., a "structural priming" task; Bock, 1986). As with traditional paradigms, this incoming linguistic stimulus is thought to activate similar linguistic forms in the mind of the listener, thus making the linguistic forms more accessible for the listener to produce (Gries, 2005; Pickering & Ferreira, 2008). Going one step forward, Balcetis and Dale (2005) also manipulated participants' perceptions of social distance and affiliation with the partner and found that—in subsequent descriptions of images—participants were more likely to converge on the same syntactic structures as partners they found to be "nice," suggesting a mediating role due to top-down social processes.

However, even with more contextually rich paradigms, social-psychological factors do not always play a strong mediating role. For example, in studies examining people with a range of social and perspective-taking abilities, such as those with autism spectrum disorder, research has found that these abilities have little to no impact on LA (Allen, Haywood, Rajendran, & Branigan, 2011; Hopkins, Yuill, & Keller, 2015; Hopkins, Yuill, & Branigan, 2017; Slocombe et al., 2013). Moreover, given a particular context, LA might be both mediated by a speaker's beliefs and simultaneously driven by automatic processes (e.g., linguistic priming), where greater mediation might be most sensitive to increased demands for communicative effectiveness (Branigan, Pickering, Pearson, & McLean, 2010).

Partially as a result of the rise of more contextually rich paradigms, researchers have also begun to adopt a more nuanced view of the processes and constraints through which LA emerges. One notable approach has been a dynamical systems perspective on interpersonal dynamics. Influenced by research on motor control and interpersonal movement dynamics (Riley, Richardson, Shockley, & Ramenzoni, 2011), recent research on nonverbal alignment (Paxton & Dale, 2017; Ramseyer & Tschacher, 2014; Wang &

[1] We recognize that the phenomenon is also known by a variety of other names, including accommodation, convergence, coordination, coupling, mimicry, and synchrony (Dale et al., 2013; Paxton, 2016; Paxton & Dale, 2013). We use here "alignment" in the generic sense of re-use of linguistic forms employed by one's interlocutor(s), inspired by the definition in Pickering and Garrod (2004).

Hamilton, 2012) and collaboration models of dialog (Brennan, Galati, & Kuhlen, 2010; Duran, Dale, & Galati, 2016; Fusaroli, Raczaszek-Leonardi, & Tylén, 2014; Paxton et al., 2016) have begun to argue for the idea that interpersonal dynamics—including LA—are emergent complex behaviors that are sensitive to contextual, individual, and interpersonal factors. This perspective argues for the central role of socially driven, top-down goals in mediating how subtle behaviors unfold during interpersonal interaction (Abney, Paxton, Dale, & Kello, 2014; Duran & Fusaroli, 2017) but also acknowledges that these influences do not (and, arguably, should not) hold in all cases—a hallmark of the language system's adaptability and context-sensitivity (Duran & Dale, 2014; Duran et al., 2016; Paxton et al., 2016).

Exploring Alignment Through Naturalistic Dialogue

Even as LA research grows to explore new questions about context and mechanisms, many studies continue to rely on tightly controlled and highly structured interactional contexts. From describing pictures (e.g., Branigan et al., 2007; Branigan, Pickering, Pearson, McLean, & Brown, 2011) to navigating a map or maze (e.g., Foltz et al., 2015; Garrod & Anderson, 1987; Reitter & Moore, 2014), these structured experimental tasks, by their very nature, constrain participants' linguistic behaviors and tend to emphasize cooperative turn-taking behaviors (e.g., referential communication tasks; Bangerter & Clark, 2003; Brennan & Clark, 1996). These tasks have been foundational in our understanding of how people draw upon mutual knowledge and how LA emerges, but they provide insight into only a small slice of the range of possible human behaviors.

Researchers have begun to consider the impact that the task demands of traditional experimental tasks may have on our understanding of LA, going so far as to raise the possibilities that LA may not be not as ubiquitous as commonly thought or that the mediating factors on alignment may simply be experimental artifacts (Healey et al., 2014; Hopkins et al., 2015; Howes, Healey, & Purver, 2010; Slocombe et al., 2013). As a result, there is an increasing understanding that we must examine LA in more open-ended and naturalistic dialogue to understand how it unfolds in real-world settings (Fusaroli et al., 2017).

Corpora-based studies are relatively few; have not provided entirely consistent findings or methods; and deal mostly with syntactic alignment. However, those that exist have provided valuable insights. Evidence for syntactic alignment appears to be most robust when communication focuses on specialized and formal topics (e.g., legal cross examination, broadcast interviews, online health discussions, Gries, 2005; Wang, Reitter, & Yen, 2014; school and university essays, Szmrecsanyi, 2005) or on task-based interactions (e.g., route descriptions, Reitter, Moore, & Keller, 2006; card sorting game, Slocombe et al., 2013). In these more controlled domains, there also appears to be an inverse relationship between alignment and the frequency of syntactic structures, such that greater alignment occurs across lower frequency structures. Given these structures are less predictable and thus more challenging to process, the likelihood of reusing a low-frequency structure is higher because of the proposed cognitive economy in doing so (Jaeger & Snider, 2013; Reitter et al., 2006; Scheepers, 2003).

In contrast, for less restricted natural dialogue, evidence of syntactical and lexical LA has been found to be absent or even to occur below what would be expected by chance (Howes et al., 2010). As argued by Healey, Purver, and Howes (2014) and others (Mills, 2014), language users in actual conversations have communication goals that extend beyond those that might lead to automatic alignment: The expressivity and productivity of language is routinely achieved by contrasting and expanding upon ideas, providing counterexamples, and creative deviation from shared assumptions. Of course, this does not occur in all cases. For instance, Moscoso del Prado and Du Bois's (2015) information-theoretic approach has shown evidence of syntactic alignment in spontaneous dialogue—but only when taking into account whether speakers share affective states.

Together these results suggest that the emergence of alignment is sensitive to strategic and contextual factors, but whether or not these effects generalize to real-world language use (and under what conditions) is still very much an open question (see Pietsch, Buch, & Kopp, 2012; Reitter et al., 2006). One of the challenges for corpus-based accounts is to explain how variation might be systematically delineated across a rich array of conversational and social goals.

Variation of Alignment Over Linguistic Levels and Over Time

Other challenges for evaluating LA in experimental and naturalistic interactions are centered on (a) how it varies across time and (b) whether its expression is similar across linguistic levels (lexical, syntactic, conceptual). The first of these two challenges asks whether we should expect alignment to increase or decrease as a conversation progresses (cf. Fusaroli et al., 2017; for similar issues in movement alignment, cf. Ramseyer & Tschacher, 2014). Prevailing mechanistic accounts mostly assume increased alignment as an interaction proceeds: As individuals' mental models become increasingly aligned and/or rapport grows, so should the alignment of how those models are linguistically expressed. Consistent with this account, experimental evidence shows interlocutors converging on shared expressions over interactional discourse also increase shared linguistic routines (Foltz et al., 2015; Friedberg, Litman, & Paletz, 2012). In the case of lexical alignment, this is well-evidenced in experimental tasks where conversational partners must negotiate unique referential descriptors for abstract geometric images that are inherently ambiguous in their representation. As partners converge on a shared conceptualization over a series of conversational turns, parallel convergence occurs in their lexical choices despite other options that would have been more informative (Brennan & Clark, 1996; Brennan et al., 2010).

Moreover, the idea of cumulative structural priming, whereby repeated exposure to particular syntactic structures makes their production more likely later on (Jaeger & Snider, 2008, 2013; Kaschak, Loney, & Borreggine, 2006), suggests a long-term adaptation that should result in increasing syntactic alignment over time. However, little support has been found in nascent attempts to systematically explore this possibility in task-based conversational interactions (Carbary & Tanenhaus, 2011; Foltz et al., 2015).

An assumption of increasing alignment should also be tempered with the observation that once a shared understanding has been established—indexed by high linguistic alignment—increasing or even maintaining current levels of LA over the course of a conversation may not provide further benefit. Essentially, establishing

a shared understanding may provide interlocutors with greater license for expressivity and deviation from previously aligned forms, resulting in decreasing or perhaps fluctuating LA after an initial increase from the start of the conversation. Although very few studies have explored this issue, early evidence from online communication suggests that both syntactic and lexical alignment can decrease over time (Wang et al., 2014). Given that this account is also supported by emerging empirical evidence (Fusaroli et al., 2012, 2017; Healey et al., 2014), further work should include systematic cross-corpus and cross-method investigations to add clarity to questions about temporal developments in LA throughout a conversation and their effects on the quality of the interaction.

The second challenge asks whether LA is uniquely expressed across different linguistic levels and how this interacts with its changes over time. By emphasizing a direct perception-action link that is unmediated by high-level social factors, the aforementioned priming account posits that syntactic, lexical, and conceptual alignment are expressions of the same mechanism, copresent and (arguably) boosting each other (Pickering & Garrod, 2004; Rowland, Chang, Ambridge, Pine, & Lieven, 2012). However, a dynamical systems account of LA would suggest that different linguistic levels might be shaped by different contextual pressures and could therefore exhibit different dynamics (Paxton et al., 2016): For instance, Hopkins, Yuill, and Keller (2015) argue that social goals have greater influence over lexical than over syntactic or even conceptual alignment, because speakers are more aware of their lexical choices and alignment with a partner. Likewise, syntactic alignment might be less susceptible to strategic goals and might instead be mediated by more automatic and unconscious demands (Pickering & Garrod, 2006). To resolve these differing perspectives, the study of LA needs systematic, temporally extended, and multilevel research that can assess and compare alignment along more than one dimension.

The Emergence and Functional Role of Alignment

Analogous to issues in the variation of alignment over levels and time, there are seemingly contradictory findings about the actual function of alignment across behaviors, including language. One major perspective places the emphasis on reciprocal understanding and rapport that can bond people together and improve communication, both influencing and reflecting the state of their interpersonal relationship. This perspective has been most clearly demonstrated in the domain of "nonverbal alignment"[2] (Lakin, Jefferis, Cheng, & Chartrand, 2003; Miles, Nind, Henderson, & Macrae, 2010) and in phonetic convergence (Babel, 2012; Pardo, Gibbons, Suppes, & Krauss, 2012), but recent work suggests it may play a similar role in lexical and syntactic LA (e.g., Balcetis & Dale, 2005; Coyle & Kaschak, 2012; Lev-Ari, 2015). For example, greater lexical LA has been associated to better personal relationships and even more successful negotiation resolution (Ireland & Henderson, 2014; Ireland et al., 2011).

Other studies—from both linguistic and nonlinguistic domains—seem to paint a different picture. For example, some kinds of alignment have been linked to worse joint performance (Fusaroli et al., 2012),[3] deception and conflict (Duran & Fusaroli, 2017; Ireland et al., 2011; Main, Paxton, & Dale, 2016; Paxton & Dale, 2013), in-group and out-group dynamics (Miles, Lumsden, Richardson, & Macrae, 2011; Yabar, Johnston, Miles, & Peace, 2006), and social awkwardness (Michael et al., 2015). From this work has come a new but growing perspective on interpersonal alignment that, much like the dynamical systems account (Fusaroli et al., 2014; Paxton et al., 2016; Riley et al., 2011), places a major emphasis on the contextual dependencies that influence alignment's emergence and function. Even so, to get a better grasp on these factors, greater consideration is needed of all the concerns outlined above: sensitivity in experimental and natural dialogue, its temporal dynamics, and its consistency and interplay across linguistic levels. Only by considering these together and across comparable conditions can we disentangle its functional role.

ALIGN: Analyzing Linguistic Interactions With Generalizable techNiques

One way to facilitate a comparative and theory-driven analysis of alignment is to provide easy-to-use tools for systematically evaluating LA within extended dialogue. These tools must provide comparable assessments across studies by capturing multiple levels of linguistic complexity, giving a unifying framework across different parameter choices, and providing a common basis for quantification and evaluation against controls and baselines. This is important because it not only encourages reproducible research—yielding the same alignment scores, regardless of the researcher—but would also contribute to the broader theoretical goals of the field outlined above: finding common trends across social and communicative conditions that might modulate alignment across linguistic levels and time.

To this end, we introduce ALIGN, a Python package using simple and intuitive code to quantitatively and reproducibly measure turn-by-turn alignment across syntactic, lexical, and conceptual levels of language. It combines well-established and cutting-edge components from natural language processing and uses proven analytical techniques so that researchers can quickly and easily deploy, analyze, and understand their data. In the interest of open science, the underlying ALIGN package is freely available on PyPI (https://pypi.python.org/pypi/align) and on GitHub with MIT license (https://github.com/nickduran/align-linguistic-alignment). Moreover, accompanying the package on the GitHub repository, we have created open-source tutorials within Jupyter notebooks to painlessly introduce ALIGN even to those without much programming experience. We provide the package code and example Jupyter notebooks to encourage community development and to inspire other researchers to openly share their tools and algorithms.

In what follows, we provide a detailed account of how ALIGN is implemented, how to deploy it, and how to interpret its basic results. Along the way, we provide justification for various processing choices, grounded in previous research. We then apply our

[2] As with linguistic alignment, we recognize that there are multifaceted interpretations of this phenomenon within the nonverbal domain. We use "alignment" in the generic sense as a type of interpersonal coordination of movement through interaction, akin to the broader context of interpretation used by Chartrand and Lakin (2013).

[3] Note that Fusaroli et al. (2012) report that alignment of confidence expressions is positively related to joint performance (in line with previous studies focusing on specific lexical expressions), but a more general measure of lexical alignment (or indiscriminate lexical alignment) is negatively related to joint performance.

tool to a corpus of naturalistic and extended face-to-face dialogue where multiple high-level conversational demands are present. We then end with a discussion on how ALIGN contributes to LA research and computational social science in general, with focus on best practices and recommendations for use, as well as necessary next steps to account for more nuanced linguistic information.

A General Approach for Quantifying Multilevel Linguistic Alignment in Discourse

Overview

ALIGN involves two primary phases, with the assumption that the data have been properly prepared for analysis (i.e., Phase 0). In Phase 1, the data are automatically cleaned, standardized, and correctly ordered to produce participants' speech turns that are uniformly interleaved (so that Participant 1 followed by Participant 2 followed by Participant 1, and so forth). In this phase, what was said in each turn is transformed into contiguous sequences of tokens; from these forms, continuous sequences of part-of-speech (or *POS*) tags are generated. After this, a control baseline (or *surrogate*) is created for each conversation by artificially joining interlocutors from different conversations into *surrogate pairs*—or fabricated adjacent turns between participants who never truly interacted in the experimental context.

In Phase 2, scores for lexical, syntactic, and conceptual alignment are generated for each turn-by-turn exchange in both real and surrogate conversations. For lexical and syntactic forms, each interlocutor's contiguous sequences of tokens (for lexical) and POS tags (for syntactic) are segmented into short chunks of increasing sizes, also known as "*n*-grams."[4] ALIGN then generates frequency vectors composed of the number of times unique *n*-grams occur for each interlocutor, within each turn; those vectors are then used to create a similarity score (or *cosine similarity*) by calculating the angle between the two vectors (as a single turn-by-turn exchange). It is important to note that the lexical and syntactic alignment captured here is indiscriminate insofar that no particular phrase or syntactic structure is specifically targeted.

For the conceptual analysis, content words from each interlocutor's turn are transformed into a high-dimensional semantic (HDS) representation using Google's word2vec (Řehůřek & Sojka, 2010). The resulting HDS vector is compared with a partner's turn via cosine similarity to obtain the conceptual similarity score.

The end result for each conversation is multiple levels of LA—all based on the same cosine scale—that occur at a local, turn-by-turn level over time. This allows for evaluations of how alignment occurs across linguistic levels, across time, and across various structural sizes (i.e., based on *n*-gram size). For each of these, ALIGN also captures directionality between interlocutors (e.g., alignment scores based on Participant 1's turns followed by Participant 2, or Participant 2's turns followed by Participant 1). Crucially, the use of control surrogate pairs allows the researcher to assess whether the phenomena observed are actually due to the interaction dynamics or to other confounds (e.g., word frequency, constraints given by the task at hand). ALIGN also allows for these analyses to be performed over the entire conversation (rather than partners' consecutive turns), although we do not discuss this here.

Phase 0: Preparing Conversational Data for ALIGN

ALIGN is optimized for dialogue corpora composed of extended back-and-forth exchanges between two people, whether these come from laboratory experiments or naturally occurring data. It requires a simple dataset of individual conversations with each saved as an $N \times 2$ matrix. Each N row corresponds to a speech turn of the current speaker, temporally ordered based on the speech turn's occurrence in the conversation, alternating between speakers. A minimal set of information is required for each turn: identifier of the current speaker (column labeled as "Participant") and word-level transcription of the utterance (column labeled as "Content"). Table 1 provides an example of the necessary corpora format for ALIGN compatibility.

ALIGN also assumes that the filename of each conversation contains the dyad and conversation identifier codes unique to each study. Formatting and transcription can be accomplished with several open-source software options (e.g., Praat: Boersma, 2001; Transana: Woods & Fassnacht, 2017; ELAN: Sloetjes & Wittenburg, 2008; ANVIL: Kipp, 2012). We provide an example of how to generate Table 1 with Praat (v. 6.0.36) in the Appendix.

Phase 1: Preprocessing at the Utterance Level

Transcription standardization. ALIGN presents several independent preprocessing options:

- removing all numbers, punctuation, and other non-ASCII alphabet characters;
- removing common speech fillers (e.g.,"um," "huh," "yeah");
- removing additional, user-specified words or fillers;
- performing basic automatic spell correction.[5] (Jurafsky & Martin, 2009);
- removing turns shorter than a user-specified number of words; and
- checking that each row in conversation alternates between unique speakers, and if not, merging contiguous turns from the same speaker into a single turn, preserving order.

We recommend that each researcher make standardization choices based on the best-practices in their field and with their research goals in mind. For example, backchanneling—that is, short utterances providing support and scaffolding to another speaker's utterance—may be of interest to some researchers; choosing to remove turns shorter than one word would remove this information (if available). Alternatively, if a researcher is specifically interested in exploring syntactical structure, removing short turns may add clarity to the analysis, because short turns tend to provide little substantial content or syntactical variation when transformed into *n*-grams.

[4] For those unfamiliar with *n*-grams, we will provide a brief demonstration using the sentence preceding this footnote. Each *n*-gram is bracketed for clarity of the demonstration. In that sentence, the initial lexical *bi*grams of the current sentence would be ["for lexical"], ["lexical and"], and ["and syntactic"]; the POS bigrams would be ["IN JJ"], ["JJ CC"], and ["CC JJ"]. The initial *tri*grams would be ["for lexical and"] and ["lexical and syntactic"]; the POS trigrams would be ["IN JJ CC"] and ["JJ CC JJ"].

[5] Method derived from Peter Norvig's code and training corpus publicly available at: http://norvig.com/spell-correct.html.

Table 1
Example of ALIGN Data File Input

Participant	Content
1	Okay so do you think marijuana should be legalized or not?
2	Yeah, I think it should.
1	Uh, why do you think it should be legalized?
2	I think it should be legalized because a lot of people use it illegally . . .
1	Uh, personally I think it should not be legalized. # because well hmm . . .
2	No absolutely not. Honestly that, no. I feel like uh alcohol . . .

Note. The ALIGN data file requires minimally two pieces of information: an identifier for the speaker ("Participant" column) and a word-level transcription of the utterance ("Content" column). Note that order of rows correspond to the order in which each turn was spoken, alternating between speakers.

Tokenize, lemmatize, and tag parts of speech. In order to correctly identify the lexical and syntactic units in each utterance, the word-level transcription is analyzed word by word (tokenized) and lemmatized (optional)—in other words, removing grammatical markers such as number and tense from each word (e.g., "are" becomes "be;" "dogs" becomes "dog"). Finally, each lemma is tagged according to its part of speech (e.g., noun, verb). These procedures are necessary to ensure control over what is considered alignment (e.g., making "dog" an instance of alignment to "dogs," because the base morpheme is reused).

To implement these three steps, ALIGN relies on the Natural Language Toolkit (NLTK, Version 3.2.5; Bird, Klein, & Loper, 2009). Tokenization and lemmatization are performed by the NLTK WordNet Lemmatizer library. POS tagging is implemented by two alternative part-of-speech taggers using the Penn Treebank tagset (Marcus, Marcinkiewicz, & Santorini, 1993): NLTK's default "averaged perceptron tagger" and the Stanford log-linear tagger (Toutanova, Klein, Manning, & Singer, 2003).[6] See Table 2 for an example of the output of this process.

Phase 2: Compute Turn-by-Turn Alignment Scores

Lexical and syntactic alignment. For each conversational turn (i.e., each row in Table 2), lexical and POS sequences are converted into *n*-grams up to a user-specified length, ranging by default from uni- to quadgrams. The frequency of each *n*-gram within each turn level is then computed and represented in vector form (see Table 3 for an example of bigram conversion for one conversational turn exchange).

For syntactic vectors, ALIGN by default includes only the POS *n*-grams that do not share an underlying lexical representation—that is, verbatim repetition of lexical units can be removed. This default option provides a stricter test of syntactic alignment, as it removes the "lexical boost" that occurs with repetition of lexical content between adjacent turns (Cleland & Pickering, 2003; Healey et al., 2014; Pickering & Branigan, 1998; Reitter, Keller, & Moore, 2011; Rowland et al., 2012). However, researchers can always override this option and include lexically shared POS *n*-grams if desired. One reason for overriding the default is that researchers may prefer to simply partial out a variable of lexical alignment when doing statistical modeling or may want to explicitly evaluate the contribution of shared linguistic content to syntactic alignment values within and across tasks.

Lexical and syntactic alignment scores are then generated on a turn-by-turn basis by taking the cosine similarity between interlocutors' contiguous turns, resulting in a score ranging from 0 to 1 for each *n*-gram vectorized structure, with higher scores indicating greater relatedness and thus greater alignment across turns (see Table 4). In addition to this straightforward interpretation of alignment, cosine similarity has the advantage of controlling for differences in utterance length by normalizing counts across comparison utterances.

Beyond taking advantage of its interpretive and statistical properties, we also chose to use cosine similarity because of its long history in natural language processing and related areas of information retrieval and text mining (Jurafsky & Martin, 2009; Tan, Steinbach, & Kumar, 2005). It is a proven measure that remains popular due to its ease of use, effectiveness, and ability to be applied in diverse domains, including existing measures of conversational LA (Hopkins et al., 2015; Manning, Raghavan, & Schütze, 2008).

ALIGN also allows measures of lexical and syntactic alignment to be conducted on a wide range of lexical and syntactic structures, rather than selecting a few target structures (Foltz et al., 2015). Moreover, syntactic alignment based on *n*-gram sequences—rather than on nonterminal syntactic rules—has proven to be an efficient and simple way for capturing the inherent ordering of elements in syntax from the smallest window (bigrams) to increasing windows of complexity (e.g., tri- and quadgrams; Dale & Spivey, 2006; Hopkins et al., 2015). Given the minimal assumptions involved, ALIGN can be easily generalized to fit many research needs, can be easily interpreted by researchers, and can be easily compared across studies if authors share their analysis parameters.

Conceptual alignment. To go beyond lexical alignment, the ALIGN method also quantifies the conceptual alignment of utterances based on their proximity to each other in a high-dimensional semantic space (Bengio, Ducharme, Vincent, & Jauvin, 2003; Osgood, Suci, & Tannenbaum, 1964). This process begins by loading in a prebuilt semantic space or by building a semantic space on-the-fly from the conversational corpus under analysis. Building a semantic space derived from the conversational corpus has the potential of being more sensitive to intrinsic word meaning differences, but it also necessitates a large number of utterances and unique words. Because of the computational demands needed to separate word meanings in a large vector space, standard corpora for building word2vec spaces contain at least 30,000 unique words (e.g., TASA corpus contains 37,000 unique words, text8 corpus contains over 600,000 unique words) across thousands of individual documents or conversational turns. These minimal requirements may be significantly larger than the corpora available to most experimental researchers in this domain. To easily accom-

[6] Although these two are provided in the standard ALIGN flow, researchers could easily implement alternative POS taggers as desired. For example, NLTK's "maxent_treebank tagger" is available as an alternative option to the default averaged perceptron tagger. However, an advantage of Stanford's log-linear tagger is that non-English language models are also available for POS tagging, a feature that ALIGN was designed to easily integrate.

Table 2
Example of ALIGN Data File Output from Phase 1

Participant	Content	Token	Lemma	Tagged Stan token	Tagged Stan lemma
1	So do you think marijuana should be . . .	So, do, you, think, marijuana, should, be,	So, do, you, think, marijuana, should, be,	(so, RB), (do, VBP), (you, PRP), (think, VB), (marijuana, NN), (should, MD), (be, VB),	[(so, RB), (do, VBP), (you, PRP), (think, VB), (marijuana, NN), (should, MD), (be, VB),
2	I think it should	I, think, it, should	I, think, it, should	(i, FW), (think, VB), (it, PRP), (should, MD)	(i, FW), (think, VB), (it, PRP), (should, MD)
1	Why do you think it should be legalized	Why, do, you, think, it, should, be, legalized	Why, do, you, think, it, should, be, legalize	(why, WRB), (do, VBP), (you, PRP), (think, VB), (it, PRP), (should, MD), (be, VB), (legalized, VBN)	(why, WRB), (do, VBP), (you, PRP), (think, VB), (it, PRP), (should, MD), (be, VB), (legalize, VB)
2	I think it should be legalized because a lot of . . .	I, think, it, should, be, legalized, because, a, lot, of,	I, think, it, should, be, legalize, because, a, lot, of,	(i, LS), (think, VB), (it, PRP), (should, MD), (be, VB), (legalized, VBN), (because, IN), (a, DT), (lot, NN), (of, IN),	(i, LS), (think, VB), (it, PRP), (should, MD), (be, VB), (legalize, VB), (because, IN), (a, DT), (lot, NN), (of, IN),
1	Personally I think it should not be legalized . . .	Personally, I, think, it, should, not, be, legalized,	Personally, I, think, it, should, not, be, legalize,	(personally, RB), (i, FW), (think, VB), (it, PRP), (should, MD), (not, RB), (be, VB), (legalized, VBN),	(personally, RB), (i, FW), (think, VB), (it, PRP), (should, MD), (not, RB), (be, VB), (legalize, VB),
2	No absolutely not honestly that no I feel like . . .	[No, absolutely, not, honestly, that, no, I, feel, like,	[No, absolutely, not, honestly, that, no, I, feel, like,	(no, RB), (absolutely, RB), (not, RB), (honestly, RB), (that, IN), (no, DT), (i, FW), (feel, VB), (like, IN),	(no, RB), (absolutely, RB), (not, RB), (honestly, RB), (that, IN), (no, DT), (i, FW), (feel, VB), (like, IN),

Note. In addition to the original data from Phase 0 (see Table 1), Phase 1 now appends to the data file new columns that include the tokenized, lemmatized, and POS-tagged linguistic data. Each POS tagger yields a separate column. In the default case, this will yield one column for the averaged perceptron tagger and one for the Stanford Log-Linear Tagger (below: "Tagged Stan token," "Tagged Stan lemma"). Note the columns for the averaged perceptron tagger are not shown because of space limitations.

modate researchers whose corpora may not meet the size requirement, ALIGN defaults to the use of a semantic space based on the freely available Google News corpus,[7] which includes over 3 billion words.

Next, on a turn-by-turn basis, ALIGN transforms each word from an utterance into a high-dimensional semantic vector (HDSV). Using simple additive composition, the HDSVs of each word in a sentence are combined to generate a new "utterance-level" projection in the semantic space. Utterances are taken as conceptually related if they occur in similar regions of this space (Foltz, Kintsch, & Landauer, 1998; Landauer & Dumais, 1997; Mikolov, Sutskever, Chen, Corrado, & Dean, 2013). Like the lexical and syntactic forms, we can then compare the conceptual content of utterances using cosine similarity; importantly, however, the range of semantic similarity values will differ from lexical and syntactic similarity scores due to the nature of this particular analysis. Specifically, semantic scores can range from −1 (i.e., completely opposite conceptual content) to +1 (i.e., completely similar content).

Table 3
Example n-Gram Counters for Lexical and Syntactic Structures Generated by ALIGN for a Turn Pair Between Interlocutors

	Lexical
P1	Counter({u'do you': 1, u'be legalized': 1, u'so do': 1, u'legalized or': 1, u'think marijuana': 1, u'marijuana should': 1, u'you think': 1, u'or not': 1, u'should be': 1})
P2	Counter({u'i think': 1, u'it should': 1, u'think it': 1})
	Syntactic
P1	Counter({(u'MD', u'VB'): 1, (u'RB', u'VB'): 1, (u'VB', u'NNS'): 1, (u'CC', u'RB'): 1, (u'VB', u'PRP'): 1, (u'PRP', u'VB'): 1, (u'VB', u'VBN'): 1, (u'NNS', u'MD'): 1, (u'VBN', u'CC'): 1})
P2	Counter({(u'MD', u'VB'): 1, (u'RB', u'VB'): 1, (u'VB', u'NNS'): 1, (u'CC', u'RB'): 1, (u'VB', u'PRP'): 1, (u'PRP', u'VB'): 1, (u'VB', u'VBN'): 1, (u'NNS', u'MD'): 1, (u'VBN', u'CC'): 1})

To build a high-dimensional semantic space and to derive word and utterance vectors, we used Gensim's Python implementation of Google's word2vec algorithm (Version 3.1.0; Řehůřek & Sojka, 2010). The original algorithm is based on a neural network architecture that attempts to converge on optimal parameters to maximize the probability that a word comes from a particular context within a given corpus. Given the highly technical nature of the learning model (which uses what is known as a "skip-gram approach with negative sampling"), we refer interested readers to seminal work on the topic: Mikolov, Chen, Corrado, and Dean (2013) and Mikolov et al. (2013). However, knowledge of these technical aspects is not necessary to use ALIGN. Because ALIGN is built using Gensim—an easy-to-use Python package that is well-maintained, easily deployable, and intuitive—researchers can apply it to their data without extensive backgrounds in natural language processing (NLP) or information retrieval.

Essentially, for our case, word2vec takes a corpus of conversations and examines how each word is distributed across a sufficiently large number of relevant turns (because we are interested in comparing similar language across conversational turns). Its goal is to maximize the probability that (a) any two words that occur in

[7] Available for download: https://code.google.com/archive/p/word2vec/.

Table 4
Example of ALIGN Data File Output From Phase 2

Order	Direction	Token Bi-	Token Tri-	Lemma Bi-	Lemma Tri-	Conceptual	Stan token Bi-	Stan token Tri-	Stan lemma Bi-	Stan lemma Tri-
0	1 > 2	.000	.000	.000	.000	.390	.192	.000	.192	.000
1	2 > 1	.436	.289	.436	.289	.547	.000	.000	.000	.000
2	1 > 2	.267	.173	.267	.173	.631	.000	.000	.000	.000
3	2 > 1	.159	.062	.156	.062	.877	.500	.159	.537	.235
4	1 > 2	.089	.000	.105	.010	.879	.497	.142	.583	.196
5	2 > 1	.000	.000	.023	.000	.774	.259	.024	.341	.118
6	1 > 2	.000	.000	.024	.000	.806	.277	.025	.254	.050

Note. In addition to the data from Phase 0 and Phase 1 (see Table 2), Phase 2 now appends to the data file new columns that include markers for each turn-by-turn comparison: turn order ("Order" column), directionality of comparison between interlocutors' utterances ("Direction" column), and multiple cosine similarity measures for lexical, conceptual, and syntactic alignment at each of the user-specified *n*-gram sizes. Note the columns for the averaged perceptron tagger are not shown because of space limitations.

the same kinds of linguistic contexts are close together in the high-dimensional space (HDS); and (b) any two words that rarely occur in the same kinds of linguistic contexts are pushed further apart. In building this space, ALIGN provides users options to remove or retain high- and/or low-frequency words and, if retained, to set cutoff thresholds for what constitutes high or low frequency.

Specifying cutoffs for high- and low-frequency words is common practice in creating semantic space models because these types of words add significant noise and instability to the model. High-frequency words (e.g., function words, common verbs) occur in so many different turns and types of linguistic contexts that they are not well-differentiated within the semantic space, instead tending to capture syntactic information; low-frequency words (e.g., proper nouns) do not occur enough times in enough unique turns to have a firm place in the semantic space. In other words, extremely low-frequency words occur too rarely to be reliably informative, and extremely high-frequency words occur too often to be uniquely informative (Mikolov et al., 2013; Rohde, Gonnerman, & Plaut, 2006). By default, ALIGN removes all words that occur three standard deviations over the mean and any word that appears only once.

When the semantic space is built and each word given a vector representation within that space, words within an utterance are combined. Cosine similarity is then computed across utterances, turn-by-turn over the conversation. Like lexical and syntactic alignment, ALIGN preserves information on turn order and provides scores for directionality of alignment between participants for conceptual alignment.

As we noted above, building a high-dimensional semantic space for word2vec requires a sufficiently large number of unique lexical items in a sufficiently large number of (in our case) unique individual turns. As a result, researchers should be aware of the size of their corpus before proceeding. If the experimental data collected are relatively small or sparse, researchers may instead create a semantic space out of other text (e.g., Wikipedia articles, TV transcripts) or use an existing word2vec space (e.g., the Google News corpus). However, researchers should be careful in selecting existing spaces or creating new ones from other sources: The conceptual representations that emerge will be dependent upon the structure of the dataset used to create it. For example, the word "medicine" would likely have a different vector in a space created by advanced medical textbooks than by parent–child utterances, and the word "media" would look quite different in a corpus of artists' conversations than in a corpus of journalists' conversations.

Directionality. Lastly, the turn-by-turn analysis also provides information about the directionality of alignment. Specifically, ALIGN generates separate scores for how Participant 2 responds to Participant 1 (in Table 4, "1>2" value in "Direction" column) and how Participant 1 responds to Participant 2 (in Table 4, "2>1" value in "Direction" column). In this way, researchers can examine potential differences in how conversational partners linguistically follow one another across time and across linguistic levels.

Creating a Baseline Comparison for Turn-by-Turn Alignment

One of the critical considerations in determining whether observed alignment is greater than what might be expected by chance is to compare it with a relevant baseline. One approach is to randomly recombine the conversational turns within a dyad and then generate alignment values from the reordered turns ("shuffled baseline"). Doing so maintains distributional properties of how words are used in entire utterances but disrupts the temporal sequencing of the utterances (Louwerse, Dale, Bard, & Jeuniaux, 2012). However, this approach can be problematic, as it does not account for structural constraints in tasks where there is a natural event sequence due to the temporal ordering (e.g., how language is used at the beginning of a conversation will differ from the middle and end).

To account for these structural constraints, another approach is to randomly pair unaltered transcripts from members of different dyads and then generate alignment values from the fabricated dyad ("surrogates" or "surrogate partners"). This ensures that the ordering in which each turn was uttered by the surrogates in their original conversations is preserved (cf. pseudointeractions in Bernieri, Reznick, & Rosenthal, 1988; also see Richardson & Dale, 2005) and can capture the temporal or other constraints of the experiment on interaction (see Louwerse et al., 2012). Moreover, the surrogate method also has the advantage in that it preserves general distributions of word frequencies within a task condition and allows for an aggregated alignment score between a single speaker and multiple surrogate partners, thereby minimizing

any statistical aberrations of high or low alignment scores that might occur in single pairings (or single shufflings).

To generate a surrogate baseline, a pseudo-turn-by-turn exchange is created, whereby the original turn order is rank-ordered (e.g., Turn 1, Turn 2, Turn 3) and then matched across partners so that, for example, surrogate Partner 1's "Turn 1" is followed by surrogate Partner 2's "Turn 2" and so on. This ensures that complete conversations from one participant are combined with complete conversations from another participant, with the number of turns determined by the shortest of the two original conversations. This procedure is repeated for all possible pairs within each condition, except for the "real" partners (i.e., those who truly interacted in the experiment). If preferred, a smaller subset can instead be generated, in which each partner is randomly paired with just one other individual with whom they did not interact. Moreover, if the user does not wish for turn order to be preserved, there is an option to randomly shuffle turn order in the reconstructed surrogate pairings; however, for the most conservative baseline, we recommend the default approach of preserving turn order.[8]

An Example Application of ALIGN: Measuring LA in Deception and Disagreement

We now provide a novel application of ALIGN's LA quantification in order to demonstrate how empirical conversational data can be analyzed and interpreted within the ALIGN framework. To do so, we use a corpus collected previously in Duran and Fusaroli (2017). Whereas earlier analyses of this corpus focused on participants' movement and speech rate patterns, the current article presents the first analysis of participants' lexical, syntactic, and conceptual linguistic behaviors.

One of the strengths of this target dataset is that it involves a conversational domain with complex social and informational goals. In this "Devil's Advocate" paradigm, one participant is surreptitiously instructed at the beginning of the experiment[9] to deceive their conversational partner as convincingly as possible while ostensibly disagreeing or agreeing about ideologically sensitive matters. Unlike previous studies that use confederates or constrain partners within the experimental context (e.g., following scripted questions, computer-mediated communication), both partners were allowed to speak face-to-face with minimal restrictions and were unfamiliar with the goals of the study beforehand. We briefly describe the corpus here simply to situate the example application of ALIGN; additional detail on the data collection setup can be found in the original article (Duran & Fusaroli, 2017).

Data Overview

Modifying a paradigm created by Paxton and Dale (2013), pairs of participants were brought into the experiment and—before meeting one another—individually completed a survey about topics that were controversial at the time[10] (e.g., abortion, legalization of marijuana, gay and lesbian marriage). The survey asked participants to write their opinion briefly in their own words and to indicate on a 1–4 Likert-style scale how strongly they held those opinions. Although participants were not told this at the outset of the experiment, their responses to these opinion questions would form the basis of their conversation prompts throughout the experiment.

Each dyad was asked to converse for eight minutes for each of two separate conversations, with each conversation prompt revealed immediately before the conversation began. Each conversation was designed to elicit either agreement or disagreement between the partners (as a randomly assigned between-dyads condition) about the topic at hand. Critically to the current paradigm, one of these conversations involved deception, and the other truth, with order counterbalanced across dyads (as a within-dyad condition).

Partners' true opinions were evaluated by covertly comparing their written explanations of their opinions for questions about which each participant rated feeling 3 = *somewhat strongly* or 4 = *very strongly* on the Likert-style scale. In only one of the two conversations, one randomly selected partner was designated the "Devil's Advocate" (DA) and was covertly instructed to espouse an opinion opposite to their true opinions for the duration of that conversation only. The DA was only informed of this assignment immediately before the target conversation and was told to take great care in not revealing their status as a DA to their partner (hereafter referred to as the "naïve" partner/participant). The DA was also given a few minutes upon being given this instruction to privately plan talking points before holding the conversation.

In truthful conversations, agreement was simply based on congruency of written opinions, and disagreement on the incongruency of opinions. For deception conversations, agreement was based on initially incongruent opinions, and disagreement on initially congruent opinions. That is, for deceptive agreement conversations, dyads were asked to discuss a topic for which the DA had written an opinion differing from that of the naïve participant; for deceptive disagreement conversations, dyads were asked to argue about a topic for which the DA had written a similar opinion to that of the naïve participant.

Overall, the design yielded 24 dyads who held agreement conversations with each other and 24 dyads who held disagreement conversations with each other, with each dyad contributing one truthful and one deceptive conversation. To demonstrate ALIGN, we analyze here only a subset of the entire corpus: the deceptive conversations during either agreement and disagreement.

ALIGN Parameters

We prepared the analysis for this experiment following the order presented in the section title A General Approach for Quantifying Multilevel Linguistic Alignment in Discourse. For Phase 0, each conversation was prepared and converted to the appropriate CSV datasheet format (see Table 1; also see the Appendix for extended example). The key decision that researchers must make at this

[8] It should be noted that there is no clear consensus as to which baseline is theoretically the most appropriate given a particular dataset. There have been recent attempts to outline and compare possibilities (e.g., Lancaster, Iatsenko, Pidde, Ticcinelli, & Stefanovska, 2018; Moulder, Boker, Ramseyer, & Tschacher, 2018), but the focus tends to be on continuous physiological and movement signals. Some early work in a linguistic domain (Louwerse et al., 2012) has briefly touched on these issues with categorical time series, but greater work is needed to better specify theoretical assumptions.

[9] In the style of a naïve confederate, an experimental construct in social psychology that asks naïve participants to behave in researcher-proscribed ways (cf. Ollendick & Schmidt, 1987).

[10] Data were collected between 2012 and 2013.

point is how to operationalize "speech turns" when creating the datasheet since each turn will have its own row.

Deciding on how to carve up dialogue into speech turns is neither trivial nor a unique problem when evaluating naturalistic interactions (Sacks, Schegloff, & Jefferson, 1974), and much depends on the nature of the dialogue and overarching research questions (Heritage, 1998). For the current study, we followed Duncan's (1972) simple criteria that a "speech turn" is an utterance that begins when one speaker takes up the conversational floor and ends when that speaker relinquishes the floor to his or her conversational partner. Such turns include cases where one speaker takes the floor before the other speaker has finished speaking (overlapping turns) but does not include backchannels, in which the partner who does not possess the floor speaks for social or metaconversational purposes (e.g., "mm-hmm," "yeah") while the partner who does hold the conversational floor is still speaking.

It should be noted that extensions beyond this operationalization of turn-taking are possible given ALIGN's flexible method for handling data input. For example, had there been more extended backchannels in the current data, we could have broken up a single turn from a speaker into two rows at the point where the partner begun to speak, with an interleaving row containing the partner's backchannel utterance. Alternatively, if we had reasons to believe that greater alignment would be seen in the DA when examining only what the naïve participant had said immediately prior to relinquishing the conversational floor, we could have truncated the naïve participant's turns accordingly.

Next, for Phase 1, ALIGN preprocessed each utterance. Given we had no a priori reasons to change the default settings, the preprocessing created a standardized dialogue in which each utterance included at least two words, all non-ASCII characters and speech fillers were removed, and all words were automatically spell-corrected. This stage also ensured that rows in the conversational datasheet alternated between speakers (as motivated above).

For Phase 2, we generated turn-by-turn alignment scores across all conversations through the process described in the previous section. For the current study, we chose to use ALIGN's default parameters: Lexical alignment was measured from lemmatized word sequences, and syntactic alignment was measured using POS-tagged word forms without duplicate lemmatized sequences across utterance comparisons. POS tags were generated using the Stanford tagger.[11] Again, while these default settings aim to provide a conservative baseline, they can be easily adjusted if needed to accommodate other research goals or assumptions.

Surrogate partners were created by pairing all possible pairs, excluding the actual partners (e.g., starting with Participant 1 and pairing with all others, then Participant 2 and pairing with all others, etc.). To create condition-sensitive baselines, all pairings came from the same condition (e.g., conversations marked by agreement and deception). Surrogate alignment scores were generated using the same parameter settings as used when generating alignment scores on actual partners. It should also be noted that because each participant now had multiple simulated conversations, to generate a composite similarity score for each turn, we averaged across all simulated conversations for each participant.

Open Science and Replicability

An accessible step-by-step tutorial of how to replicate this procedure can be found on ALIGN's GitHub repository (https://github.com/nickduran/align-linguistic-alignment). The tutorial contains all the necessary information for accessing the transcript data and codebook, generating ALIGN scores, and replicating statistical models as reported in the following section.

The data and codebook are hosted on a protected-access repository on the Interuniversity Consortium for Political and Social Research (http://dx.doi.org/10.3886/ICPSR37124.v1; Duran, Paxton, & Fusaroli, 2018), allowing any verified researcher on the Interuniversity Consortium for Political and Social Research network to freely access the de-identified data for this study. Because of the sensitivity of participants' self-disclosures in this study, this data-sharing model allows us to balance the needs of our participants with the importance of open science.

Statistical Analysis and Model Specifications

Given the great scarcity of prior research on linguistic alignment in conversations involving deception or conflict, our analyses are ultimately exploratory in nature. These analyses serve as an opportunity to highlight the theory-driven applications of ALIGN by assessing multilevel linguistic alignment in a novel experimental context. For current purposes, we focus on whether alignment is expressed (relative to a surrogate baseline) by either the DA or naïve participant to their partner. In doing so, we are able to explore the impact that individuals' social goals (i.e., truth-telling vs. motivated deception) and joint conversation goals (i.e., agreement vs. disagreement) have on the emergence of linguistic alignment across lexical, syntactic, and conceptual levels.

We generated two sets of linear mixed-effects models, with one set evaluating the standardized syntactic, lexical, and conceptual alignment scores of the DA (Participant 1) responding to a naïve partner (Participant 2; i.e., DA-following), and a second set of models evaluating the standardized syntactic, lexical, and conceptual alignment scores of the naïve partner (Participant 2) responding to the DA partner (Participant 1; i.e., naïve-following). In each set, we examined the fixed factor of data type: whether scores are based on real partners versus surrogate partners (contrast codes: Real = 0.5, Surrogate = −0.5), with particular focus on interactions between conflict (Agreement = 0.5, Disagreement = −0.5) and turn order (changes over time in alignment; mean-centered).

Moreover, because ALIGN also generates syntactic and lexical alignment scores based on *n*-grams of varying length, we also repeated the above for *n*-grams of Size 2 (bigrams) and Size 3 (trigrams). This analysis provides insight into how alignment persists over increasing alignment structure sizes (i.e., greater phrase specificity from bi- to tri-grams). It is important to again note that directionality of alignment (Participant 1 to Participant 2; Participant 2 to Participant 1), data type, turn order, and *n*-gram sizes are automatically generated by ALIGN and can be easily integrated into model specifications.

For additional model settings relevant to the current demonstration, we also include a random intercept for dyad identity with a

[11] The same pattern of results was also found with the NLTK's default averaged perceptron tagger.

random slope for turn order. Using Barr, Levy, Scheepers, and Tily (2013) standard notation, we report here the single-equation notation for the structure of each of our analytical models:

$$cos_{dl} = \beta_0 + D_{0df} + \beta_1 t_{df} + \beta_2 c_{df} + \beta_3 o_{df} + \beta_4 t_{df} c_{df} + \beta_5 t_{df} o_{df} + D_{1df} o_{df} + e_{df} \quad (1)$$

Equation (1) estimates the cosine similarity *cos* of a given linguistic level (and, when appropriate, *n*-gram size) for dyad *d* with follower *f*. To do so, it estimates the fixed (global) coefficients *ß* for each of our fixed effects: the main effect of data type *t* (i.e., real or baseline data) and its interactions with conflict *c* (i.e., agreement or disagreement) and turn order *o* (i.e., standardized number). The model also includes the lower-level main effects for conflict and turn order, error *e*, and a random intercept for dyad identity *D* with a time-sensitive random slope structure *o*. Written in R-typical pseudovariable notation, our model outlined in Equation (1) would be implemented in lmer code syntax as provided in Equation (2):

$$\begin{aligned} \text{dependent.variable.alignment} \sim\ & \text{data.type} + \text{conflict} + \text{turn.order} \\ & + \text{data.type} : \text{conflict} \\ & + \text{data.type} : \text{turn.order} \\ & + (1 + \text{turn.order} | \text{dyad}) \end{aligned}$$

with error *e* implicitly included by the statistical package. In our example, this model structure is used for bigram lexical, trigram lexical, bigram syntactic, trigram syntactic, and conceptual alignment scores separately for DA—and naïve—following alignment (as we report below).

To ensure that each overall model was statistically significant compared with a null model with only random effects, we also report the results of a likelihood ratio test between the two model types. All analyses were run using the lme4 package (v. 1.1–17; Bates, Maechler, Bolker, & Walker, 2015) in the R programming environment (R Core Team, 2018). For all full models, we report coefficients of the standardized predictors, their standard errors, and derived *p* values for each of the predictors. Overall model fit (R^2) is computed as variance explained by fixed and random factors together, using a version of Nakagawa and Schielzeth's (2013) "conditional R^2" method, extended by Johnson (2014), and that is implemented in the MuMIn R statistical package (v. 1.40.4; Barton, 2018). This overall fit statistic accounts for variances at multiple levels (e.g., for each random factor and the residual variance) that is a hallmark of mixed-effects modeling.[12]

Results

We first report results corresponding to patterns of DA-following alignment and then patterns of naïve-following alignment. Tables 5 and 7 provide descriptive data and Tables 6 and 8 show the main effects of data type (real vs. surrogate) and interactions with conflict (agreement vs. disagreement) and turn order (time). It is important to note that given data type is a constitutive variable within an interaction term, the conditional main effect of data type should be interpreted as its effect (on linguistic alignment) when the other constitutive variables, conflict, and turn order are equal to zero (Brambor, Clark, & Golder, 2006).

For simplicity and clarity of text, we do not reiterate the parameter estimates reported in the tables here. While the tables include all effects included in each model, any effect not mentioned in the text did not reach statistical significance in that model. Additionally, while we do not report exact *p* values in the tables, the complete outputs of all models are available on ALIGN's GitHub repository.

DA-Following Results (See Table 6)

Lexical alignment. The overall bigram model fit with fixed and random effects was statistically significantly different from the model fit with random effects only, $\chi^2(3) = 92.248$, $p < 0.001$, $R^2 = 0.118$. There was a main effect of data type, such that real conversations had significantly higher lexical alignment than would be expected by chance. Although the interaction between data type and time (turn order; see Figure 1) did not reach statistical significance at the chosen alpha ($p < .05$), an important next step would be to determine whether there was a more pronounced rate of decline for real partners compared with surrogate ones.

The overall trigram model fit with fixed and random effects was statistically significantly different from the model fit with random effects only, $\chi^2(3) = 39.768$, $p < .001$, $R^2 = 0.048$. We again found a main effect of data type, such that greater alignment persisted in the real conversations compared with surrogate conversations.

Conceptual alignment. The overall conceptual alignment model fit with fixed and random effects was statistically significantly different from the model fit with random effects only, $\chi^2(3) = 60.197$, $p < .001$, $R^2 = 0.158$, with a main effect for data type alone. Again, the model fit comparisons found greater alignment in real conversations compared with surrogate conversations.

Syntactic alignment. We found no statistically significant effects for bigrams for syntactic alignment. However, the overall trigram model fit—which was statistically significantly different from the model fit with random effects only, $\chi^2(3) = 14.254$, $p = .003$, $R^2 = 0.090$—identified a main effect for data type, such that greater alignment occurred in the real conversations compared with surrogate conversations.

Naïve-Following Results (See Table 8)

Lexical alignment. A statistically significant difference between the overall model fit and the null model fit was found for bigrams, $\chi^2(3) = 63.020$, $p < .001$, $R^2 = 0.072$, and trigrams, $\chi^2(3) = 26.156$, $p < .001$, $R^2 = 0.023$. In both the overall bigram and trigram models, a main effect was found for data type, indicating greater lexical alignment in real conversations compared with surrogate conversations.

Conceptual alignment. The overall conceptual alignment model fit with fixed and random effects was statistically significantly different from the model fit with random effects only, $\chi^2(3) = 37.614$, $p < .01$, $R^2 = 0.163$, with a statistically signifi-

[12] Generating an appropriate effect size statistic for hierarchical linear models has unique challenges given the potential for unexplained variance at each level of the hierarchy (Jaeger, Edwards, Das, & Sen, 2017). Here, we have chosen a widely used method to compute R-squared that provides a single approximation; however, for a more complete picture, it will be necessary to examine multiple R-squared measures in an integrative framework that allows for a more comprehensive decomposition of variance (see Rights & Sterba, 2018 for a leading account).

Table 5
Observed Mean and Standard Error of Real and Surrogate "DA-Following" Alignment Scores for Lexical and Syntactic Alignment (Bi- and Trigrams), as Well as Conceptual Alignment, Separated by Agreement and Disagreement Conflict Conditions

	DA-Following				
	Lexical		Conceptual	Syntactic	
Variables	Bi-	Tri-	—	Bi-	Tri-
Agree					
Real	.082 (.006)	.020 (.002)	.688 (.007)	.280 (.007)	.086 (.004)
Surrogate	.045 (.002)	.006 (.0005)	.636 (.005)	.274 (.005)	.077 (.002)
Disagree					
Real	.077 (.006)	.019 (.002)	.673 (.007)	.267 (.007)	.083 (.004)
Surrogate	.037 (.002)	.005 (.0004)	.611 (.005)	.260 (.005)	.071 (.002)

cant effect for data type: Real conversations showed greater conceptual alignment compared with surrogate conversations.

Syntactic alignment. No statistically significant effects for syntactic alignment were found.

Brief Discussion of the Example Application

To demonstrate ALIGN's use in naturalistic corpora, we applied our method to a corpus involving deceptive conversations with goals of agreement and disagreement and analyzed lexical, semantic and syntactic alignment. Generally, we found that lexical, semantic, and syntactic alignment in conversations tended to occur more than we would expect simply by chance (see Figure 2). However, not all forms of alignment were equally present: Both DA and naïve participants exhibited lexical and semantic alignment (i.e., likely to follow), broadly supporting previous findings on lexical alignment in communication (e.g., Brennan & Clark, 1996). Conversely, while DA participants reliably showed syntactic alignment at the trigram level, we did not find that naïve participants showed more syntactic alignment than would occur by chance, nor did we find any significant changes in any type of alignment over time by either participant. Interestingly—and contrasting with measures of nonverbal alignment in this corpus (Duran & Fusaroli, 2017)—we observed no effects of disagreement and agreement.

Although these results are inarguably exploratory, they can provide us with preliminary hints for important theoretical questions about alignment and highlighting follow-up questions. Perhaps most strikingly, our results suggest that the emergence of alignment is sensitive both to linguistic form and to contextual pressures, supporting a view of interaction as a complex adaptive system (e.g., Fusaroli et al., 2014; Paxton et al., 2016; Riley et al., 2011). Future work should continue to explore interpersonal communication in complex settings in order to better understand the context-sensitivity of human interaction.

The current results may also add nuance to critical issues as they relate to multiple pressures on interaction and how they are expressed in form-specific channels. For example, deception requires a strategic intent to appear natural, yet also involves increased cognitive load to overcome a true representation of opinion. Such pressures may have differential effects on the DA participant's behaviors: The strategic intent unintentionally may lead to a decrease in lexical alignment as participants attempt to appear more "truth-like," while other linguistic modes—like syntactic alignment—remain high throughout a conversation given their greater sensitivity to cognitive load and (less so) strategic control. Further work should continue to explore the role of pressures like communicative intent and cognitive load on naturalistic, fully interactive contexts to better understand their varied effects on unique linguistic levels.

Our results also add nuance to continuing conversations about the function and emergence of alignment across behavior types. Contrary to previous work on body movement during argument (Paxton & Dale, 2013, 2017), we found no effect of conversation goals (i.e., agreement vs. disagreement) on lexical alignment. More interestingly, the current findings on linguistic alignment did not show the same complex interactions (i.e., between goals and

Table 6
DA-Following: Results From Mixed Effects Models With Factors Data Type (Real vs. Surrogate), Conflict (Agreement vs. Disagreement), and Time (Continuous Turn-Ranked)

	DA-Following				
	Lexical		Conceptual	Syntactic	
Model terms	Bi-	Tri-	—	Bi-	Tri-
Type	$.536^{***}$ (.056)	$.363^{***}$ (.057)	$.420^{***}$ (.056)	.079 (.056)	$.199^{***}$ (.057)
Type × Conflict	−.054 (.111)	−.055 (.115)	−.193 (.112)	−.020 (.111)	−.097 (.113)
Type × Time	$-.010^{O}$ (.005)	−.005 (.006)	−.008 (.006)	−.006 (.005)	−.006 (.005)

Note. We report the β with associated *p*-value, and standard error of the coefficient.
^{O} $p < .07$. *** $p < .001$.

Table 7
Observed Mean and Standard Deviation of Real and Surrogate "Naïve-Following" Alignment Scores for Lexical and Syntactic Alignment (Bi- and Trigrams), as Well as Conceptual Alignment, Separated by Agreement and Disagreement Conflict Conditions

	Naïve-Following				
	Lexical		Conceptual	Syntactic	
Variables	Bi-	Tri-	—	Bi-	Tri-
Agree					
Real	.080 (.006)	.022 (.003)	.686 (.007)	.275 (.007)	.083 (.004)
Surrogate	.042 (.002)	.006 (.0005)	.637 (.005)	.273 (.005)	.077 (.002)
Disagree					
Real	.070 (.006)	.018 (.002)	.669 (.007)	.267 (.007)	.078 (.003)
Surrogate	.037 (.002)	.005 (.0003)	.617 (.005)	.264 (.005)	.073 (.002)

deception) as the interlocutors' body movement patterns within the same corpus (Duran & Fusaroli, 2017). Similarly, even within linguistic behaviors, we found different effects across lexical, syntactic, and conceptual levels. Taken together, these suggest that different behaviors—and specific levels even within those behaviors—may have unique alignment dynamics, perhaps even with their own outcomes and associated markers. By studying different behaviors within a single interaction context, future work should investigate the signatures and effects of alignment within and across behaviors (also see Oben & Brône, 2016; Paxton et al., 2016).

Finally, recent work has begun to acknowledge that the simple *presence* of alignment is not always beneficial to a given interaction or joint goal (Fusaroli et al., 2012; Main et al., 2016). While we have here explored the presence of alignment under various individual- and contextual-level pressures, the nature of the corpus makes it difficult to establish a metric for success. Although our exploratory results can shed new light on multilevel alignment over time, future work should investigate the degree to which the kinds of patterns observed in the present study contribute to success metrics.

General Discussion

The presence of turn-by-turn linguistic alignment (LA) in extended dialogue promises to shed new light on the interpersonal dependencies that shape cognition and language use. This places particular importance on providing flexible quantitative metrics for LA that can be extended to different types of conversational texts and that can be easily used by researchers of diverse backgrounds. To that end, we here have introduced one such set of straightforward procedures: ALIGN, an open-source tool that simultaneously evaluates syntactic, lexical, and conceptual alignment using a combination of established and cutting-edge natural language processing techniques.

These procedures allow for a "first-pass" understanding of critical issues surrounding LA, including the presence and relative strength of turn-by-turn alignment across linguistic levels, phrases of increasing specificity (through *n*-grams), individual differences, communication contexts, and time. Moreover, ALIGN provides researchers with multiple options for creating surrogate baselines and removing lexical influences on syntactic alignment to accommodate a range of experimental, theoretical, and analytical perspectives. ALIGN is implemented in Python code to provide researchers with an all-in-one tool for evaluating multilevel LA without labor-intensive hand-coding or extensive experience in computational linguistics. ALIGN can thus be readily and reproducibly applied to experimentally created corpora (e.g., our empirical application) as well as naturally occurring and "big data" resources (e.g., online community boards, social media).

By opening up these sorts of analyses to even bigger scales and broader contexts, ALIGN addresses a need in LA research for a systematic comparison of alignment across studies. Such cross-study comparison is needed to create a better understanding of the common social and communicative demands that contribute to patterns of alignment. Ultimately, this should help provide new inroads into key debates surrounding the purpose and adaptiveness

Table 8
Naïve-Following: Results from Mixed Effects Models With Factors Data Type (Real vs. Surrogate), Conflict (Agreement vs. Disagreement), and Time (Continuous Turn-Ranked)

	Naïve-Following				
	Lexical		Conceptual	Syntactic	
Model terms	Bi-	Tri-	—	Bi-	Tri-
Type	.467*** (.058)	.303*** (.060)	.352*** (.057)	.036 (.056)	.095 (.057)
Type*Conflict	.061 (.116)	−.081 (.119)	−.058 (.115)	.043 (.113)	.070 (.113)
Type*Time	−.004 (.006)	−.001 (.006)	.003 (.006)	−.003 (.006)	−.004 (.006)

Note. We report the β with associated *p*-value, and standard error of the coefficient.
*** $p < .001$.

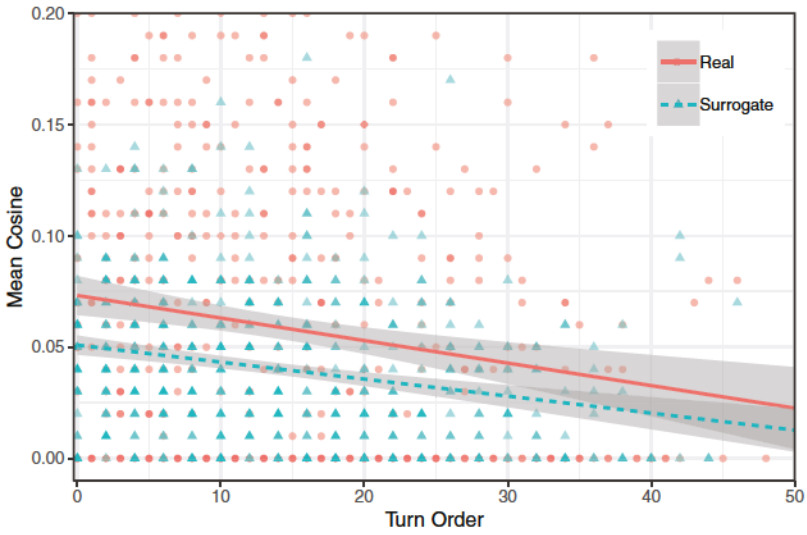

Figure 1. Lexical (lemma) bigram alignment decreases over time for DA-following directionality. Linear fits for each dyad type (Real = red [solid line; circles]; Surrogate = blue [dashed line; triangles]) included over all data points. See the online article for the color version of this figure.

of alignment. To that end, ALIGN can clear some of the obstacles by introducing intuitive and standardized measures that can be extended to naturalistic experimental contexts and naturally occurring data. As with other areas of cognitive science (Goldstone & Lupyan, 2016; Griffiths, 2015; Paxton & Griffiths, 2017), the growing availability of naturally occurring and big data sets is poised to revolutionize our theoretical exploration of interpersonal alignment in larger and more naturalistic contexts, and we hope that ALIGN will improve theory-driven exploration of these new resources with its low barrier to entry, its open-source development, and its accessible implementation of current best practices.

Areas for Future ALIGN Development

Although ALIGN allows multiple levels of LA to be assessed, the depth and complexities of alignment requires the development of additional techniques that can provide more varied and more detailed assessments. For example, with questions surrounding prediction error in structural priming (Fine, Jaeger, Farmer, & Qian, 2013; Jaeger & Snider, 2013) or for more explicit accounts of how people align particular aspects of language (Ellis, 2005; Mitchell, Myles, & Marsden, 2013), it becomes necessary to examine alignment at the level of isolated syntactic and lexical forms. As is often the case, this requires the use of annotated corpora or specialized computational parsers that are not currently implemented within ALIGN. To address this, we are currently creating solutions that might allow for more direct comparison. These include recording the frequency with which specific *n*-grams appear and the rate at which they are aligned across turns; another would be to go beyond *n*-grams, allowing users to target specific phrases or syntactic structures with customized regular expressions (e.g., POS structures for alternations between active vs. passive; prepositional vs. double-object datives).

Another assessment for in-depth LA is to look at the persistence of alignment between conversational turns at increasing temporal lags (i.e., "decay rate;" e.g., Healey et al., 2014; Reitter et al., 2011). This has had particular importance in understanding the role of working memory in alignment and how it might contribute to facilitating alignment across linguistic levels (Reitter et al., 2011). Whether the turn-to-turn alignment results reported would hold across longer time scales is an open question, and intersects with a growing interest in understanding multiscale alignment in other behavioral domains (Abney et al., 2014; Xu & Reitter, 2017). We believe such analysis could be easily integrated with ALIGN procedures. Indeed, in the current version of ALIGN, we have made an option available to generate alignment scores between immediately contiguous conversational turns (as is currently done) and turns systematically removed in time. However, a detailed description of this functionality is outside the scope of the current article.

Follow-up research should also compare the sensitivity and discriminability of cosine similarity measures (such as the one used in the ALIGN) with other, more probability-based measures, including the hierarchical alignment model (HAM; Doyle, Yurovsky, & Frank, 2016), RepDecay (Xu & Reitter, 2015), and Local LA (LLA; Fusaroli et al., 2012). In one recent attempt focused on probability-based measures (Xu & Reitter, 2015), variations in outcome do show relative advantages, but these advantages tend to depend on the theoretical questions being asked, such as whether the focus is on identifying individual differences or whether similar alignment values are expected across linguistic levels. Another exploration of this issue has taken a critical view of the features being isolated and whether baselines in word frequencies and an individual's propensity to use certain word classes can influence alignment scores (Doyle et al., 2016). From this initial work, it appears that such considerations do increase performance, most notably when applied to sparse data sets (e.g., Twitter conversations). Moving forward, it will be important to continue to explore these baseline considerations and to update ALIGN procedures to comport with evolving best-practices.

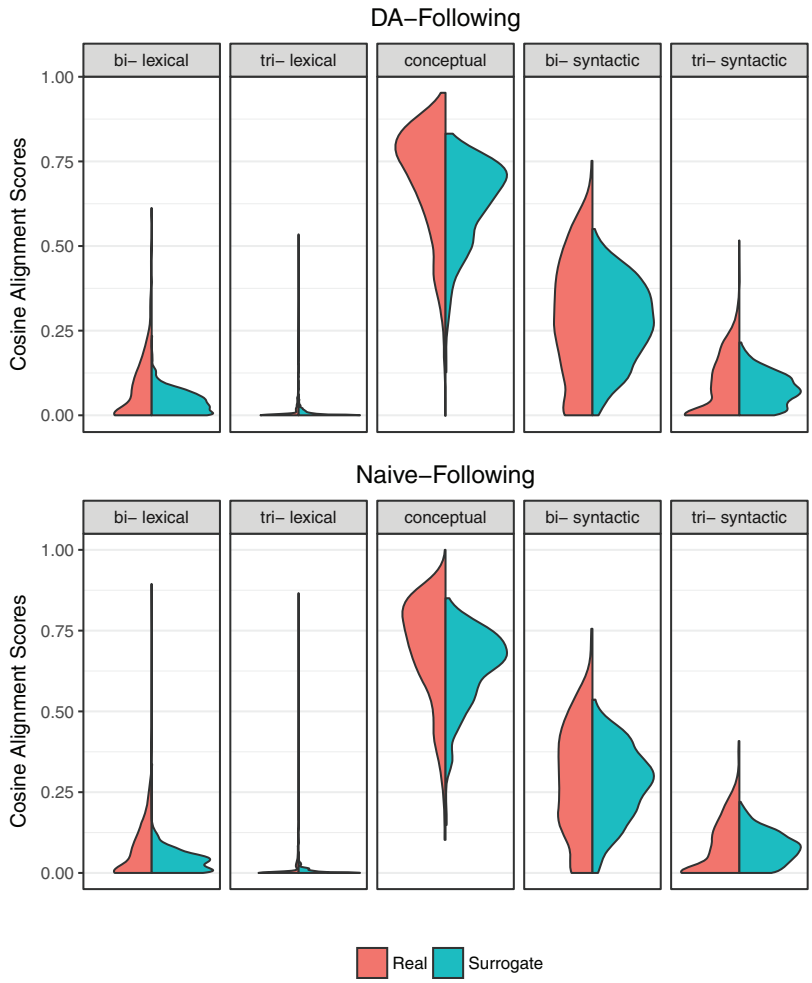

Figure 2. Split violin plots showing distributions of cosine scores of each dyad type (Real = red [distribution on the left]; Surrogate = blue [distributions on the right]) for lexical and syntactic alignment (bi- and trigrams), as well as conceptual alignment, for "DA-Following" directionality (top panel) and "Naïve-Following" directionality (bottom panel). Scores are collapsed across agreement and disagreement conflict conditions given no there were no statistically observed effects for conflict. See the online article for the color version of this figure.

ALIGN was developed with a deep commitment to the open science community. As such, we invite others to join us as we continue to develop the project. The source code for the tool is available on GitHub (https://github.com/nickduran/align-linguistic-alignment) with detailed documentation, and a stable version is available for distribution through PyPI (https://pypi.python.org/pypi/align). ALIGN is developed in Python, a programming language with an extensive history and community of individuals engaged in natural language processing applications, and we have built upon the Python packages ntlk and genism to implement a number of our core functions.

We explain these procedures in detail in the documentation hosted on ALIGN's GitHub repository. We have also made available the "Devil's Advocate" corpus reported in this article for users, not only to replicate the results but also to provide a concrete example of how to use ALIGN in theory-driven research. ALIGN additionally includes other analysis examples with accompanying corpora to further demonstrate how our methodological approach can be used to study language dynamics in a range of areas (e.g., child-caregiver interactions[13]). We are also committed to continuing to develop ALIGN itself, with future plans to convert code to Python 3 and the option to use additional NLP packages, such as OpenNLP. Moreover, by sharing our code on GitHub, we have provided a means for future community development, and we welcome interested researchers to work with us to improve ALIGN for the language community.

[13] We currently have published with the ALIGN package the "Kuczaj Corpus" by Stan Kuczaj (licensed under a Creative Commons Attribution-ShareAlike 3.0 Unported License), which consists of extended conversational interactions between a single child and his caregiver. Additional Jupyter notebook tutorials using two corpora from previous studies on alignment—CALLFRIEND (Canavan & Zipperlen, 1996) and the Map Task corpus (Anderson et al., 1991)—are also in development.

Conclusion

In the exciting new frontier of large-scale and naturally occurring data sets, we have developed ALIGN to provide an easily extensible tool for those studying patterns of multilevel LA. By combining well-established and cutting-edge natural language processing tools with current best-practices in an open-source distribution framework, ALIGN equips researchers with an easily accessible tool for wide-reaching quantitative exploration of general patterns of LA. It thus has the potential to open the door to pursuing and comparing reproducible analyses of alignment across experimental contexts and data sources, where both laboratory-based and real-world data sets can be weaved together for a richer understanding of LA.

References

Abney, D. H., Paxton, A., Dale, R., & Kello, C. T. (2014). Complexity matching in dyadic conversation. *Journal of Experimental Psychology: General, 143,* 2304–2315. http://dx.doi.org/10.1037/xge0000021

Allen, M. L., Haywood, S., Rajendran, G., & Branigan, H. (2011). Evidence for syntactic alignment in children with autism. *Developmental Science, 14,* 540–548. http://dx.doi.org/10.1111/j.1467-7687.2010.01001.x

Anderson, A., Bader, M., Bard, E., Boyle, E., Doherty, G. M., Garrod, S., . . . Weinert, R. (1991). The HCRC map task corpus. *Language and Speech, 34,* 351–366. http://dx.doi.org/10.1177/002383099103400404

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics, 40,* 177–189. http://dx.doi.org/10.1016/j.wocn.2011.09.001

Balcetis, E., & Dale, R. (2005). An exploration of social modulation of syntactic priming. In B. G. Bara, L. Barsalou, & M. Bucciarelli, (Eds.), *Proceedings of the 27th Annual Meeting of the Cognitive Science Society* (pp. 184–189). Austin, TX: Cognitive Science Society.

Bangerter, A., & Clark, H. H. (2003). Navigating joint projects with dialogue. *Cognitive Science, 27,* 195–225. http://dx.doi.org/10.1207/s15516709cog2702_3

Barton, K. (2012). *MuMIn: Model selection and model averaging based on information criteria* (R package version 1.40.4). Retrieved from https://cran.r-project.org/package=MuMIn

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68,* 255–278. http://dx.doi.org/10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67,* 1–48. http://dx.doi.org/10.18637/jss.v067.i01

Bengio, Y., Ducharme, R., Vincent, P., & Jauvin, C. (2003). A neural probabilistic language model. *Journal of Machine Learning Research, 3,* 1137–1155.

Bernieri, F., Reznick, J. S., & Rosenthal, R. R. (1988). Synchrony, pseudosynchrony, and sissynchrony: Measuring the entrainment process in mother-infant interactions. *Journal of Personality and Social Psychology, 54,* 243–253. http://dx.doi.org/10.1037/0022-3514.54.2.243

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with python*. Sebastopol, CA: O'Reilly Media.

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology, 18,* 355–387. http://dx.doi.org/10.1016/0010-0285(86)90004-6

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5,* 341–345.

Brambor, T., Clark, W. R., & Golder, M. (2006). Understanding interaction models: Improving empirical analyses. *Political Analysis, 14,* 63–82. http://dx.doi.org/10.1093/pan/mpi014

Branigan, H. P., & Pickering, M. J. (2017). Structural priming and the representation of language. *Behavioral and Brain Sciences, 40,* e313. http://dx.doi.org/10.1017/S0140525X17001212

Branigan, H. P., Pickering, M. J., McLean, J. F., & Cleland, A. A. (2007). Syntactic alignment and participant role in dialogue. *Cognition, 104,* 163–197. http://dx.doi.org/10.1016/j.cognition.2006.05.006

Branigan, H. P., Pickering, M. J., Pearson, J., & McLean, J. F. (2010). Linguistic alignment between people and computers. *Journal of Pragmatics, 42,* 2355–2368. http://dx.doi.org/10.1016/j.pragma.2009.12.012

Branigan, H. P., Pickering, M. J., Pearson, J., McLean, J. F., & Brown, A. (2011). The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. *Cognition, 121,* 41–57. http://dx.doi.org/10.1016/j.cognition.2011.05.011

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1482–1493. http://dx.doi.org/10.1037/0278-7393.22.6.1482

Brennan, S. E., Galati, A., & Kuhlen, A. K. (2010). Two minds, one dialog: Coordinating speaking and understanding. *Psychology of Learning and Motivation, 53,* 301–344. http://dx.doi.org/10.1016/S0079-7421(10)53008-1

Canavan, A., & Zipperlen, G. (1996). *CALLFRIEND American English-non-southern dialect*. Philadelphia, PA: Linguistic Data Consortium.

Carbary, K., & Tanenhaus, M. (2011, July). Conceptual pacts, syntactic priming, and referential form. In K. van Deemter, A. Gatt, R. van Gompel, & E. Krahmer (Chairs), *Production of referring expressions: Bridging the gap between computational, empirical and theoretical approaches to reference*. Workshop conducted at the 33rd Annual Meeting of the Cognitive Science Society, Boston, MA.

Chartrand, T. L., & Lakin, J. L. (2013). The antecedents and consequences of human behavioral mimicry. *Annual Review of Psychology, 64,* 285–308. http://dx.doi.org/10.1146/annurev-psych-113011-143754

Cleland, A. A., & Pickering, M. J. (2003). The use of lexical and syntactic information in language production: Evidence from the priming of noun-phrase structure. *Journal of Memory and Language, 49,* 214–230. http://dx.doi.org/10.1016/S0749-596X(03)00060-3

Coyle, J. M., & Kaschak, M. P. (2012). Female fertility affects men's linguistic choices. *PLoS ONE, 7,* e27971–e27977. http://dx.doi.org/10.1371/journal.pone.0027971

Dale, R., Fusaroli, R., Duran, N., & Richardson, D. C. (2013). The self-organization of human interaction. *Psychology of Learning and Motivation, 59,* 43–95. http://dx.doi.org/10.1016/B978-0-12-407187-2.00002-2

Dale, R., & Spivey, M. J. (2006). Unraveling the dyad: Using recurrence analysis to explore patterns of syntactic coordination between children and caregivers in conversation. *Language Learning, 56,* 391–430. http://dx.doi.org/10.1111/j.1467-9922.2006.00372.x

Doyle, G., Yurovsky, D., & Frank, M. C. (2016). *A robust framework for estimating linguistic alignment in twitter conversations*. New York, NY: ACM Press.

Duncan, S. (1972). Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology, 23,* 283–292. http://dx.doi.org/10.1037/h0033031

Duran, N. D., & Dale, R. (2014). Perspective-taking in dialogue as self-organization under social constraints. *New Ideas in Psychology, 32,* 131–146. http://dx.doi.org/10.1016/j.newideapsych.2013.03.004

Duran, N., Dale, R., & Galati, A. (2016). Toward integrative dynamic models for adaptive perspective taking. *Topics in Cognitive Science, 8,* 761–779. http://dx.doi.org/10.1111/tops.12219

Duran, N. D., & Fusaroli, R. (2017). Conversing with a devil's advocate: Interpersonal coordination in deception and disagreement. *PLoS ONE, 12,* e0178140. http://dx.doi.org/10.1371/journal.pone.0178140

Duran, N. D., Paxton, A., & Fusaroli, R. (2018). *Conversational transcripts of truthful and deceptive speech involving controversial topics,*

Central California, 2012. Ann Arbor, MI: Inter-university Consortium for Political and Social Research.

Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition, 27,* 305–352. http://dx.doi.org/10.1017/S027226310505014X

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS One, 8,* e77661. http://dx.doi.org/10.1371/journal.pone.0077661

Foltz, A., Gaspers, J., Meyer, C., Thiele, K., Cimiano, P., & Stenneken, P. (2015). Temporal effects of alignment in text-based, task-oriented discourse. *Discourse Processes, 52,* 609–641. http://dx.doi.org/10.1080/0163853X.2014.977696

Foltz, P. W., Kintsch, W., & Landauer, T. (1998). The measurement of textual coherence with latent semantic analysis. *Discourse Processes, 15,* 85–307. http://dx.doi.org/10.1080/01638539809545029

Friedberg, H., Litman, D., & Paletz, S. B. F. (2012). *Lexical entrainment and success in student engineering groups*. New York, NY: IEEE.

Fusaroli, R., Bahrami, B., Olsen, K., Roepstorff, A., Rees, G., Frith, C., & Tylén, K. (2012). Coming to terms: Quantifying the benefits of linguistic coordination. *Psychological Science, 23,* 931–939. http://dx.doi.org/10.1177/0956797612436816

Fusaroli, R., Rączaszek-Leonardi, J., & Tylén, K. (2014). Dialog as interpersonal synergy. *New Ideas in Psychology, 32,* 147–157. http://dx.doi.org/10.1016/j.newideapsych.2013.03.005

Fusaroli, R., Tylén, K., Garly, K., Steensig, J., Christiansen, M., & Dingemanse, M. (2017). Measures and mechanisms of common ground: Backchannels, conversational repair, and interactive alignment in free and task-oriented social interactions. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proceedings of the 39th Annual Meeting of the Cognitive Science Society* (pp. 2055–2060). Austin, TX: Cognitive Science Society.

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition, 27,* 181–218. http://dx.doi.org/10.1016/0010-0277(87)90018-7

Goldstone, R. L., & Lupyan, G. (2016). Discovering psychological principles by mining naturally occurring data sets. *Topics in Cognitive Science, 8,* 548–568. http://dx.doi.org/10.1111/tops.12212

Gries, S. T. (2005). Syntactic priming: A corpus-based approach. *Journal of Psycholinguistic Research, 34,* 365–399. http://dx.doi.org/10.1007/s10936-005-6139-3

Griffiths, T. L. (2015). Manifesto for a new (computational) cognitive revolution. *Cognition, 135,* 21–23. http://dx.doi.org/10.1016/j.cognition.2014.11.026

Healey, P. G. T., Purver, M., & Howes, C. (2014). Divergence in dialogue. *PLoS ONE, 9,* e98598–e6. http://dx.doi.org/10.1371/journal.pone.0098598

Heritage, J. (1998). Conversation analysis and talk: Analyzing distinctive turn-taking systems. In S. Cmejrkova, J. Hofmanova, O. Mullerova, & J. Svetla, (Eds.), *Dialogue analysis VI* (pp. 3–17). Tubingen, Germany: Niemeyer. http://dx.doi.org/10.1515/9783110965049-001

Hopkins, Z., Yuill, N., & Branigan, H. P. (2017). Inhibitory control and lexical alignment in children with an autism spectrum disorder. *Journal of Child Psychology and Psychiatry, 58,* 1155–1165. http://dx.doi.org/10.1111/jcpp.12792

Hopkins, Z., Yuill, N., & Keller, B. (2015). Children with autism align syntax in natural conversation. *Applied Psycholinguistics, 37,* 347–370. http://dx.doi.org/10.1017/S0142716414000599

Howes, C., Healey, P., & Purver, M. (2010). Tracking lexical and syntactic alignment in conversation. *Processes of Cognitive Science, 9,* 2004–2009.

Ireland, M. E., & Henderson, M. D. (2014). Language style matching, engagement, and impasse in negotiations. *Negotiation and Conflict Management Research, 7,* 1–16. http://dx.doi.org/10.1111/ncmr.12025

Ireland, M. E., Slatcher, R. B., Eastwick, P. W., Scissors, L. E., Finkel, E. J., & Pennebaker, J. W. (2011). Language style matching predicts relationship initiation and stability. *Psychological Science, 22,* 39–44. http://dx.doi.org/10.1177/0956797610392928

Jaeger, B. C., Edwards, L. J., Das, K., & Sen, P. K. (2017). An R^2 statistic for fixed effects in the generalized linear mixed model. *Journal of Applied Statistics, 44,* 1086–1105. http://dx.doi.org/10.1080/02664763.2016.1193725

Jaeger, T. F., & Snider, N. (2008). Implicit learning and syntactic persistence: Surprisal and cumulativity. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Meeting of the Cognitive Science Society* (pp. 1061–1066). Austin, TX: Cognitive Science Society.

Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition, 127,* 57–83. http://dx.doi.org/10.1016/j.cognition.2012.10.013

Johnson, P. C. (2014). Extension of Nakagawa & Schielzeth's R^2_{GLMM} to random slopes models. *Methods in Ecology and Evolution, 5,* 944–946. http://dx.doi.org/10.1111/2041-210X.12225

Jurafsky, D., & Martin, J. H. (2009). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Kaschak, M. P., Loney, R. A., & Borreggine, K. L. (2006). Recent experience affects the strength of structural priming. *Cognition, 99,* B73–B82. http://dx.doi.org/10.1016/j.cognition.2005.07.002

Kipp, M. (2012). Multimedia annotation, querying and analysis in ANVIL. In M. Maybury (Ed.), *Multimedia information extraction* (pp. 351–368). New York, NY: Wiley. http://dx.doi.org/10.1002/9781118219546.ch21

Lakin, J. L., Jefferis, V. E., Cheng, C. M., & Chartrand, T. L. (2003). The chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry. *Journal of Nonverbal Behavior, 27,* 145–162. http://dx.doi.org/10.1023/A:1025389814290

Lancaster, G., Iatsenko, D., Pidde, A., Ticcinelli, V., & Stefanovska, A. (2018). Surrogate data for hypothesis testing of physical systems. *Physics Reports, 748,* 1–60. http://dx.doi.org/10.1016/j.physrep.2018.06.001

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review, 104,* 211–240. http://dx.doi.org/10.1037/0033-295X.104.2.211

Lev-Ari, S. (2015). Selective grammatical convergence: Learning from desirable speakers. *Discourse Processes, 53,* 657–674. http://dx.doi.org/10.1080/0163853X.2015.1094716

Louwerse, M. M., Dale, R., Bard, E. G., & Jeuniaux, P. (2012). Behavior matching in multimodal communication is synchronized. *Cognitive Science, 36,* 1404–1426. http://dx.doi.org/10.1111/j.1551-6709.2012.01269.x

Main, A., Paxton, A., & Dale, R. (2016). An exploratory analysis of emotion dynamics between mothers and adolescents during conflict discussions. *Emotion, 16,* 913–928. http://dx.doi.org/10.1037/emo0000180

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511809071

Marcus, M. P., Marcinkiewicz, M. A., & Santorini, B. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics, 19,* 313–330.

Michael, J., Bogart, K., Tylén, K., Krueger, J., Bech, M., Østergaard, J. R., & Fusaroli, R. (2015). Training in compensatory strategies enhances rapport in interactions involving people with Möbius syndrome. *Frontiers in Neurology, 6,* 213. http://dx.doi.org/10.3389/fneur.2015.00213

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013, May). *Efficient estimation of word representations in vector space*. Paper presented at

the International Conference on Learning Representations, Scottsdale, AZ. Retrieved from https://arxiv.org/abs/1301.3781

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems, 26,* 3111–3119.

Miles, L. K., Lumsden, J., Richardson, M. J., & Macrae, C. (2011). Do birds of a feather move together? Group membership and behavioral synchrony. *Experimental Brain Research, 211,* 495–503. http://dx.doi.org/10.1007/s00221-011-2641-z

Miles, L. K., Nind, L. K., Henderson, Z., & Macrae, C. N. (2010). Moving memories: Behavioral synchrony and memory for self and others. *Journal of Experimental Social Psychology, 46,* 457–460. http://dx.doi.org/10.1016/j.jesp.2009.12.006

Mills, G. (2014). Dialogue in joint activity: Complementarity, convergence, and conventionalization. *New Ideas in Psychology, 32,* 158–173. http://dx.doi.org/10.1016/j.newideapsych.2013.03.006

Mitchell, R., Myles, F., & Marsden, E. (2013). *Second language learning theories* (3rd ed.). London, UK: Routledge.

Moscoso del Prado, M., & Du Bois, J. W. (2015, July). *Syntactic alignment is an index of affective alignment: An information-theoretical study of natural dialogue*. Proceedings of the 37th Annual Meeting of the Cognitive Science Society, CogSci 2015, Pasadena, CA.

Moulder, R. G., Boker, S. M., Ramseyer, F., & Tschacher, W. (2018). Determining synchrony between behavioral time series: An application of surrogate data generation for establishing falsifiable null-hypotheses. *Psychological Methods, 23,* 757–773. http://dx.doi.org/10.1037/met0000172

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution, 4,* 133–142. http://dx.doi.org/10.1111/j.2041-210x.2012.00261.x

Oben, B., & Brône, G. (2016). Explaining interactive alignment: A multimodal and multifactorial account. *Journal of Pragmatics, 104,* 32–51. http://dx.doi.org/10.1016/j.pragma.2016.07.002

Ollendick, T. H., & Schmidt, C. R. (1987). Social learning constructs in the prediction of peer interaction. *Journal of Clinical Child Psychology, 16,* 80–87. http://dx.doi.org/10.1207/s15374424jccp1601_10

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1964). *The measurement of meaning*. Chicago: University of Illinois Press.

Pardo, J. S., Gibbons, R., Suppes, A., & Krauss, R. M. (2012). Phonetic convergence in college roommates. *Journal of Phonetics, 40,* 190–197. http://dx.doi.org/10.1016/j.wocn.2011.10.001

Paxton, A. (2016). *Coordination: Theoretical, methodological, and experimental perspectives* (Doctoral dissertation). University of California, Merced, CA. Retrieved from http://escholarship.org/uc/item/5tx5s7zh

Paxton, A., & Dale, R. (2013). Argument disrupts interpersonal synchrony. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 66,* 2092–2102. http://dx.doi.org/10.1080/17470218.2013.853089

Paxton, A., & Dale, R. (2017). Interpersonal movement synchrony responds to high- and low-level conversational constraints. *Frontiers in Psychology, 8,* 1135. http://dx.doi.org/10.3389/fpsyg.2017.01135

Paxton, A., Dale, R., & Richardson, D. C. (2016). Social coordination of verbal and nonverbal behaviours. In P. Passos, K. Davids, & J. Y. Chow (Eds.), *Interpersonal coordination and performance in social systems* (pp. 259–274). New York, NY: Routledge.

Paxton, A., & Griffiths, T. L. (2017). Finding the traces of behavioral and cognitive processes in big data and naturally occurring datasets. *Behavior Research Methods, 49,* 1630–1638. http://dx.doi.org/10.3758/s13428-017-0874-x

Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language, 39,* 633–651. http://dx.doi.org/10.1006/jmla.1998.2592

Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: A critical review. *Psychological Bulletin, 134,* 427–459. http://dx.doi.org/10.1037/0033-2909.134.3.427

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27,* 169–190. http://dx.doi.org/10.1017/S0140525X04000056

Pickering, M. J., & Garrod, S. (2006). Alignment as the basis for successful communication. *Research on Language and Computation, 4,* 203–228. http://dx.doi.org/10.1007/s11168-006-9004-0

Pietsch, C., Buch, A., & Kopp, S. (2012). Measuring syntactic priming in dialogue corpora. In B. Stolterfoht & S. Featherston (Eds.), *Empirical approaches to linguistic theory: Studies in meaning and structure. Studies in generative grammar* (pp. 29–42). Berlin, Germany: Mouton de Gruyter.

Ramseyer, F., & Tschacher, W. (2014). Nonverbal synchrony of head- and body-movement in psychotherapy: Different signals have different associations with outcome. *Frontiers in Psychology, 5,* 979. http://dx.doi.org/10.3389/fpsyg.2014.00979

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Řehůřek, R., & Sojka, P. (2010, May). Software framework for topic modelling with large corpora. In R. Witte, H. Cunningham, J. Patrick, E. Beisswanger, E. Buyko, U. Hahn, . . . A. R. Coden, (Chairs.), *New Challenges for NLP Frameworks*. Workshop conducted at the 7th International Conference on Language Resources and Evaluation, Valletta, Malta.

Reitter, D., Keller, F., & Moore, J. D. (2011). A computational cognitive model of syntactic priming. *Cognitive Science, 35,* 587–637. http://dx.doi.org/10.1111/j.1551-6709.2010.01165.x

Reitter, D., & Moore, J. D. (2014). Alignment and task success in spoken dialogue. *Journal of Memory and Language, 76,* 29–46. http://dx.doi.org/10.1016/j.jml.2014.05.008

Reitter, D., Moore, J. D., & Keller, F. (2006). Priming of syntactic rules in task-oriented dialogue and spontaneous conversation. In R. Sun, & N. Miyake, (Eds.), *Proceedings of the 28th Annual Meeting of the Cognitive Science Society* (pp. 685–690). Austin, TX: Cognitive Science Society.

Richardson, D. C., & Dale, R. (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science, 29,* 39–54.

Rights, J. D., & Sterba, S. K. (2018). Quantifying explained variance in multilevel models: An integrative framework for defining R-squared measures. *Psychological Methods*. Advance online publication. http://dx.doi.org/10.1037/met0000184

Riley, M. A., Richardson, M. J., Shockley, K., & Ramenzoni, V. C. (2011). Interpersonal synergies. *Frontiers in Psychology, 2,* 38. http://dx.doi.org/10.3389/fpsyg.2011.00038

Rohde, D. L., Gonnerman, L. M., & Plaut, D. C. (2006). An improved model of semantic similarity based on lexical co-occurrence. *Communications of the ACM, 8,* 627–633.

Rowland, C. F., Chang, F., Ambridge, B., Pine, J. M., & Lieven, E. V. (2012). The development of abstract syntax: Evidence from structural priming and the lexical boost. *Cognition, 125,* 49–63. http://dx.doi.org/10.1016/j.cognition.2012.06.008

Sacks, H., Schegloff, E., & Jefferson, G. (1974). A simplest systematics for the organisation of turn-taking for conversation. *Language, 50,* 696–735. http://dx.doi.org/10.1353/lan.1974.0010

Scheepers, C. (2003). Syntactic priming of relative clause attachments: Persistence of structural configuration in sentence production. *Cognition, 89,* 179–205. http://dx.doi.org/10.1016/S0010-0277(03)00119-7

Schoot, L., Heyselaar, E., Hagoort, P., & Segaert, K. (2016). Does syntactic alignment effectively influence how speakers are perceived by

their conversation partner? *PLoS ONE, 11,* e0153521. http://dx.doi.org/10.1371/journal.pone.0153521

Slocombe, K. E., Alvarez, I., Branigan, H. P., Jellema, T., Burnett, H. G., Fischer, A., . . . Levita, L. (2013). Linguistic alignment in adults with and without Asperger's syndrome. *Journal of Autism and Developmental Disorders, 43,* 1423–1436. http://dx.doi.org/10.1007/s10803-012-1698-2

Sloetjes, H., & Wittenburg, P. (2008). Annotation by category: ELAN and ISO DCR. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, . . . D. Tapias, (Eds.), *Proceedings of the 6th International Conference on Language Resources and Evaluation* (pp. 816–820). Paris, France: European Language Resources Association.

Szmrecsanyi, B. (2005). Language users as creatures of habit: A corpus-based analysis of persistence in spoken English. *Corpus Linguistics and Linguistic Theory, 1,* 113–150. http://dx.doi.org/10.1515/cllt.2005.1.1.113

Tan, P. N., Steinbach, M., & Kumar, V. (2005). *Introduction to data mining*. Boston, MA: Pearson Addison Wesley.

Toutanova, K., Klein, D., Manning, C., & Singer, Y. (2003). Feature-rich part-of-speech tagging with a cyclic dependency network. In M. Hearst, & M. Ostendorf, (Chairs.), *Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 173–180). Stroudsburg, PA: Association for Computational Linguistics.

Wang, Y., & Hamilton, A. F. D. C. (2012). Social top-down response modulation (STORM): A model of the control of mimicry in social interaction. *Frontiers in Human Neuroscience, 6,* 153. http://dx.doi.org/10.3389/fnhum.2012.00153

Wang, Y., Reitter, D., & Yen, J. (2014, June). Linguistic adaptation in conversation threads: Analyzing alignment in online health communities. In V. Demberg, & T. O'#x2019;Donnell, (Chairs.), *Cognitive modeling and computational linguistics*. Workshop conducted at the 52nd Annual Meeting of the Association for Computational Linguistics, Baltimore, MD. Retrieved from http://acl2014.org/acl2014/W14-20/pdf/W14-2007
.pdf

Weatherholtz, K., Campbell-Kibler, K., & Jaeger, T. F. (2014). Socially-mediated syntactic alignment. *Language Variation and Change, 26,* 387–420. http://dx.doi.org/10.1017/S0954394514000155

Woods, D., & Fassnacht, C. (2017). *Transana v3.10*. Madison, WI: Spurgeon Woods LLC. Retrieved from https://www.transana.com

Xu, Y., & Reitter, D. (2015). *An evaluation and comparison of linguistic alignment measures*. Stroudsburg, PA: Association for Computational Linguistics.

Xu, Y., & Reitter, D. (2017). Spectral analysis of information density in dialogue predicts collaborative task performance. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics* (pp. 623–633). Retrieved from http://www.aclweb.org/anthology/P/P17/P17-1058.pdf

Yabar, Y., Johnston, L., Miles, L., & Peace, V. (2006). Implicit behavioral mimicry: Investigating the impact of group membership. *Journal of Nonverbal Behavior, 30,* 97–113. http://dx.doi.org/10.1007/s10919-006-0010-6

(*Appendix follows*)

Appendix

Converting Duran and Fusaroli's (2017) "Devil's Advocate" Spoken Interactions Into ALIGN-Formatted Conversational Corpora

The Duran and Fusaroli (2017) study consisted of multiple dyads engaging in extended, face-to-face spoken conversations. These conversations were recorded and saved as .WAV files. Using a free and widely-used speech analysis computer program called Praat (Boersma, 2001), the continuous speech stream for each participant was manually segmented into start and stop boundaries corresponding to a participant's speech turn. A speech turn was understood to begin when one participant took up the conversational floor and ended when that participant relinquished it to his or her conversational partner. This is distinct from back-channels, where listeners speak for social or metaconversational purposes while their conversational partner continues to hold the conversational floor. In Praat, identifying the start and stop times can be done via auditory and visual cues (e.g., audio playback features, onset and offset of energy peaks in the waveform). Once segmentation was complete, a team of research assistants directly transcribed exactly what was heard into each boundary-marked region. Typos or punctuation inconsistencies during transcription were corrected within the ALIGN code protocols (as described in the Phase 1 section of the main article).

Researchers may also want to consider more automated "first-pass" methods of automatically extracting speech turns from audio files and then having a human rater correct for errors. Recent advances in speech-to-text programs (e.g., Google Speech API) also make automated transcription an increasingly viable option.

To give an example of the general approach used above, Figure A1 shows a short segment of a longer conversation where conversational turns have been manually "boundary-marked" (blue lines) and transcribed. Praat automatically generates an output file that separates out each bounded region (speech) from nonbounded regions (pauses), tagged with the start and stop times for each region type, the transcribed speech if present, and whether the region corresponds to Participant 1 or Participant 2. From here, researchers can compile this information as separate column-wise variables within a single .csv file, doing so across all dyads and participants (as shown in Table 1 of the main article).

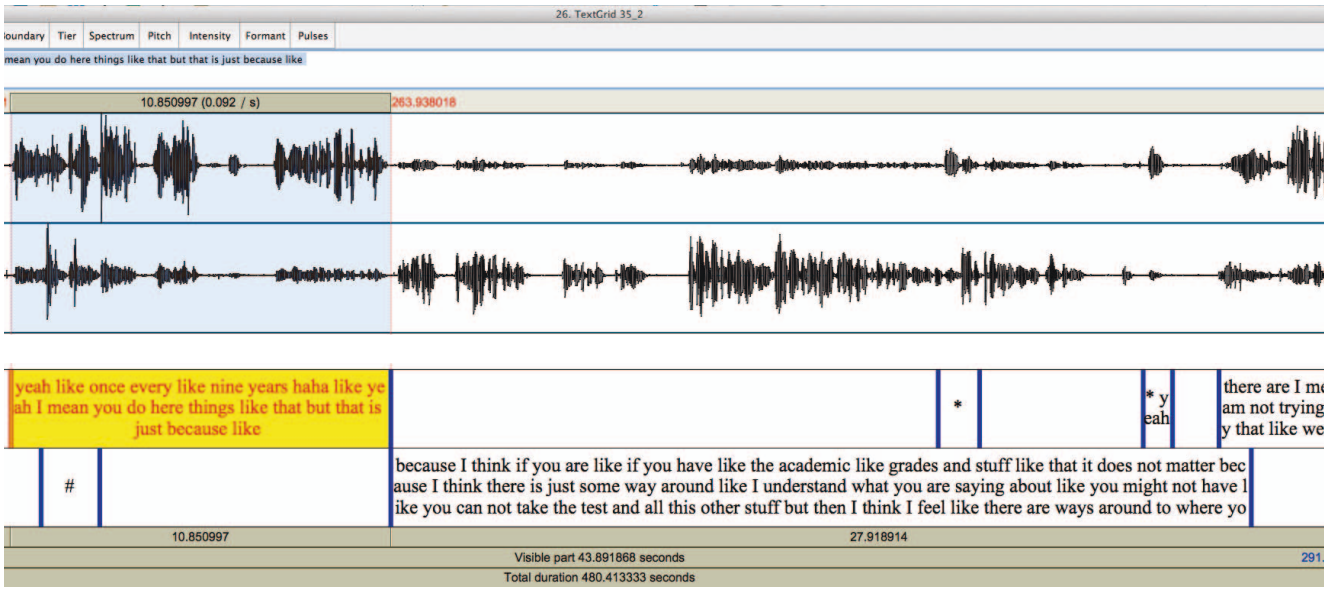

Figure A1. Example of initial data preparation to manually mark boundaries of conversational turns and transcription to get turn initiation time and duration, separated by speaker (Participant 1 and Participant 2) and other conversational conditions. See the online article for the color version of this figure.

Received February 12, 2018
Revision received November 17, 2018
Accepted November 25, 2018 ■