# When Faithful Informational Exchange is Just not Worth it: Reformulation Ability as a Predictor of Other-Initiated Repair

Thesis Submitted in Partial Fulfillment
of the Requirements
of the Jay and Jeanie Schottenstein Honors Program

Yeshiva College
Yeshiva University
September 2021

Louis Teitelbaum

Mentor: Professor Bruno Galantucci, Psychology

**Abstract**

Galantucci et al (2020) showed that when people are given instructions that they do not understand, they do not always ask clarifying questions. Why do they keep their confusion to themselves? Are people more likely to repair some kinds of misunderstandings than others? I hypothesize that people avoid clarification questions when they perceive them as requiring excessive effort; if people think that they must figure out a new way to phrase an issue in order to gain clarification, they will more probably avoid doing so. In two analyses of data from the HCRC Map Task Corpus, I find correlational evidence for an implication of this hypothesis: That individuals more able and willing to reformulate utterances (i.e. those whose utterances are less similar to any that came before) initiate clarification more frequently than those who reformulate less frequently (for whom reformulation may be more difficult). Finally, I propose an experiment to test the hypothesis more directly: Participants will receive instructions from a confederate on how to move cards from a tray onto a chessboard and then back into a tray in a setup identical to that of Galantucci et al. (2020). Here, though, the instructions will use only strings of words for individual properties, without any grammatical structure. In critical rounds, the lack of grammar will result in referential ambiguity, forcing the participant to choose whether or not to ask a clarifying question. Crucially, due to the lack of grammar, the question will require effortful reformulation. Following this thesis' main hypothesis, I predict that participants will avoid repair more often than did those in Galantucci et al.'s study, for whom repair initiation was relatively effortless.

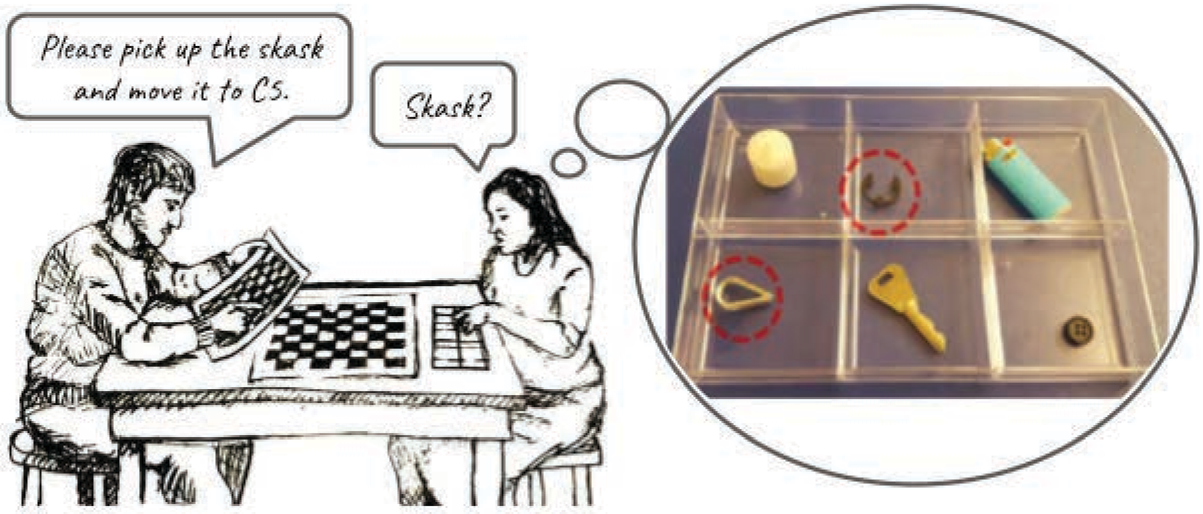
**Contents**

Imagine that you are getting instructions from a supervisor at work. You did not quite understand the last instruction, and you consider whether to ask her what she meant (i.e., initiate repair of the communication problem) or just let it slide (i.e. avoid repair)[1]. In this scenario, you and your supervisor are engaged in a purely informational exchange, the goal of which is to ensure that you perform the task correctly. Research in human communication has traditionally assumed that participants in a conversation act to establish as much mutual understanding as is required by the task at hand (e.g. Clark & Wilkes-Gibbs, 1986; Grice, 1975). In conversations such as the one described above, the necessary understanding could be achieved by initiating repair (e.g. "could you repeat what you said about our analytics?"). Nevertheless, a recent study by Galantucci and colleagues has revealed that repair avoidance is surprisingly common, even in purely informational, task-driven contexts (Galantucci et al., 2020). A confederate instructed participants to "pick up the skask" from a tray containing six objects and move it to a specific location. Even though "skask" is a non-word invented by the experimenters, 29 of the 48 participants avoided a repair which they could have easily performed with a mere repeat (e.g. "Skask?").

**Figure 1**

*Critical Trial Procedure in Galantucci et al. (2020)*

[1] Following Schegloff et al. (1977), researchers generally distinguish between other-initiated repair (OIR), such as the one described here, and self-initiated repair, in which a speaker clarifies his own utterance. Unless otherwise noted, the term "repair" in this thesis refers to the former of these two categories.




*Note*. The confederate (left) instructs a participant (right) to pick up the "skask" and move it to a specific location on a chess board. In the tray shown at right, used in critical trials of the study, four objects have well-known names (e.g. candle, key, button), while two (circled red) do not and are therefore might reasonably be identified as the "skask". In this condition, 45% of participants did not ask a clarifying question ("e.g. "skask?" as shown above).

Repair avoidance, it seems, is rather common. Why? What factors cause interlocutors to avoid conversational repair? An empirically based answer to this question could aid the development of protocols to reduce the probability of catastrophic misunderstanding in a wide range of contexts. Galantucci et al. (2020) found that repair avoidance occurred more frequently when avoiding the clarification was unlikely to result in an error in the execution of the instruction. This suggests that people's decisions to avoid repair are driven by a judgement that overt consequences are unlikely. The likelihood of overt consequences may explain why people decide to initiate repair when their default choice is avoidance, but why is avoidance the default? Wouldn't the least likelihood of overt consequences be achieved by never avoiding repair? In this thesis, I propose that the judgement of the overt cost of repair avoidance is counteracted by a parallel judgement of a more covert consequence of the alternative: cognitive effort. In other words, people avoid repair because repair takes effort.

The first four sections of the thesis outline important background material for the studies presented below. "Why ask for Clarification" illustrates the potential cost of repair avoidance by reviewing available literature on the benefits of repair in conversation. "Methodological Concerns", acknowledges difficulties in labelling behavior as repair avoidance and identifies requirements for a valid measurement of repair avoidance. "A Note on Asymmetrical Informational Interactions" defends the focus on such interactions in the studies below.

Study 1 explores three alternative operationalizations of cognitive effort in repair taken from the literature (Albert & de Ruiter, 2018; Dideriksen et al., 2020; Dingemanse et al., 2015), and one novel operationalization. These operationalizations are applied to data from the HCRC Map Task Corpus (Anderson et al., 1991) and are used to show evidence for the hypothesis that individuals who are more willing to reformulate their repair initiations initiate repair more frequently overall (the Individual Effort Hypothesis).

Study 2 implements a computational approach to measure individual willingness to reformulate utterances, using TF-IDF sentence similarity scores (Spärck Jones, 1972) within the Map Task Corpus. This new operationalization is shown to predict repair frequency, lending further support for the hypothesis that variation in repair avoidance can be partially explained in terms of individual reformulation willingness or ability.

The final section of the thesis proposes an experiment to test directly the impact of cognitive effort saving on repair avoidance. Other potential future directions for research are then discussed, as well as the general implications of the research presented here.

## Background

### Why Ask for Clarification?

Before addressing reasons why people might avoid conversational repair, let us ask a more basic question: why do people repair at all? A clear understanding of this will make plain why repair avoidance may lead to problems. In the article that first introduced the term "repair", Schegloff et al. (1977, p. 381) declared that "the organization of repair is the self-righting mechanism for the organization of language use in social interaction". Naturally then, any complete account of the purpose of conversational repair must also address the purpose of "the organization of language use in social interaction" in general. A discussion of the purpose of language is beyond the scope of this thesis, but I will nevertheless review evidence for three advantages of conversational repair.

Perhaps most obviously, repair promotes shared understanding (otherwise referred to as faithfulness). It does this by providing negative evidence of understanding (Clark & Brennan, 1991), thereby encouraging the utterer of the trouble source to offer clarifying information. Some repair strategies go even further by offering possibilities for what the speaker may have meant or shaping participants' shared conceptual framework to ease further communication, as we shall see shortly. An individual's motivation to understand may arise from personal factors (e.g. curiosity or personal goals related to the information being conveyed), situational factors (e.g. anticipation of consequences if information is not properly conveyed), or social considerations (e.g. a desire to demonstrate that one cares about what one's interlocutor is talking about). Regardless of the motivation for understanding, repair has been experimentally shown to be effective in improving communicative faithfulness (Mills, 2014b). As such, it is more frequent in contexts that demand greater mutual understanding (Colman & Healey, 2011), and more specific[2] repair initiations (i.e. more rigorous ones) seem to be related to communicative success in those contexts (Dideriksen et al., 2020)[3]. Conversely, repair is not necessary in non-informational interactions such as phatic communion (commonly known as "small talk"; Malinowski, 1923), because participants have no need to maintain shared understanding. One informal conversationalist observed by Galantucci et al. (2018) therefore had no problem responding to a nonsense phrase interjected by his interlocutor by laughing briefly, saying "all right", and continuing with the conversation, despite reporting afterward that he had noticed the strange utterance.

[2] The terms "specific" and "restricted" are used interchangeably in this thesis.

[3] Some of these results should be treated with caution. In support of the argument presented here, Dideriksen et al. found that restricted offers were positively related to performance in the Map Task. However, they also found that more open repair strategies were mostly unrelated or negatively related to performance in the Map Task. These conflicting data likely reflect the situationally-dependent usage of repair: on the one hand, increased repair may in general help communication. On the other hand, dyads who have difficulty communicating in general may attempt to remedy the problem with increased repair, to limited success. Furthermore, as discussed below, individuals who have trouble generating more specific repairs (and thus display a higher relative frequency of open-type repairs) may lack skills critical to communicative faithfulness in general.

The pro-faithfulness effects of repair are obviously due in part to the role of individual repair sequences in recovering information from an utterance that would otherwise be misunderstood or missed entirely, as in the following exchange.[4]

G *until you you get over the top of the slate mountain*

F ***over the top of the***

G *slate mountain*

If F had not initiated repair, she may never have known what she was meant to get over the top of. F's repair initiation therefore has a local pro-faithfulness effect. Most researchers understand this to be the primary purpose of repair (e.g. Dingemanse et al., 2015). Nevertheless, people often initiate repair and subsequently realize that they understand the trouble turn before the repair sequence can be completed:

G *now eh if you go follow the line of the lake follow the line l-- of the lake left*

F ***going west?***

F ***right***

F has confirmed her understanding of "left" as "west" before G has even had a chance to respond (cf. Kendrick, 2015b, p. 6) ! In such cases, the repair would seem to be a waste of time; if repair is not necessary to repair trouble, how else can it be explained?

First, it is possible that repair functions as a buffer (the buffer theory), buying time for a conversationalist while they process the preceding utterances and cognitively catch up to

[4] All quoted dialogues are from the HCRC Map Task Corpus (Anderson et al., 1991) unless otherwise noted. In these dialogues, partner G (Giver) is instructing partner F (Follower) on how to navigate a particular route on a map. Both partners are looking at roughly similar versions of the map, but only G can see the route. Pronouns used to refer to either participant are arbitrary and may not reflect their actual gender.

the conversation. Some empirical evidence makes this possibility difficult to accept. In particular, the ability to pause a conversation does not help overhearers better comprehend references (Schober & Clark, 1989), and listeners are able to adapt very quickly to comprehending rapid speech, even when it is compressed to less than half of its normal duration (Dupoux & Green, 1997). Nevertheless, given the now widely recognized fact that real-time interlocutors engage in interactive conceptual alignment in a way that passive listeners do not (see Pickering & Garrod, 2004; Levinson S. C., 2016), such findings do not altogether disprove the buffer theory.

Even without the buffer theory, it is likely that repair has benefits to communicative faithfulness beyond the content of the individual repair sequence. Consider the following exchange.

G *just straight down to the right of the buffalo.*

F ***to the right?***

G *y--*

G *so that… so that you're on… you're on the right of the buffalo.*

With her repair initiation, “to the right?”, F seeks to verify her comprehension of G’s instruction. G begins to respond with a simple confirmation, presumably “yeah”; after all, F’s “to the right” was a verbatim repetition of the original instruction. Nevertheless, G interprets the repair initiation as negative evidence of understanding in general and attempts to remedy this with a further clarification: “so that… you’re on the right side of the buffalo”, lest F think that she should travel to the right of her current location. We see, therefore, that repair initiation facilitates communicative faithfulness both for its immediate content (i.e. it signals mis- or non-understanding of a particular referent) and for its broader context (by prompting further clarification). Similarly, a growing body of experimental work by Patrick Healey and colleagues suggests that other-initiated repair drives the development of more efficient (i.e. shorter and more generalizable) referential expressions over the course of a conversation (Healey, 2008; Healey et al., 2003; 2018a; Mills, 2014b). In the following exchange, for example, F's repair initiation may have a global pro-faithfulness effect in solidifying a shorter, mutually agreed-upon way to refer to the iron bridge.

G *go up as if you're can cross over the iron bridge or whatever it is you've got there*

F ***so i g-- i'm going to cross over the iron bridge?***

G *cross over the bridge*

When either G or F refers to "the bridge" again later in the conversation, they will have already established this reference as common ground, and the other will understand what feature is being indicated.

Of course, the role of other-initiated repair in developing efficient referential expressions (lexical alignment) points to another of its beneficial effects: effort saving. If repair allows conversational partners to agree more quickly on shorter shared expressions, it may also save them time and energy. This hypothesis is supported by the high frequency of repair even in the least informationally demanding circumstances; in informal face-to-face conversation between family and friends, other-initiated repair occurs on average once every 1.4 minutes across a wide variety of languages (Dingemanse et al., 2015).

In summary, conversational repair promotes communicative faithfulness both locally and throughout the conversation. It also may sometimes function as a buffer to allow conversationalists to process speech before moving on with the conversation, and may also save effort for interlocutors.

**Why People Avoid Asking for Clarification: Effort Saving**

If conversational repair is a boon to faithfulness, why would people such as the participants in the "skask" study (Galantucci et al., 2020) avoid repair with such high frequency? This thesis will explore one possible answer to this question: in language, as in general, people don't like to work any harder than they have to (Clark & Brennan, 1991). Researchers have long recognized effort-saving as influencing various aspects of conversational repair behavior[5]. Generally, this effort saving has been described as collaborative. In other words, individual participants are willing to put in some of their own effort in order to lessen the combined effort of all participants (Clark & Wilkes-Gibbs, 1986; Clark & Brennan, 1991). For example, repair initiators prefer to initiate the most specific repair type possible, thereby minimizing the effort required for the interlocutor to identify the problem and present the solution (Dingemanse et al., 2015)[6]. This preference is immediately visible in the following exchange:

[5] As Clark and Brennan (1991, p. 226) note, Grice himself organized his maxims of quantity and manner around a form of cognitive effort saving: the ideal of conveying the least amount of information sufficient for current purposes.

[6] For the purposes of this proposal, "initiator" refers to the person initiating repair, while "initiation" refers to that person's utterance. The speaker of the trouble turn will be referred to as such, or as "the interlocutor". "Completion" refers to the achievement of mutually accepted understanding, usually finalized by the speaker of the trouble turn. "Repair sequence" refers to the entire interaction between the repair initiator and his interlocutor which resolves an identified mis- or non-understanding, a process that may include multiple repair initiations.

G *well i want you to do a back-- a backwards "c" as if you're you're going round farmland and cutting the river cutting across the stream just below the dead tree*

F *right*

F *okay*

G *so that ins-- stop when you get to the bottom of the dead tree*

F ***so remind me what i'm... i'm… i've got to go underneath the dead tree?***

G *no*

G has uttered a confusing instruction and F, it seems, has not understood. F therefore initiates repair with a general request for clarification, "so remind me what I'm…"[7]. In the middle of producing this request, however, F aborts and instead produces a specific offer, "I've got to go underneath the dead tree". By producing a specific offer instead of a request, F is saving effort for G, who now can give a simple "yes" rather than reiterating her entire instruction or trying to guess what it is that F did not understand. Even if G rejects F's specific offer, as in fact she does, G can tailor his clarification to F's particular misunderstanding, which she has now expressed as an offer. The preference for using the most specific possible repair initiator also minimizes the joint effort of speaker and interlocutor, as demonstrated by measurements of the length of conversational turns (Dingemanse et al., 2015; Dideriksen et al., 2020). In the above example, the effort that F exerted in generating her specific offer was likely worth it for the greater effort it saved G.

[7] Within the traditional framework (e.g. Dingemanse et al., 2015), this initiation would generally be considered a "specific request" as opposed to an open-type one such as "what?" or "huh?". Nevertheless, G's request here is quite general; its only referential specificity is in indicating that she is referring to an earlier instruction, not G's most recent one.

In light of this connection between joint effort saving and repair behavior, it seems likely that effort saving would drive repair *avoidance* as well. Studies of corpus data have already observed that in cases where there is intervening material between the trouble source and the initiation, or where the trouble source is long, open-type repairs (e.g. “what?”) are less likely (Dingemanse et al., 2015). This is presumably because in these cases, open-type repairs would not be sufficient to identify the referent, as in a scene from *The Love Match*, by Henry Cockton, in which Chump bursts in on Mr. Jowles and initiates the following conversation (p. 74):

> “Mr Jowles,” said he, “What do you mean?”
>
> “What do I mean?”
>
> “Aye; what do you mean?”
>
> “What do I mean by what?”

At this point, a considerable amount of time has elapsed since the trouble source was uttered, and when Chump utters an open-type repair initiation, Jowles has no idea which of his utterances he is now meant to clarify. This is perhaps an extreme example, but it demonstrates a crucial point: indicating a particular utterance to be clarified—as well as what about it needs to be clarified—is no simple task. In order to effectively initiate repair in the above conversation, Chump would have had to say something like, “What clothing were you thinking of when you told the general that the man was dressed like a gentleman’s servant?”. Presumably, restricted-type repairs like this one will not always be possible (or worthwhile) to generate. To formulate such an utterance requires time, physical effort, and cognitive effort, including consideration of the interlocutor’s bases of grounding, a process which has been demonstrated to require attentional resources (Apperly et al., 2006) and be relatively effortful (Lin et al., 2010). When the cost of this effort outweighs the benefit of clarification, people may avoid repair altogether. The aforementioned studies of corpora, which found a decreased proportion of open-class initiations in cases of difficult reference, cannot comment on those times when repair was avoided altogether. We, however, can hypothesize that the need for more restricted-type initiations in such cases sometimes lead people to avoid repairs altogether.

Cognitive effort has been demonstrated to be a driving factor of behavior in information acquisition (Simon, 1956), attention allocation (Sims, 2006; Vul et al., 2014), and numerous decision-making heuristics (Gabaix 2017; Lieder & Griffiths, 2020). Likewise, it has been implicated as a primary driver of linguistic behavior (Hawkins, 2004; Kemp & Regier, 2012; Regier et al., 2007; Zaslavsky et al., 2018; Zipf, 1949). Most strikingly, Engelhardt et al. (2013) found that approximately one-third of the variance in self-initiated repairs can be accounted for by individual differences in inhibitory control (a particular cognitive ability). They also showed significant correlations between repair disfluencies (presumably reflective of difficulty with sentence formulation; cf. Williams & Korko, 2019) and two intelligence subtests. It stands to reason, then, that other-initiated repair avoidance also occurs as a result of the cognitive effort associated with formulating a repair initiation. Consider again the example given at the beginning of this thesis: you have received an ambiguous instruction from your supervisor at work, and you consider whether to ask her what she meant or just let it slide. This type of hesitation may reflect social considerations or time constraints, but may also be driven by a desire to avoid the effort of explaining what, exactly, you did not understand (or what additional information you might need). Does the cognitive effort required to reformulate certain problems in communication deter people from initiating repair?

At this point, it behooves us to briefly consider the nature of cognitive effort in general. Effort is on the one hand easy to intuit—we have all had the experience of "just not feeling like it"—and on the other hand notoriously difficult to define precisely. Some have theorized that cognitive effort is analogous to muscular exertion; certain tasks use up resources more quickly than others, and when these resources are depleted, the task can no longer be performed. This analogy, however, is not supported by available evidence (Westbrook & Braver, 2015). It seems more likely that, rather than using up physical resources, effortful activities occupy more general *cognitive* resources and therefore entail an opportunity cost. In other words, the decision to engage in an effortful cognitive activity precludes involvement in other tasks that may be important (see Westbrook & Braver, 2015 for a review). In language, for example, the benefits of an effortful utterance must outweigh the costs of sacrificing cognitive capacity in the middle of conversation. This model of cognitive effort suggests that the most effortful activities will be those that require use of the most generalized cognitive functions, such as working memory or inhibitory control, which are required for a wide variety of important tasks. As we have seen, some experimental evidence already suggests the importance of such generalized functions in generating novel repairs (Engelhardt et al., 2013).

In the section above entitled "Why Ask for Clarification?" I outlined critical positive outcomes of repair initiation for communicative faithfulness, as well as for cognitive effort-saving. These benefits may be integrated with the putative cognitive costs associated with repair initiation to formulate a cost-benefit model of repair avoidance, for which the coefficients need to be ascertained for any given context. If the probability of the speaker initiating repair is $P_{\text{Repair}}$, and costs and benefits of possible courses of action are represented by *C* and *B*, respectively, the model would be expressed as follows:

$$P_{\text{Repair}} = ( PB_{\text{Repair}} - PC_{\text{Repair}}) - ( PB_{\text{Repair Avoidance}} - PC_{\text{Repair Avoidance}} )$$

Or, more specifically:

$$P_{\text{Repair}} = ((PB_{\text{Comprehension}} + PB_{\text{Cognitive Effort}} )_{\text{Repair}} - (PC_{\text{Cognitive Effort}} )_{\text{Repair}})$$

$$- ((PB_{\text{Cognitive Effort}} )_{\text{Repair Avoidance}} - (PC_{\text{Comprehension}} + PC_{\text{Cognitive Cost}} )_{\text{Repair Avoidance}})$$[8]

In other words, the probabilities of incurring cost and benefits as a result of repair initiation are weighed against the probabilities of incurring costs and benefits as a result of repair avoidance. If the net benefit of repair avoidance is greater than that of repair, the speaker will be unlikely to ask for clarification. This model allows us to make a number of testable hypotheses. These will be the foci of the two corpus studies and the experimental procedure proposed below.

---

[8] $(PB_{\text{Comprehension}})_{\text{Repair}}$ : The benefit of repair to comprehension times the probability of that benefit if the speaker initiates repair.

$(PB_{\text{Cognitive Effort}})_{\text{Repair}}$ : The benefit of cognitive effort saved by repair times the probability of that benefit if the speaker initiates repair.

$(PC_{\text{Cognitive Effort}})_{\text{Repair}}$ : The cost of cognitive effort expended in repair times the probability of that benefit if the speaker initiates repair.

$(PB_{\text{Cognitive Effort}})_{\text{Repair Avoidance}}$ : The benefit of cognitive effort saved by repair avoidance times the probability of that benefit if the speaker avoids repair.

$(PC_{\text{Comprehension}})_{\text{Repair Avoidance}}$ : The cost of repair avoidance to comprehension times the probability of that benefit if the speaker avoids repair.

$(PC_{\text{Cognitive Effort}})_{\text{Repair Avoidance}}$ : The cost of cognitive effort incurred by repair avoidance times the probability of that benefit if the speaker avoids repair.

**Methodological Concerns: Content Deafness and the Difficulty of Labelling "Problems"**

Before testing hypotheses in answer to the question of why people avoid asking for clarification, I will briefly address the possibility that the question itself is based on false premises: perhaps people do not avoid repair at all. Indeed, when regarding a conspicuous lack of repair of a communicative problem, we can imagine explanations that would not constitute "avoidance".

In another study by Galantucci et al. (2018), participants had informal one-on-one conversations with a partner, ranking five "would you rather" questions (e.g. "Would you rather live the rest of your life on a tree or in a cave?") in order of humorousness and oddness. In the middle of the conversation, the partner—who was in fact a confederate of the lab—uttered the words "colorless green ideas sleep furiously". One participant responded with "you never played would you rather?" Another just said "yeah" and continued with the conversation. When interviewed soon after the confederate uttered the nonsense sentence, 20 of the 30 participants–including the two just mentioned–insisted that no such sentence had been introduced into their conversation. Of the ten who thought that such a sentence had been introduced, only one could correctly recognize it in a list with 20 other nonsensical sentences. This phenomenon has been dubbed "content deafness" (Galantucci et al., 2018. cf. Fenn et al., 2011; Simons & Levin, 1998) and it presents a difficult methodological problem in the study of repair avoidance: sometimes people do not perceive problems in their communication. In such cases, it cannot be said that people are avoiding repair; they simply misperceived or did not hear. "Noticing" is a response that is difficult to define and still more difficult to measure; self-report measures, for example, cannot distinguish between problems of memory and perception [9]. Did the 20 participants who denied that "colorless green ideas sleep furiously" had been uttered actually not perceive it as a problem in the first place, or had they merely forgotten? The objectionability of labelling "problems" that were never explicitly pointed out by conversants is one of the reasons cited by conversation analysts to defend their unwillingness to theorize about mental states or to use informational measures of communicative success (Albert & de Ruiter, 2018). Problems identified by researchers may not be perceived as such by conversants. Conversely, problems perceived as such by conversants may not be identified by researchers; speakers often repair minor infelicities in talk even when they are not essential to informational faithfulness (see Keysar, 2007).

Before Galantucci et al. (2020) provided convincing evidence that the participants in their study perceived problems as such and avoided repair nonetheless, the objectionability of labelling "problems" may have cast considerable doubt on a basic premise of this thesis—that repair avoidance is common. Now that the prevalence of repair avoidance has been demonstrated experimentally, research such as this thesis may proceed to investigate its causes. Nevertheless, the three studies presented here will necessarily be concerned with overcoming the above-mentioned difficulties with internal validity. These difficulties will be addressed below with a careful choice of conversational context (asymmetrical informational interactions with overt consequences of misunderstanding), an operational definition of repair avoidance that does not label "problems", and, in Study 3, experimental control.

[9] About problems with self-report in general see Nisbett & Wilson, 1977. In the context of communicative trouble see Micklos et al., 2020.

**A Note on Asymmetrical Informational Interactions**

The studies below—both the Map Task Corpus analyses and the proposed experimental procedure—deal with highly asymmetrical informational interactions. That is, conversations in which the participants work together to build a shared understanding of information (rather than, for example, social lubrication) and in which one partner possesses most of the relevant information. The reason for my focus on this type of interaction is practical; it is the paradigmatic setting of other-initiated repair (and the setting in which it is most common; Dideriksen et al., 2020). Other-initiated repair is (among other things) a cooperative method to remedy mis- or non-understanding in a context where one partner (the initiator) recognizes the problem but does not have access to the solution without the help of his partner (the interlocutor). Its paradigmatic case is therefore one in which all the pressure for understanding is placed on one partner, who therefore must recognize communicative problems, whereas the other partner has all the access to the information being conveyed. Given that my focus is on repair *avoidance*, it is important that I deal with such contexts, in which it can be reasonably claimed that the (potential) initiator will not understand critical information in cases where repair fails or is avoided.

Of course, most human communication—and likewise most conversational repair—does not occur in such extreme contexts. As Dingemanse and colleagues (2015) have shown, other-initiated repair occurs on average once every 1.4 minutes even in informal conversation between family and friends, i.e. in conversations in which 1. the focus is not primarily informational, 2. both participants share substantial common ground in advance of the conversation, 3. information is not held asymmetrically, and 4. little pressure is put on participants to understand communicated information accurately. It is my hope that my investigations into asymmetrical informational contexts can shed light on the psychological underpinnings of repair behavior generally, but it will be the job of future researchers to determine whether any of the findings presented here are truly generalizable.

Nevertheless, many important types of human communication *are* asymmetrical informational interactions very much like those under study here. Oral interaction between scientists and specialized journalists, which precedes the writing of science popularization texts targeted for the lay reader, is one example of such interactions (Ciapuscio, 2003).

## Study 1: Reformulation Effort and Helpfulness in Repair Initiations

In the above sections I have proposed that repair avoidance may be driven by cognitive effort saving. If this is so, might individual differences in the propensity to avoid repair be explained by individual differences in the propensity to expend cognitive effort? The bulk of Study 1 below will examine this question. In particular, I hypothesize that individuals who are more likely to reformulate in attempts to repair (i.e. who are more willing or able to expend cognitive effort in reformulation) will avoid repair less frequently, and by extension achieve greater faithfulness in conversation. This prediction is based on two assumptions. The first of these is that the ability to efficiently reformulate utterances is relatively stable within individuals. Given that reformulation of utterances likely involves well-documented aspects of general intelligence (Engelhardt et al., 2013), this is almost certainly true. This idea also complements work of Krauss & Glucksberg (1977) and Keysar (2007) which suggest that the generation of non-egocentric references (i.e. effective references to anything that is eminently known by the initiator but not his interlocutor) is cognitively demanding and must be learned. Repair initiations often require non-egocentric references, because the initiator refers to problems that are well known to him, but may not be obvious to the speaker of the trouble source (see Alexander et al., 1997). The second necessary assumption is that *willingness* to exert cognitive effort in reformulation is stable within individuals. This is also likely to be true; the disposition to engage with and enjoy cognitively demanding tasks is well documented as a stable trait and predicts higher academic achievement and standardized testing scores, among other adaptive outcomes (Cacioppo et al., 1996; Westbrook & Braver, 2015).

If it is true that the ability and willingness to exert cognitive effort in reformulation are stable within individuals, and it is true that repair avoidance is partially a result of the cost of cognition, as our model proposes, we can expect that those who are more able and willing to reformulate will initiate repair more often than those for whom reformulation is difficult (Individual Effort Hypothesis). In support of the Individual Effort Hypothesis, Dideriksen et al. (2020) found that relative frequencies of the two more open repair types (restricted request and open request) within repairs were negatively correlated with overall repair frequency in a joint decision-making game (the Alien Game)[10]. Since more open repair types probably take less effort in general than restricted offers (Dingemanse et al., 2015), this may indicate that those who in general expend less effort in their repair initiations also initiate repair less often. Conversely, Dideriksen et al. (2020) found in their analysis of a Danish Map Task corpus (in which one participant instructs the other on how to draw a route on a map) that restricted offers (likely the most effortful repair type) were positively correlated with

[10] Dideriksen et al. did not explicitly express the hypothesis in question.

frequency of repair in general. In other words, Alien game players who tended toward using open strategies in their repairs used fewer repairs altogether, and Map Task players who tended toward using more specific repair strategies initiated more repair altogether. These results lend some initial support to the Individual Effort Hypothesis, but are nevertheless far from conclusive given the nature of the measurement. First, Dideriksen et al. failed to replicate the above-described pattern of results in all of their analyses. Furthermore, as will be discussed below, the division of repair initiation into three categories (open request, restricted request, and restricted offer) yields at best a crude proxy for cognitive effort. Generating more precise measures in the pursuit of testing my hypothesis will be a primary focus of this analysis.

**Operationalizing Repair Avoidance**

In a preexisting corpus, we cannot reliably identify problems in conversation unless these problems are explicitly identified by participants (Albert & de Ruiter, 2018; see "Methodological Concerns" above). Consequently, we cannot quantify the extent to which participants avoid initiating repair of such problems. Nevertheless, the asymmetrical informational interaction affords us a singular advantage: Almost all utterances by the information giver are instructions that can be clarified (see "A Note on Asymmetrical Informational Interactions" above). The Map Task, as will be described below, has the additional advantage of being relatively difficult; shapes of curves on a map are nearly impossible to represent verbally. For these reasons, we can surmise that repair would be appropriate and helpful following the majority of giver utterances in a Map Task corpus[11]. In other words, the ratio of follower repair initiations to giver utterances in a conversation is likely to be a reasonable proxy for the extent to which the follower in that conversation did *not* avoid repair. This measure can be improved slightly by removing common, stereotypical utterances such as "yeah", which are unlikely to convey any information to the follower that would warrant further clarification, from the count of giver utterances[12]. I will therefore operationalize repair avoidance in a conversation as the inverse of the ratio of follower repair initiations (see Appendix C) to giver utterances, not counting common utterances:

$$\textit{Repair Avoidance Index} = \frac{\textit{Giver Utterances} - \textit{Common Utterances}}{\textit{Follower Repair Initiations}}$$

**Operationalizing Effort in Repair: Current Methods**

As we have seen, cognitive effort saving is likely to have a primary effect on other-initiated repair. Indeed, the most prominent current findings on repair initiation focus primarily on its potential to reduce cognitive effort in order to explain observed behavior (e.g. Dingemanse et al., 2015; Dideriksen et al., 2020). Surprisingly though, none of this research examined data using criteria that are directly designed to measure cognitive effort.

[11] This assumption is also supported by evidence that repair drives interactive alignment (and probably, therefore, communicative faithfulness) even when inserted artificially in conversation (Healey et al., 2003).

[12] Utterances not counted as giver utterances for this purpose were as follows: "aye", "erm no", "ehm no", "eh oh right", "nope", "er right okay", "'kay", "yep okay", "yep yep", "ehm well", "erm well", "uh-huh uh-huh", "alright", "and then", "oh right", "um mmhmm", "uh-huh right", "right so", "so", "okay so", "right um", "mm oh aye", "okay erm", "so ehm", "um", "right er", "mm", "ehm", "no", "no no", "no no no", "erm", "but", "now", "well", "oh right okay", "mmhmm", "yeah", "okay", "okay yeah", "right okay", "right", "uh-huh", "right-fine", "uh-huh yeah", "yep", "okay then". These were identified subjectively while reading through a sample of 15 conversations in the corpus. All instances of these utterances in the corpus were removed computationally while computing giver utterances.

Since the field of psychology that studies repair behavior is young and primarily derivative of the field of sociology called Conversation Analysis (Albert & de Ruiter, 2018), researchers have tended to use existing categorizations of repair initiators developed by the latter field. The most prominent categorizations of repair initiators used in Conversation Analysis (now the most prominent categorizations in the corresponding psychological field) are based on repair helpfulness (also called "strength", i.e. how helpful is a repair initiator in completing the full repair; see Albert & de Ruiter, 2018; Schegloff et al., 1977). Maximal helpfulness is achieved by offering a complete repair on the level of the speech act (i.e. verifying what the interlocutor is *doing* with his utterance; Holtgraves, 2002, pp. 5-33; see Searle, 1969) and requiring only its confirmation by the interlocutor[13] (e.g. "so I should close the door?"). Initiations of lesser helpfulness require the speaker of the trouble source to provide the necessary information. The least helpful initiators do not identify what needs to be repaired at all[14]. The literature has accordingly maintained the distinction between offers and requests, because offers need only be affirmed, while requests, by definition, request more information. This approach has therefore most often yielded three categories: open

[13] "Other-correction", though it is generally not in direct reference to the speech act, may also fall into this category (see Kendrick, 2015a), as in the following:

G *do you have white water*
F *er well that'll be the rapids*
G *yeah that's true*

[14] To be clear, this does not necessarily mean that the interlocutor must guess blindly at the problem following a repair initiation. A good deal of trouble source and type identification can be performed by the interlocutor, without explicit indication by an initiator, on the basis of conversational grounding and an awareness of what he has just uttered. For this reason, other-initiated and self-initiated repairs are not entirely distinct; both generally involve the cooperation of the interlocutor.

requests (e.g. “huh?”), restricted requests (e.g. “who?”), and restricted offers (e.g. “did you say Mr. Green?”).

Psychologically oriented researchers, in adopting categories like the one outlined above, posit that initiation strength is a reasonable proxy for cognitive effort. After all, stronger repair initiators do tend to be longer (Dingemanse et al., 2015), and generating longer utterances takes more effort to plan (Levinson, 2016). Nevertheless, the equation of effort and helpfulness is a rough one and has numerous exceptions. For example, many restricted offer sequences (purportedly the most costly type for the initiator) consist only of a verbatim repetition and a single-syllable affirmation (e.g. “Jack invited me to his party.” “Jack?” “Yeah.”). Assuming facilitation by a short-term phonological store of the previous turn (Baddeley & Hitch, 1974), such a sequence can be expected to have minimal cognitive cost for both participants. Indeed it is well documented that short, primed utterances can be generated much more quickly than those not primed (Levinson, 2016) and may even forgo any syntactic processing (Reitter & Moore, 2007). When testing the principle of least collaborative effort, a short, primed restricted offer should be viewed as fundamentally different from another restricted offer that exacts reformulation costs from the participants (e.g. “I read that referential expressions attenuate with time.” “You mean words I use get shorter the more I say them?” “I guess so, yeah”). This latter type presumably requires much more effort both on the part of the initiator, who must reformulate her interlocutor’s utterance and take the time to produce a longer initiation, and on the part of the interlocutor, who must check the match between the original and the new formulation. Given that almost half of repair initiations feature repetition of the trouble source (Dingemanse et al., 2015), the distinction between repair initiations that reformulate the problem source and those that do not is critical. For this reason, I focus here on developing new ways of measuring cognitive effort more directly, giving special attention to the effort of syntactic and semantic reformulation, which likely entail the most use of generalized cognitive function.

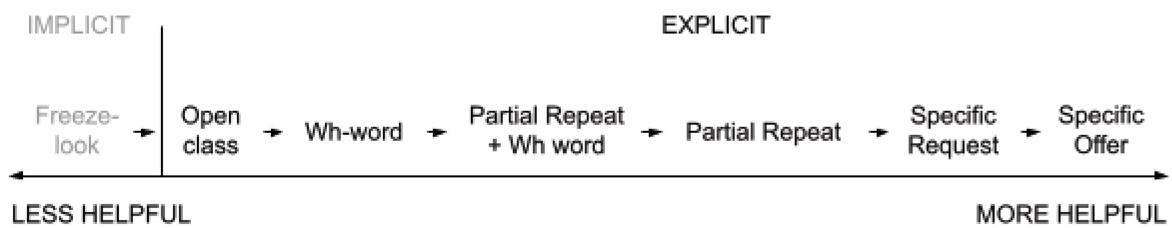
In order to ground this research firmly in existing literature, and in order to empirically assess the advantages and disadvantages of various operationalizations of cognitive effort in repair, I will first examine three alternative operationalizations taken from the literature:

**Orthographic length.** I recorded the orthographic length of each repair initiation as the number of characters, including spaces, in the turn as it appears in the HCRC Map Task Corpus transcript. Orthographic length of utterances is known to be highly correlated with turn duration and information content (Piantadosi et al., 2011), and has been used as a proxy for cognitive effort in repair initiation by Dingemanse et al. (2015).

**Helpfulness (Expanded Scale)**. I measured helpfulness (also called specificity or strength) of the repair initiation along the spectrum provided by Albert and de Ruiter (2018; see also Clift, 2016; Manrique & Enfield, 2015; Sidnell, 2011), with one additional level. As shown in Figure 2, this yielded six levels, from least to most helpful: 1. Open Class (e.g. huh?), 2. Wh-word (e.g. where?), 3. Wh-word + Partial repeat (e.g. close to what?), 4. Partial Repeat, 5. Specific Request (e.g. to the right or left?), and 6. Specific Offer (e.g. to the lake?). Note that this categorization does not map linearly onto the one more commonly used by researchers in this field (open request/specific request/specific offer), because most partial repeats would be categorized in the latter as specific offers and therefore *more* helpful than specific requests.

**Figure 2**

*Six Level Helpfulness Scale Used in this Study (cf. Albert & de Ruiter, 2018; Clift, 2016)*




**Helpfulness (Three Levels).** In addition to the above six level scale, I also measured helpfulness on the three level scale most commonly used in psychological literature derivative of Conversation Analysis (e.g. Dingemanse et al., 2015; Dideriksen et al., 2020). From least to most helpful, the three levels are: 1. Open Request (e.g. huh?), 2. Specific Request (e.g. to the right or left?), and 3. Specific Offer (e.g. to the lake?). In contrast to the expanded scale, the three level helpfulness measure is not sensitive to partial repeats. Consequently, partial repeats are almost always coded as specific offers in this measure.

**Operationalizing Reformulation Effort: A New Method**

How can we conceptualize reformulation? First, as an effortful divergence from the norm. This norm is lexical entrainment; conversants seek to develop conceptual pacts wherein concepts and referents are always described in the same way. For this reason, while variability of word-choice is high between conversations, it is relatively low within a conversation (Brennan & Clark, 1996; Garrod & Anderson, 1987). Achieving lexical entrainment is a process; when two people repeatedly discuss the same object, they come to use the same terms (Brennan, 1996). Priming and memory play a primary role in this process, as demonstrated by the fact that people show greatest lexical alignment directly after a term is made salient with an overt reformulation by the interlocutor (Brennan, 1996). Thus it seems that reformulation in other-initiated repair is both effortful (in that it diverges from the automatic norm of lexical entrainment) and purposeful (in that it compels the interlocutor to adopt the newly formulated conceptualization).

Reformulation is also, of course, intimately connected with the fundamental cognitive difficulty in language production in general. One prominent model of language production that captures this cognitive demand was proposed by Levelt (1989, 1999). His model consists of three main stages: conceptualization, formulation, and articulation, across which a non-linguistic conceptual representation is elaborated lexically, syntactically, and phonologically to produce an utterance (Bock & Levelt, 1994; Ferreira & Engelhardt, 2006). By separating conceptualization and formulation, this model provides one important guideline for operationalizing reformulation effort: utterances that introduce new conceptual frameworks for information entail greater effort than those that only differ lexically. Using this framework, one might rate reformulation effort on a three point scale: 1. identical repeat, 2. lexical difference, 3. conceptual difference. Consider for example the following repair initiations.

| | | |
|---|---|---|
| 1) | G | *pass pass below the ghost town* |
| | F | *right okay* |
| | G | *but above the carved wooden pole and above the stone creek* |
| | F | *right okay* |

G *right*

**F** ***so between the ghost town and sort of like a forty-five degree line or something***

G *so you're just*

G *yeah*

2) G *and down below the springs*

**F** ***walk walk below the springs***

G *uh-huh*

The first of these exchanges introduces a new conceptual framework (“between”; “forty-five degree”), while the second says essentially the same thing as its referent, but with a different word (“walk”).

In the particular context of repair, an additional operationalization of reformulation can be gleaned from Clark & Wilkes-Gibbs (1986), who distinguish between "expansion" and "replacement". Expansion means that the initiator adds content to the interlocutor's noun phrase as a request for confirmation, while replacement means that the initiator substitutes the interlocutor's noun phrase with another one.

Though the above categorizations seem to be good first steps toward an operationalization of reformulation in repair that is rooted in prominent models of cognitive effort in language production, I have opted in the present study to use a more loosely defined, qualitative measure of reformulation. This is because priming is complex, and lexical or syntactic similarities with any of the preceding context—not just the trouble source turn—are likely to impact processing. In the HCRC Map Task Corpus, which will be used in this study, many repair initiations combine verbatim elements from multiple previous turns into a single syntactic structure, making it difficult to clearly define the boundary between a repeat and expansion or between lexical and conceptual difference; most cases are an amalgamation. In fact, many initiations include partial repeats of the utterances of the initiator rather than the interlocutor. These too should be considered as having low reformulation, as at the time that the "decision" of whether or not to initiate repair is being made, the cognitive effort has already been invested and the cost will be low. For these reasons, reformulation will be evaluated in this study on a five point scale representing the similarity between the repair initiation turn and the preceding turns ("Reformulation Rating"). The reformulation rating scale ranged from 1 (verbatim partial repeat) to 5 (no similarity). Roughly speaking, 2 was awarded only to initiations with minor lexical differences, whereas 4 and 5 were only awarded to those with significant conceptual contributions beyond that which was previously uttered.

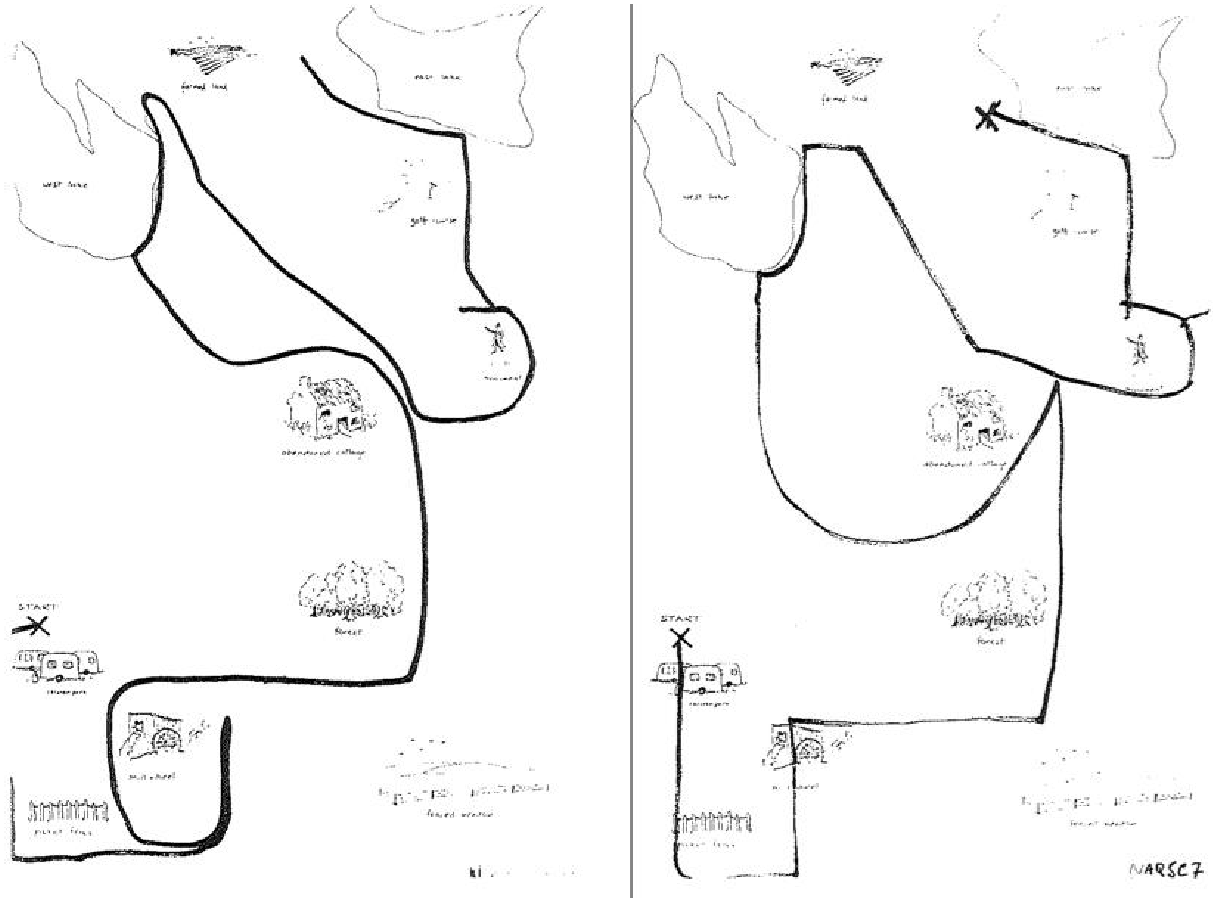
**Method**

The data in this study are taken from the HCRC Map Task Corpus (Anderson et al., 1991). This corpus consists of one hundred and twenty-eight dialogues between 64 students of the University of Glasgow, Scotland (32 male and 32 female), carrying out the map task (Brown et al., 1984). Participants were recruited with a familiar partner, who they had known for an average of about two years (Boyle et al., 1994). Participants' ages ranged from 17 to 30 years with a mean of 20 years. In the task, participants were randomly assigned to play the role of either Instruction Giver or Instruction Follower. Givers were given a map with labelled landmarks and a route, represented by a curving line. Followers were given a map similar to that of the giver, but with slightly different (e.g. missing, extra, differently named) landmarks and no route. The giver's task was to describe the route so that the instruction follower could draw it on his own map. Both subjects were informed of the slight inconsistency between their maps and encouraged to talk freely and ask questions if there was anything they were not sure about. The maps each followed one of four different basic plans designed to provide routes of roughly equal difficulty. Likewise, though discrepancy between the giver's and the follower's map varied at different points along the path, all maps had the same amount of total discrepancy.

Each participant in a quad took part in four conversations, twice as Giver and twice as Follower, once in each case with their familiar partner and once with an unfamiliar partner. Fuller details of the design are given in Anderson et al. (1991).

**Figure 3**

*Two follower maps from the HCRC Map Task Corpus.*




The HCRC Map Task corpus was chosen for five reasons. First, the task has a clear goal which is informational and shared among all participants. Second, the transfer of task-relevant information is entirely unidirectional, from the giver to the follower. This creates a controlled environment for investigating other-initiated repair, as the follower is nearly always the repair initiator, while the giver is the utterer of the trouble source. Third, the map task provides a well-defined measure of communicative success: the extent to which the instruction follower's route matches the instruction giver's route. This degree of fit is measured in square centimeters between the two routes, such that a larger number represents less successful communication. Fourth, the corpus includes full orthographic transcriptions of all 128 conversations. This makes organizing data much easier, and allows coding to be easily re-checked. Lastly, each participant acted as Follower twice, each time with a different giver. This allows us to make inferences about the role of individual cognitive abilities in repair behavior or communicative success. If reformulation ability and willingness are indeed stable individual traits, we can expect to see greatest consistency of repair behavior within trials with the same follower.

***Independent Variables***

I collected data by listening to recorded clips of conversations from the Map Task Corpus while looking at the corresponding transcript. To minimize concerns of data mismanagement, I initially listed each incident of repair initiation within its immediate context in a word processor document. I then recorded scores for each newly coded variable alongside the text of its corresponding repair initiation. The document was ordered by appearance of the sequence in the corpus and was checked against an SPSS spreadsheet regularly. Each time a repair initiation was uttered by the follower, I recorded the following four dimensions of the repair sequence, each either designed to be or commonly assumed to be analogous to cognitive effort, as discussed above.

- *Reformulation rating* (1-5)
- *Orthographic length*
- *Helpfulness* (Expanded Scale)
- *Helpfulness* (Three Levels).

The above four variables were then averaged within each conversation and the means were entered into a second spreadsheet.

***Dependent Variables***

- *Repair Avoidance Index*, as described above
- *Deviation Score.* The area—in square centimeters—between the route given to the giver and that drawn by the follower is referred to as the deviation score. As described by Reitter & Moore (2007; 2014) and others, this is an inverse measurement of communicative success; the lower the deviation score between the two routes, the more effective the pair's communication was during the task.

**Table 1**

*Descriptive Statistics for Variables in Study 1*

| Variable | Mean | SD | Min | Max |
|---|---|---|---|---|
| Reformulation Rating (by Initiation) | 3.71 | 1.32 | 1 | 5 |
| Reformulation Rating (Follower Avg.) | 3.58 | .51 | 2.30 | 4.31 |
| Orthographic Length (by Initiation) | 40.09 | 27.29 | 5 | 234 |
| Orthographic Length (Follower Avg.) | 39.71 | 8.52 | 25.93 | 56.92 |
| Helpfulness (Six Levels; by Initiation) | 4.60 | .82 | 0 | 5 |
| Helpfulness (Six Levels; Follower Avg.) | 4.55 | 0.19 | 4.18 | 4.88 |
| Helpfulness (Three Levels; by Initiation) | 1.79 | .43 | 0 | 2 |
| Helpfulness (Three Levels; Follower Avg.) | 1.80 | .144 | 1.41 | 2.00 |
| Repair Avoidance Index | .77 | .10 | .55 | .91 |
| Deviation Score | 90.91 | 56.02 | 11 | 227 |

***Hypotheses***

*The Individual Effort Hypothesis:* Within followers, greater average repair initiation effort will predict lower repair avoidance indices. This effect will be strongest for reformulation rating, which most directly measures cognitive effort.

*The Effective Communication Hypothesis:* Within followers, greater average repair initiation effort will predict greater communicative success as measured by average deviation score on the map task.

## Analysis and Results

### *Reliability of Measures*

Bivariate correlations were calculated among all four measures of repair initiation effort, with individual initiations as cases. Subjective reformulation rating, orthographic length of repair initiations, and the six level helpfulness scale were all found to correlate strongly with each other (Table 2). Unsurprisingly, the six level and three level helpfulness scales were also very strongly correlated. Surprisingly, however, the three level helpfulness scale (used in Dideriksen et al., 2020 and others) was not found to significantly correlate with reformulation rating or orthographic length at all (see Table 2).

**Table 2**

*Pairwise Correlations Among Measures of Repair Initiation Effort*

| Variable | Reformulation Rating | Orthographic Length | Helpfulness (Six Levels) | Helpfulness (Three Levels) |
|---|---|---|---|---|
| Reformulation Rating | 1 | **.51***** N=342 | **.34***** N=342 | **-.03** N=342 |
| Orthographic Length | **.51***** N=342 | 1 | **.30***** N=342 | **-.05** N=342 |
| Helpfulness | **.34***** | **.30***** | 1 | **.76***** |

| | | | | |
|---|---|---|---|---|
| (Six Levels) | N=342 | N=342 | | N=531 |
| Helpfulness (Three Levels) | **-.03** N=342 | **-.05** N=342 | **.76*****<br>N=531 | 1 |

***Note.*** Because some variables are ordinal and none are normally distributed, correlations are expressed with Spearman's ρ. Three asterisks indicate that a correlation is significant at the p<.001 level (2-tailed).

These associations can be observed even more dramatically when data are aggregated within each follower, as shown in Table 3.

**Table 3**

*Pairwise Correlations Among Measures of Repair Initiation Effort, Averaged Within Followers*

| Variable | Reformulation Rating | Orthographic Length | Helpfulness (Six Levels) | Helpfulness (Three Levels) |
|---|---|---|---|---|
| Reformulation Rating | 1 | **.63*** N=12 | **.76**** N=12 | **-.36** N=12 |
| Orthographic Length | **.63*** N=12 | 1 | **.55**$^{\dagger}$ N=12 | **-.14** N=12 |
| Helpfulness (Six Levels) | **.76**** N=12 | **.55**$^{\dagger}$ N=12 | 1 | **.49**$^{\dagger}$ N=16 |
| Helpfulness (Three Levels) | **-.36** N=12 | **-.14** N=12 | **.49**$^{\dagger}$ N=16 | 1 |

*Note.* Data were first aggregated by taking the mean of each variable within conversations, so that averages were not biased by the conversation length. These conversation-level means were then averaged by follower. Aggregate scores for a given follower are therefore not averages of all repair initiations uttered by that follower, but averages of average scores for the two conversations in which that follower participated. Correlations are expressed with Pearson's *r*. One asterisk indicates that correlation is significant at the p<.01 level (2-tailed). A superscript dagger indicates a marginally significant correlation (p<.10).

One-way ANOVAs demonstrate the relative consistency of results within followers (Table 4). This further supports the understanding that these measures reflect stable individual traits.

**Table 4**

*One-way ANOVAs Show Consistency of Measures Within Followers*

| | Reformulation Rating | Orthographic Length | Helpfulness (Six Levels) | Helpfulness (Three Levels) | Repair Avoidance Index |
|---|---|---|---|---|---|
| $F_{Between\ Followers}$ | **2.68**$^{\dagger}$ | **2.42**$^{\dagger}$ | **1.14** | **3.43*** | **2.18*** |
| $F_{Between\ Givers}$ | .87 | .63 | .55 | 1.63 | **2.56*** |

*Note*. Variables are averaged by conversation. Each participant acted as Giver twice and follower twice. Because repair avoidance index is largely coded computationally, it was possible to analyze a larger sample (N=48 conversations).

### ***The Individual Effort Hypothesis***

To test whether greater follower average repair initiation effort predicts lower repair avoidance indices within conversations, four separate simple linear regression models were fit—one for each independent variable. The results of this analysis are shown in Table 5.

**Table 5**

*Standardized Coefficients of Simple Linear Regression Models Predicting Repair Avoidance*

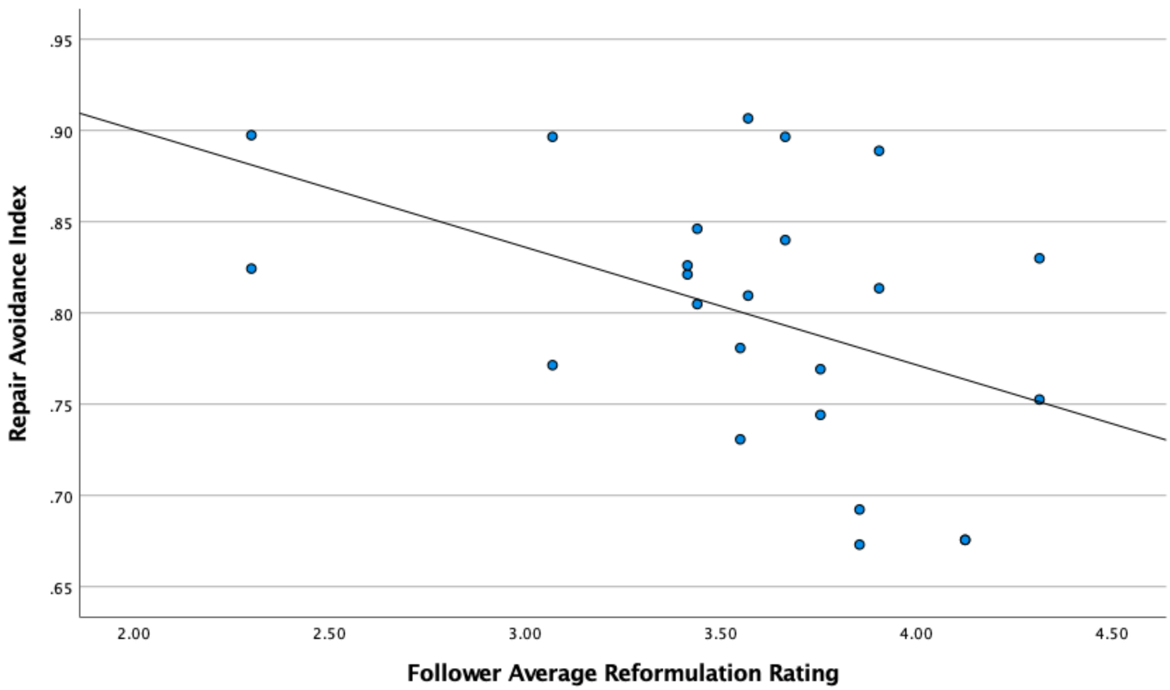
| Predictor | Standardized Beta | Sample Size |
|---|---|---|
| Follower Average Reformulation Rating | **-.45*** | N=24 |
| Follower Average Orthographic Length | **-.05** | N=24 |
| Follower Average Helpfulness (Six Levels) | **-.15** | N=32 |
| Follower Average Helpfulness (Three Levels) | **-.37*** | N=32 |

***Note***. Independent variables are averaged within followers, while the dependent variable is averaged within conversations. Each follower participated in two conversations.

All model slopes were in the direction predicted by the Individual Effort Hypothesis: followers displaying higher average reformulation rating, orthographic length, or helpfulness of repair initiations avoided repair less frequently. The Individual Effort Hypothesis also predicted that this effect would be strongest for reformulation rating, which most directly measured cognitive effort (see above sections on operationalizing effort in repair). This was also supported by the analysis. Figure 4 shows a scatterplot of repair avoidance by follower average reformulation rating.

**Figure 4**

*Follower Average Reformulation Rating of Repair Initiations Predicts Repair Avoidance*




***The Effective Communication Hypothesis***

As with the Individual Effort Hypothesis, four separate simple linear regression models were fit—one for each independent variable—now predicting map task deviation score (i.e. the inverse of communicative faithfulness). None of the effects approached significance. These data therefore offer no support for the Effective Communication Hypothesis: followers displaying higher average reformulation rating, orthographic length, or helpfulness of repair initiations did not perform better in the map task, though they did initiate repair more frequently. In fact, there was also no significant correlation between repair avoidance and deviation score ($r$=.16, p=.38, N=32).

**Discussion**

Though the sample size was relatively small, results of Study 1 lend considerable support to the analysis of cognitive effort in repair presented earlier in this thesis. First, the extent to which a repair initiation reformulates previous utterances seems likely to be the best indicator of cognitive effort expended in generating that initiation. A six level helpfulness scale similar to that proposed by Albert and de Ruiter (2018) correlates strongly with level of reformulation. Orthographic length, as utilized by Dingemanse et al. (2015) also shows moderate correlation with both of these measures. On the other hand, a three level helpfulness scale as used by Dideriksen et al. (2020) and others does not seem to correlate with these measures at all. This may indicate that it is too crude a measure to accurately capture cognitive effort expended in repair initiation. The failure of the three level helpfulness scale to register the distinction between partial repeats and fully reformulated offers may be particularly responsible for its lack of correlation with other measures.

The above described correlations between measures of cognitive effort on the level of individual repair initiations hold true when averaged within individual participants as well. Furthermore, all four measures are relatively consistent within individual participants. This consistency constitutes critical evidence for the validity of the constructs involved in the Individual Effort Hypothesis. It seems likely that the expense of cognitive effort in repair is contingent on individual differences as well as situational ones. Though this study is concerned with individual differences in repair initiators, it is important to emphasize the dyadic nature of conversation. This can be seen especially in the observation that, while repair avoidance is relatively consistent within followers, it is also relatively consistent within givers. This may reflect differences in the way givers structure their instructions, or differences in the initial clarity of those instructions, leading some givers to require more repair than others in order to reach sufficient mutual understanding.

Results from Study 1 lend considerable support to the Individual Effort Hypothesis: Within followers, greater average repair initiation effort (as measured by reformulation rating, orthographic length, or helpfulness) does predict lower repair avoidance indices. This effect does seem to be strongest for reformulation rating, which most directly measures cognitive effort.

Thus far, the results of Study 1 consistently support the cost-benefit model of repair avoidance outlined in the “Background” section. In particular, cognitive effort has been reasonably implicated as a cost driving people to initiate repair less frequently. On the other hand, the benefit associated with the cost-benefit model remains elusive: less repair avoidance does not seem to be associated with greater communicative success as measured by average deviation score on the Map Task. This replicates Dideriksen et al.’s (2020) finding that repairs were mostly unrelated to performance in the Map Task (in another corpus). As noted by Dideriksen et al., this may indicate a pattern in which dyads flexibly modulate their conversational structure in a nuanced way according to the contextual demands. Perhaps factors other than repair are more important in determining success in the Map Task. Future research should investigate different conversational contexts and different measures of communicative success.

### *Limitations*

The current study presents some limitations. First and foremost, the primary independent variable under study, reformulation rating, was coded subjectively and was not clearly defined. Though I took steps to focus on individual repair initiations without noticing the total number of initiations in a conversation (which partially defines its repair avoidance index), this design may be subject to coding bias. In fact, these steps to avoid bias may even have contributed to it, as coding of reformulation rating was done after all repair initiations had been identified and extracted from the transcript. This likely led to some cases in which some context that would have impacted ratings was left out of the excerpted initiations and missed. Furthermore, because coding was done in multiple sittings, sometimes separated by a week or more, it may be that my criteria changed from one sitting to another.

Even if my criteria for rating reformulation were consistent, they did not account for time elapsed or interfering utterances between trouble source and repair initiation. These factors are known to impact priming and working memory, and likely therefore impact cognitive effort involved in utterance formulation (Frings et al., 2015). Likewise it does not account directly for word length, which is known to impact the capacity of the phonological loop which stores auditory information in the short term (Baddeley & Hitch, 1974). A better operational definition of cognitive effort in repair initiation would account for these and other factors recognized by the literature.

Additional concerns emerged from the nature of the Map Task Corpus itself. First, because of the abundance of written labels in the map task (which were sitting right in front of participants' eyes; see Figure 3) many initiations identified as highly reformulated (because of references to landmark names not yet mentioned in the conversation) may not have been so cognitively difficult. Furthermore, the prevalence of Standard Scottish English in the corpus (see Anderson et al., 1991) may have biased my ratings as a native speaker of American English.

Though the definition of repair initiation was much more clearly delineated than that of reformulation (see Appendix B), the operationalization of repair avoidance may also be imprecise. In particular, it is possible that higher repair frequency represents greater chunking of individual repairs. Because multiple initiations in a single repair sequence were counted separately, the current study cannot distinguish between many small repair sequences and a few lengthy ones. Future research should therefore compare differential effects of coding of one repair sequence as one initiation.

Lastly, Study 1 was deeply exploratory in that four separate independent variables were tested and the dependent variable was not fully operationalized before data analysis began. The exploratory nature of the study renders *p* values nearly useless in assessing the true extent to which results support my hypotheses. With the constructs and hypotheses now more clearly defined and understood, future studies can provide more rigorous answers.

Most of the limitations outlined above resulted from the subjective approach to coding, which both increased risk of biased data and necessitated a small sample size due to the time constraints on the production of this thesis of coding by hand. Study 2 will address many of these limitations—and one additional limitation not yet acknowledged—by developing a computational approach to measuring reformulation in dialogue.

## Study 2: Generalized Propensity to Reformulate

In Study 1, I analyzed four independent variables, each aiming to measure the ability and/or willingness of a given participant to reformulate utterances while acting as Follower. This, in theory, is a general ability or willingness, not confined to repair initiations (see "Why People Avoid Asking for Clarification" above). The coding paradigm in Study 1, which was limited to repair initiations for all four independent variables, was therefore flawed. One major limitation was the disregard for those instances in which followers took control entirely, without reference to a previous instruction by the giver, as in the following exchange.

G just go a couple of steps

F mmhmm

**F have i to pass pass by the waterhole**

G pass pass the pass

G well

G it's stony desert that's what i've got here

F erm

G but

F now

F i'm starting off from the s-- just above to the north of the stony desert

G mmhmm

F erm

**F have i to pass by the savannah**

G no

Strictly speaking, the two bolded utterances are not repair initiations because they are not aiming to clarify any previous utterance; the follower in this conversation is merely guessing at her route. This behavior is a sort of hyper-repair—aiming to clarify the giver's instructions before they are even uttered (notice that both of the bolded utterances introduce a new reference framework, and both are specific offers). Because such hyper-repair utterances were not coded as repair initiations following criteria in Study 1, their reformulation rating, orthographic length, and helpfulness were not recorded. Thus it is very likely that this follower's average reformulation rating, orthographic length, and helpfulness as recorded for their repair initiations were not representative of their behavior in general. Indeed, after excluding two conversations in which this behavior appeared especially often, the pairwise correlation between follower average character length and frequency of repair rose from .03 to .32, and the correlation between follower average helpfulness (six level scale) and frequency of repair rose from .14 to .58 ($p=.005$). It thus became clear that an alternative operationalization of reformulation willingness was needed—one that took into account all utterances in a conversation.

Such an algorithm would perform an approximate calculation of lexical priming effects to measure the extent to which any given utterance is made effortless by its similarity to previous utterances. Within participants, lower average similarity between utterances and preceding speech in a conversation could then be construed as greater average reformulation of previous material. Just as in Study 1, I hypothesize that followers with higher average reformulation of previous material will also initiate repair more often.

This approach to the question of cognitive effort in repair resolves many of the concerns with the validity of Study 1. Most straightforwardly, it minimizes the possibility for coding bias or subjective inconsistencies in coding. With a fully algorithmic operationalization of the independent variable (reformulation) the only remaining subjectivity lies in the counting of repair initiations in each conversation, which determines the denominator of the dependent variable (repair avoidance). Even this subjectivity is negligible; the definition of repair initiation is clearly delineated (see Appendix C) and bias is minimized with an algorithmically determined numerator (see "Operationalizing Repair Avoidance" above). Furthermore, the algorithmic approach enables us to account for peculiarities of the Map Task Corpus, rather than relying on my sensibilities as an American English speaker (how it does this will be explained below). The primarily algorithmic approach also makes coding much less time consuming. For this reason, Study 2 will analyze a considerably larger sample than Study 1. Finally, Study 2 can take a more confirmatory stance than Study 1; the dependent variable and the hypothesis are already defined, and the independent variable is newly developed based on theoretical criteria. This simplicity allows me to specify a method of analysis in advance and compute a *p* value for the extent to which the null hypothesis should be rejected.

**Computational Approach to Measuring Reformulation in Dialogue**

Using the Python programming language, I developed an algorithm to calculate the average lexical similarity between a participant's utterances and those uttered previously in the conversation[15]. The source code for this algorithm is available in Appendix A[16]. Using transcripts, the algorithm first computes term frequency–inverse document frequency (TF-IDF) scores for each word in each utterance of the entire HCRC map task corpus. TF-IDF is a commonly used statistic in natural language processing that weighs each term in an utterance by its frequency within that utterance (TF), and by its inverse log-scaled

---

[15] Thank you to my brother, Walter, for devoting many hours to helping with the technicalities of programming.

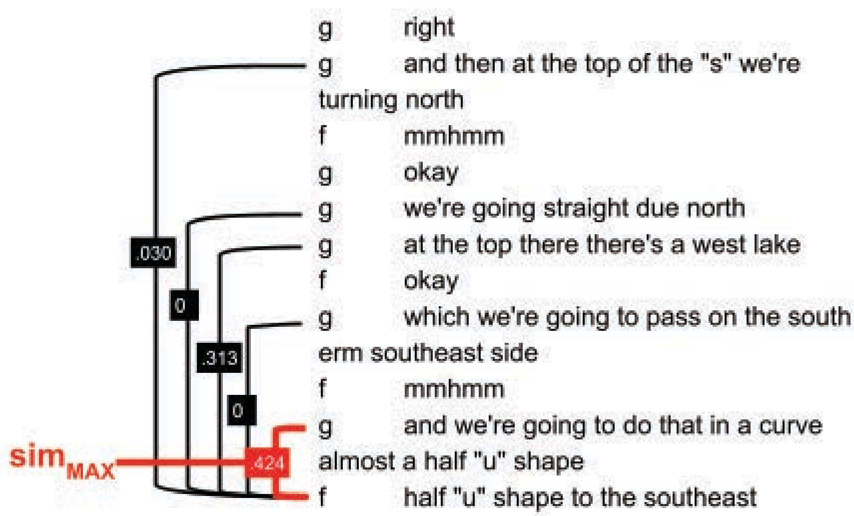
[16] An equivalent implementation of the algorithm written in the R programming language is available upon request from the author.

frequency in the entire corpus (IDF), such that the most frequent terms (e.g. and, the, it) in the corpus are devalued, and the rarest terms are given greater weight (Spärck Jones, 1972; see Achananuparp et al., 2008 for a review of this and similar methods). This weighting roughly parallels lexical priming effects in spoken language, which are stronger for lower frequency words (Pace-Sigge & Patterson, 2017). This frequency effect on lexical access follows a roughly logarithmic pattern (Murray & Forster, 2004). TF-IDF was computed within the Map Task Corpus rather than a larger, generalized corpus of English language to account for effects of context (see Coane & Balota, 2010), regional dialect (primarily Standard Scottish English; see Anderson et al., 1991), and visual priming of words in map labels (see Figure 3; labeled landmark names appear more frequently in dialogue and are therefore attributed less importance, irrespective of their prevalence in the English language generally).

Using TF-IDF cosine similarity (a commonly used document similarity metric that eliminates effects of sentence length; Achananuparp et al., 2008), the algorithm compares the lexical similarity of each utterance with each of the preceding utterances in the conversation. Whichever similarity score is highest within these is taken as the similarity score for that utterance ($sim_{MAX}$). $Sim_{MAX}$ therefore reflects the extent to which a given utterance is a low-effort reiteration of a preceding utterance in the conversation (Figure 5).

**Figure 5**

*Sentence Similarity Measurements used in Computing $Sim_{MAX}$ for a Follower Utterance*




*Note.* Numbers in boxes represent cosine similarities between utterances.

The mean of all $sim_{MAX}$ scores computed for utterances of a given participant in a conversation (i.e. the giver or the follower) is considered the average similarity score of that conversation.

**Method**

Unlike the four measures of cognitive effort used in Study 1, which are specific to follower repair behavior, average similarity can be computed for both participants in a conversation—Giver and Follower. Of course, because the giver's role involves introducing new information to the follower, the giver's average similarity score is likely to be lower. Nevertheless, the opportunity to measure participants' behavior when acting in both roles allows for greater generalizability to stable individual differences (since each individual takes turns as both Giver and Follower). For this reason, aggregate similarity for a participant is defined as the average of scores from four conversations: two in which the participant acted as Follower, and two in which they acted as Giver.

As in Study 1, I hypothesize that there will be a positive association between participant aggregate similarity and the repair avoidance of that participant when she is acting as Follower, such that those individuals whose utterances are less similar to preceding utterances (i.e. those who reformulate more) will avoid repair less often, and vice versa.

The extent to which a giver reiterates his instructions may affect the follower's need to ask for clarification. To control for this, the effect of individual differences in reformulation ability on repair avoidance will be assessed with multiple regression:

$$\textit{Repair Avoidance Index} = \beta_0 + \beta_1(\textit{Aggregate Similarity}_F) + \beta_2(\textit{Average Similarity}_G) + \varepsilon$$

The null hypothesis will be rejected if the effect of Follower aggregate similarity on repair avoidance when controlling for Giver average similarity ($\beta_1$) is positive with a *p* value less than .05[17].

**Table 6**

*Descriptive Statistics for Variables in Study 2*

| Variable | Mean | SD | Min | Max |
|---|---|---|---|---|
| Follower Similarity (Conversation Avg.) | .51 | .08 | .29 | .70 |
| Participant Aggregate Similarity | .62 | .04 | .55 | .70 |
| Repair Avoidance Index | .75 | .09 | .55 | .91 |

**Results**

Participant aggregate similarity significantly predicted repair avoidance ($\beta_1$=.41, p=.001; see Table 7). The null hypothesis is therefore rejected.

[17] I will sometimes refer to "participant aggregate similarity" as "Follower aggregate similarity" to make clear that "Giver average similarity" is referring to the measurement of a different person.

**Table 7**

*Coefficients of Multiple Regression Predicting Repair Avoidance*

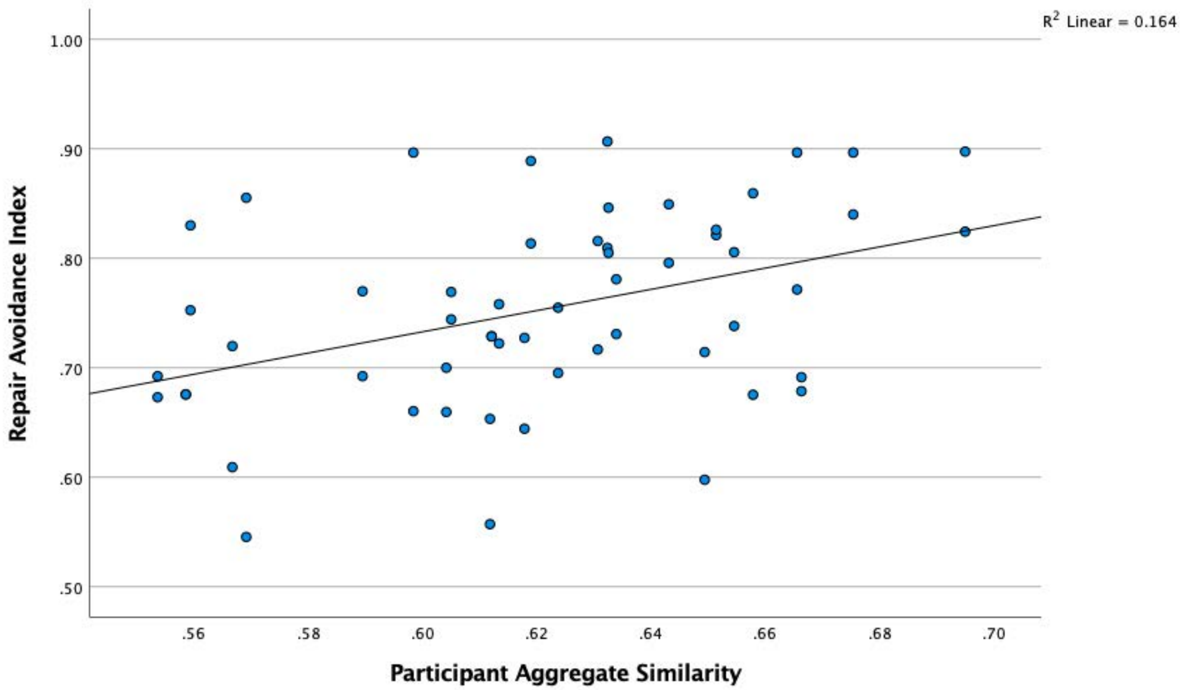
| Predictor | Standardized Beta | Sig. |
|---|---|---|
| Follower Aggregate Similarity | **.41** | .001 |
| Giver Similarity (Conversation Average) | **.23** | .030 |

*Note.* Dependent Variable: Repair Avoidance Index. N=56

Giver similarity, as averaged within each conversation, also significantly predicted follower repair avoidance, but there was no evidence of multicollinearity (tolerance = 1.00). There was also no significant correlation between Follower aggregate similarity and number of Giver instructions (the numerator of repair avoidance index; $r$=-.06, $p$=.66). Figure 6 shows the relationship between Follower aggregate similarity and repair avoidance without controlling for giver similarity.

**Figure 6**

*Participant Aggregate Similarity Predicts Repair Avoidance*




*Note.* An F test yielded no significant evidence of heteroskedasticity (F=.32 , *p*=.58).

**Post hoc Analysis**

Though the above results seem to support the Individual Effort Hypothesis—that individual differences in cognitive ability (or need for cognition) predict frequency of repair avoidance—the complexity of the aggregate similarity measure may be cause for skepticism. In particular, the averaging of $sim_{MAX}$ scores within participants makes it impossible to discern which type of utterances have the most effect on the measure. Is the variability in aggregate similarity mostly due to frequency of very common utterances (which are likely to be identical to an earlier utterance) or is it mostly due to utterances with very low similarity scores?

In order to better understand the results of Study 2, I conducted a series of post hoc analyses in which I slightly altered the utterance similarity algorithm and reassessed the effect of participant aggregate similarity on repair avoidance. First, I ran the algorithm with a preprocessing step which discounted common backchannels (short utterances that affirm that the listener is paying attention, e.g. “uh-huh”) from all analyses[18]. Because these utterances are common and formulaic (see Dideriksen et al., 2020), they are likely to occur many times in a conversation, and are therefore likely to have $sim_{MAX}$ scores of 1.0 (i.e. identical to a previous utterance in the conversation). Next, I ran the algorithm with a preprocessing step that discounted a much longer list of common utterances (identical to the list of utterances discounted in computing repair avoidance index; see note 12). For each of these two altered versions of the algorithm, follower similarity scores were computed as before and regression models were fit predicting repair avoidance. Table 8 shows the results of these analyses.

**Table 8**

*The Effect Disappears when Common Utterances are Discounted from Processing*

| Predictor | Standardized Beta | Sig. |
|---|---|---|
| Participant Aggregate Similarity (No Preprocessing) | **.41** | .002 |
| Participant Aggregate Similarity (Backchannels Removed) | .07 | .615 |
| Participant Aggregate Similarity (All Common Utterances Removed) | .06 | .658 |

[18] Utterances discounted for this purpose were as follows: "mmhmm", "yeah", "okay", "okay yeah", "right okay", "right", "uh-huh", "right-fine", "uh-huh yeah", "yep", "okay then".

*Note.* Dependent variable: Repair Avoidance Index. Coefficients are for simple linear regression models.

The effect of participant aggregate similarity on repair avoidance was severely mitigated when common utterances were discounted. This indicates that the effect observed in the Results section above was primarily due to variability in the frequency of very common utterances such as backchannels. In other words, participants who uttered more formulaic utterances (as a proportion of their total utterances) were less likely to initiate repair.

The huge drop in effect size after the removal of backchannels from the analysis raises a serious challenge to the internal validity of this study: Perhaps aggregate similarity is not a measure of expended cognitive effort but a measure of understanding. In other words, perhaps people who frequently indicate their comprehension with backchannels have less need for clarification questions because they can understand instructions without them! Three pieces of evidence make this possibility unlikely: First, follower similarity (conversation average) was seemingly unrelated to success in the Map Task ($r$=-.06, p=.67). If average similarity indicates the extent to which conversants indicate their comprehension of instructions, this indication has no significant relation to true comprehension. Second, Dideriksen et al. (2020) analyzed the relationship between various grounding mechanisms in six different conversational contexts (including the map task) and found no significant relationship between frequencies of backchannels and repair[19]. Third, the effect of participant

[19] Due to time constraints on the production of this thesis, I cannot report the relative frequencies of backchannels and repair for the set of conversations analyzed in this study. While common, formulaic backchannels such as the ones listed in Note 18 can be coded algorithmically, others may entail novel phrasing or partial repetition of preceding utterances.

average similarity on repair avoidance is evident even when the participant average score only includes conversations in which the participant acted as Giver, as shown in Table 9.

**Table 9**

*The Average of a Participant's Similarity Scores When Acting as Giver Predicts Their Repair Avoidance When Acting as Follower*

| Predictor | Standardized Beta | Sig. |
|---|---|---|
| Follower Average Similarity as Giver | **.22** | .09 |
| Giver Similarity (Conversation Average) | **.27** | .04 |

*Note.* Dependent Variable: Repair Avoidance Index.

If the relevant individual differences are seen even when an individual is giving information rather than receiving it, it seems unlikely that the effect is one of comprehension. Together, these pieces of evidence suggest that the effect observed in this study is not a result of the interaction between backchannels and repair.

If aggregate similarity does not reflect actual or perceived comprehension, how can we explain the disappearance of its effect on repair avoidance when common utterances are discounted? First of all, an individual's propensity to avoid exerting cognitive effort in utterance formulation, which manifests as repair avoidance, may well also manifest itself as a tendency to utter more formulaic and fewer original utterances. Because such formulaic utterances are very common, very short, and carry very little semantic information, they are likely to require very little cognitive effort to produce (see "Operationalizing Effort in Repair" above). Second, TF-IDF similarity scores may be too crude a measure to accurately reflect cognitive effort in more complex utterances. To further investigate this possibility, I developed an integrated measure of cognitive effort that accounts for both orthographic length (demonstrated above to be a useful proxy for cognitive effort; see "Operationalizing Effort in Repair") and the number of turns elapsed since the most similar previous utterance (see Sinclair et al., 2004). The reason for the latter is that utterances whose most similar previous utterance occurred many turns ago are likely to benefit less from priming and therefore are likely more effortful. The "effort" of an individual utterance was therefore defined as follows:

$$Effort = (1 - sim_{MAX}) \times \ln(Turns\ since\ sim_{MAX}) \times \ln(Orthographic\ length\ of\ turn)$$

Effort scores were then aggregated in the same manner as similarity scores above. The source code for this algorithm is available in Appendix B. The results of multiple regression analysis for participant aggregate effort predicting repair avoidance are shown in Table 10.

**Table 10**

*Integrated "Effort" Measure Predicts Repair Avoidance Better than Similarity Alone*

| Model | Predictor | Standardized Beta | Sig. |
|---|---|---|---|
| Model 1 | Follower Aggregate Effort (No Preprocessing) | **-.50** | <.001 |
| | Giver Similarity (Conversation Average) | **.32** | .007 |
| Model 2 | Follower Aggregate Effort (All Common Utterances Removed) | **-.39** | .003 |
| | Giver Similarity (Conversation Average) | **.31** | .013 |

*Note.* Giver similarity is computed without preprocessing, as in Table 9, for maximum comparability with the original result. In both models, an F test yielded some significant evidence of heteroskedasticity (F=3.03 , $p$=.09; F=3.26, $p$=.08). Parameter estimates with robust standard errors (HC3 method; Long & Ervin, 2000) were still significant to $p$<.01.

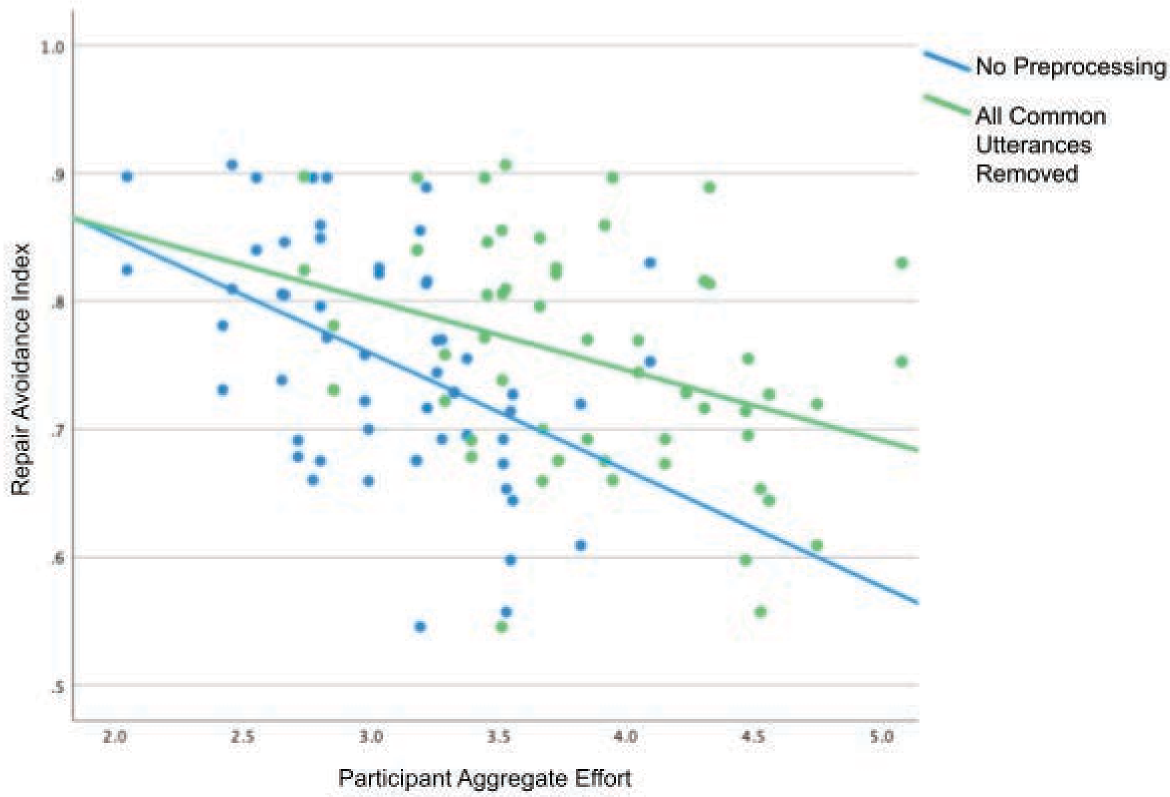
Controlling again for giver similarity[20], the integrated "Effort" measure predicts repair avoidance much better than TF-IDF similarity scores alone; the standardized coefficient of multiple regression rose[21] from .41 to -.50 (see Tables 5 and 7). Even more notably, aggregate effort predicted repair avoidance even when all common utterances were removed from processing (Table 10). This result provides additional support for the Individual Effort Hypothesis originally formulated in Study 1: Individuals with a higher propensity to expend cognitive effort in formulating utterances are more likely to initiate repair. Conversely, individuals who favor less effortful utterances tend to avoid repair of potential misunderstandings. This effect holds true even if most or all short, formulaic utterances are discounted.

**Figure 7**

*Integrated Effort Measure Predicts Repair Avoidance Even When Formulaic Utterances are Discounted*

[20] Effects are still significant or marginally so if the control variable is Giver Effort, with the amount of preprocessing corresponding to the independent variable ($\beta_{\text{No Preprocessing}}$=-.43, $p$=.001; $\beta_{\text{All Common Utterances Removed}}$=-.23, $p$=.079).

[21] Note that higher similarity corresponds with lower effort.




**Discussion**

Building on and improving Study 1, Study 2 yielded additional evidence for the Individual Effort Hypothesis. The propensity to expend effort in utterance production is a stable individual trait. This trait is evident not just in repair initiation when one is acting as Follower (as shown in Study 1), but in all kinds of utterances, even while acting as Giver. participant aggregate similarity (the algorithmically generated proxy for a participant's propensity to avoid expending reformulation effort) is strongly correlated with follower average reformulation rating (the subjectively rated measure of repair initiations used in Study 1; $r$=-.82, $p$<.001, N=24). Like the latter measure, participant aggregate similarity predicts repair avoidance when the participant is acting as Follower. In fact, standardized coefficients for the regression analyses conducted in Studies 1 and 2 were strikingly similar ($\beta_{\text{Reformulation Rating}}$=-.45; $\beta_{\text{Similarity}}$=.41). These results lend considerable support for the internal validity of Study 1 (i.e. that the subjective reformulation rating was in fact measuring effort-spending) and for the Individual Effort Hypothesis.

Upon further analysis, it became apparent that much of the effect of participant aggregate similarity on repair avoidance was attributable to formulaic utterances such as “yeah”, which are likely to have been previously uttered in a conversation. As I have argued in the section above, this does not present a challenge to the Individual Effort Hypothesis, but rather an insight into the nature of effort-spending in conversation. It seems that those people who are generally less willing to expend effort in conversation primarily express this trait as a propensity to utter short, formulaic utterances rather than longer, more original ones. This is consistent with the grounding theory classically expressed by Clark and Brennan (1991), which identifies two phases within each contribution to a conversation: In the presentation phase, A presents an utterance for B to consider, on the assumption that B will provide evidence that she understands the utterance. In the acceptance phase, B provides this positive evidence of her understanding. Acceptance phases often take the form of simple backchannels such as “yeah” or “uh huh”. These short, formulaic utterances give minimal evidence that their speaker has accepted the last presented utterance. However, utterances that constitute the acceptance phase for one contribution may double as the presentation phase for the next. Take for example the following interaction.

G *the top you should be like at the right to about two inches below them at the right-hand corner of that*

F *of the corn field*

G *the left-hand corner*

F's utterance, "of the corn field" is both positive evidence of her understanding (in that it corrects G's initial utterance) and a presentation which must in turn be accepted by G. Likewise, G could have accepted F's utterance with a backchannel like "yeah", but instead issues a new presentation, "the left-hand corner", which doubles as evidence of his acceptance. In other words, the above interaction *could* have proceeded as follows:

G *the top you should be like at the right to about two inches below them at the right-hand corner of that*

F ***okay***

F *of the corn field*

G ***yeah***

G *the left-hand corner*

This imaginary version contains two short, formulaic utterances, which would have caused average similarity ratings to shoot up for both participants. But for individuals who are more willing to expend cognitive effort in formulating original utterances, original utterances may *replace* formulaic backchannels as positive evidence of grounding. This replacement may lower the number of formulaic utterances relative to original utterances and thereby lower their average similarity scores.

The above explanation is, of course, speculative. Future research should investigate the interplay between cognitive effort expenditure and various grounding mechanisms more directly.

***Limitations***

Even though Study 2 improved on Study 1 in sample size, construct operationalization, and confirmatory stance, it is still possible to discern some challenges to its validity. First, it is possible that the TF-IDF similarity measure of cognitive effort is influenced by lexical entrainment. Lexical entrainment is the tendency of conversation partners to converge on shared terms. This tendency has recently been shown to be a stable individual trait (Tobar-Henríquez et al., 2020). Insomuch as the similarity scores used in this study measure lexical similarity between conversation partners, they may reflect this trait. This may have been a sizeable contribution, as 37% of $sim_{MAX}$ scores represented the similarity between utterances of different conversation partners (i.e. people's utterances tend to be most similar to one of their own previous utterances, but not overwhelmingly so). If the similarity scores used in this study partially reflect lexical alignment, is that effect likely to dampen or enhance the relationship between those scores and repair avoidance? Dideriksen et al. (2020) analyzed the relationship between various grounding mechanisms in four different conversational contexts over two studies and found no significant correlation between lexical alignment and repair in any of their analyses. As such, it is likely that any influence of lexical alignment in this study dampened the observed effect size and was therefore not a threat to validity. It is also possible that "lexical alignment" refers to the same general phenomenon as does the aggregate similarity used in this study. Pickering and Garrod's Interactive Alignment theory—one of the most influential explanations of lexical alignment—attributes the phenomenon to priming (i.e. recent processing of words makes them more accessible in memory). Pickering and Garrod use this model to explain why dialogue seems to require much less cognitive effort than monologue—automatic alignment processes such as lexical alignment diminish the need to formulate original utterances in order to establish mutual understanding. If this is the case, cognitive effort expenditure (as investigated in this study) may partially reflect the lack of interactive alignment. Future research should investigate the degree of similarity between different partners' utterances separately from the degree of similarity between utterances of the same partner. Is the propensity to reformulate one's own utterances related to the propensity to reformulate a partner's? Perhaps the most successful conversationalists rephrase their own utterances but align with their partner's.

The possibility of a relationship between similarity scores and lexical alignment points to another possible limitation of the measure: its disregard for similar lexicon in all but the most similar preceding utterance (see Figure 5). The algorithm used in this study was designed with the assumption that utterances that reformulate will always be reformulating a particular previous utterance. First of all, this is often not the case. For example, consider again the interaction shown in Figure 5.

G *and then at the top of the "s" we're turning north*

F *mmhmm*

G *okay*

G *we're going straight due north*

F *okay*

G *which we're going to pass* ***on the south erm southeast*** *side*

F *mmhmm*

G *and we're going to do that in a curve almost a **half "u" shape***

F ***half "u" shape to the southeast***

Note that F's utterance, "half 'u' shape to the southeast" is a summary of two of G's earlier utterances (in red). Even though the similarity between "to the southeast" and "on the… southeast side" is initially recognized by the algorithm (see Figure 5), it is ultimately ignored when "and we're going… half 'u' shape" is chosen as $sim_{MAX}$. It seems then that the average similarity measure used in this study does not account for all reiteration of semantic content. Furthermore, many utterances do not reiterate or refer to the semantic content of any previous utterance, yet are associated with a particular $sim_{MAX}$, since the algorithm is programmed to find a $sim_{MAX}$ for every utterance. For utterances that do not refer to previous turns, the choice of one $sim_{MAX}$ while disregarding the majority of similar words in earlier utterances is entirely unjustified. For these reasons, future research should develop methods to account for all similarity without devaluing particular reformulated utterances. This may be accomplished in part by weighting similarities with previous utterances by the amount of intervening material since those utterances. This and other opportunities for improvement of the algorithm are discussed below.

***Opportunities for Improvement of the Algorithm***

A particularly promising result (and a particularly strong support for the Individual Effort Hypothesis) was that of integrating similarity scores with orthographic length and the number of interfering utterances since the last prime to produce an aggregate measure of effort in conversational utterance production. This aggregate measure predicts repair avoidance more strongly than does any individual measure analyzed in this thesis, and is robust to removal of formulaic utterances in preprocessing. Nevertheless, the way in which the integrated measure was computed was a crude estimate. The true relative contributions of similarity to previous utterances, number of interfering utterances since the last prime, and orthographic length to effort are as yet unknown. Furthermore, there may be interactions between these three components, or the proper transformation for orthographic length and number of interfering utterances may be something other than logarithmic (cf. Murre & Dros, 2015). It may even be worthwhile to add a fourth component—one that estimates the complexity of an utterance (e.g. that of Si & Callan, 2001). Future analyses might remedy this by building a model that optimizes the coefficients of each of the components for the ability of the aggregate measure to predict repair avoidance. Such optimized coefficients would likely produce a more accurate estimate of the cognitive effort involved in formulating each utterance. These coefficients might be cross-validated by reviewing the available literature on utterance length, complexity, and the rate at which lexical priming fades. Possible applications for an algorithm improved in this way will be discussed in the conclusion section.

## Study 3: Proposed Experimental Procedure[22]

Studies 1 and 2 yielded significant evidence that repair avoidance is modulated by individual differences in the propensity to expend cognitive effort in utterance formulation; those who are more willing to generate novel utterances are less likely to avoid asking clarification questions. In the sections above, I have theorized that this influence of cognitive effort-saving is due to the demand for cognitive effort imposed by certain trouble sources in

---

[22] This proposal was presented as a poster at the 2021 APS Virtual Convention.

conversation. In other words, certain problems in conversation cannot be remedied without effortful reformulation of previous utterances. The cognitive effort required to reformulate these problem utterances may deter people from initiating repair. To find direct support for the causal influence of such difficult problems in conversation, we must turn to an experiment. For Study 3, therefore, I propose an experiment to test the impact of cognitive effort saving on repair avoidance more directly, on the scale of the individual trouble source. This study is part of a new line of research that examines long-standing conversation-analytic principles under controlled, laboratory conditions (following a call by Albert and de Ruiter in 2018). It is a followup to the experiment conducted by Galantucci et al. (2020; described at the beginning of this thesis). In Galantucci et al.'s experiment, the cognitive effort required on the part of the participants was trivial. A simple "skask?" sufficed to alert their interlocutor to the specific problem. In fact, ten out of fifteen participants who initiated repair in the critical round of that study did so simply by picking up an object and offering it as a possible interpretation. In the proposed study, we will replicate the conditions of the previous study while altering the communicative trouble such that participants will be forced to reformulate the instruction in a way that is likely to be cognitively demanding. We theorize that this cognitive demand will be weighed against social and informational concerns that affect communicative faithfulness. Because these concerns will remain constant, we predict that the greater cognitive demand will result in even greater repair avoidance than observed in the previous study.

**Method**

Upon arriving at the lab, the participant will be "randomly" assigned a partner, who will in fact be a confederate of our lab. This pair will perform two tasks.

**Cover Task**. In the first task, the experimenter will ask the pair to rank "would you rather" questions in order of difficulty and humorousness. In doing so, half of the pairs will work silently, facing opposite directions (Formal Settings), while half will cooperate casually (Informal Settings). This cover task manipulation was also used in Galantucci et al. (2020), with inconclusive results. Our hope is that the combined data from that study and the proposed one will allow us to draw conclusions about the influence of social pleasantry on repair avoidance (i.e. on behavior in the critical task).

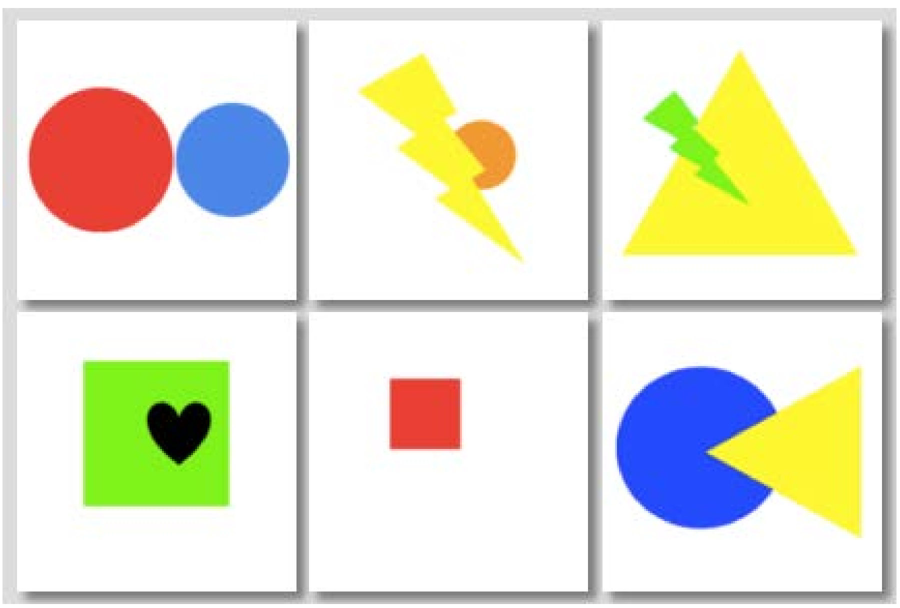
**Questionnaire 1.** The experimenter will ask each participant to answer three questions in writing, on a scale from 0 to 10: 1. How good of a job did you do in ranking the questions? 2. How good of a time did you have with your partner on this first task. 3. How good do you feel right now?

**Critical Task**. The participant and confederate will sit across from each other at a table, as shown in the drawing on the right. The experimenter will ask the participant to blindly pick one of two pieces of paper to determine his/her role. Both pieces will be labeled "mover", so that the participant will always be the mover, while the confederate will be the director. The experimenter will explain the rules: for each round of the game, the director will give instructions to the mover on how to move cards from a tray onto the chessboard and then back into the tray. The director will be given a picture of the chessboard illustrating where each object is to be placed and in what order the instructions are to be given (Figure 10).

However, the experimenter will explain, the director (i.e. the confederate) must use only words for individual properties, without any grammatical order. The experimenter will give an example of such an instruction based on an example picture. In reality, however, the confederate will have a predetermined script for each instruction. These instructions will contain all understandable sets of words, and for the first, second, and fourth rounds (each with six instructions), there will be no referential ambiguity (Figure 8).

**Figure 8**

*Mundane Round Cards*

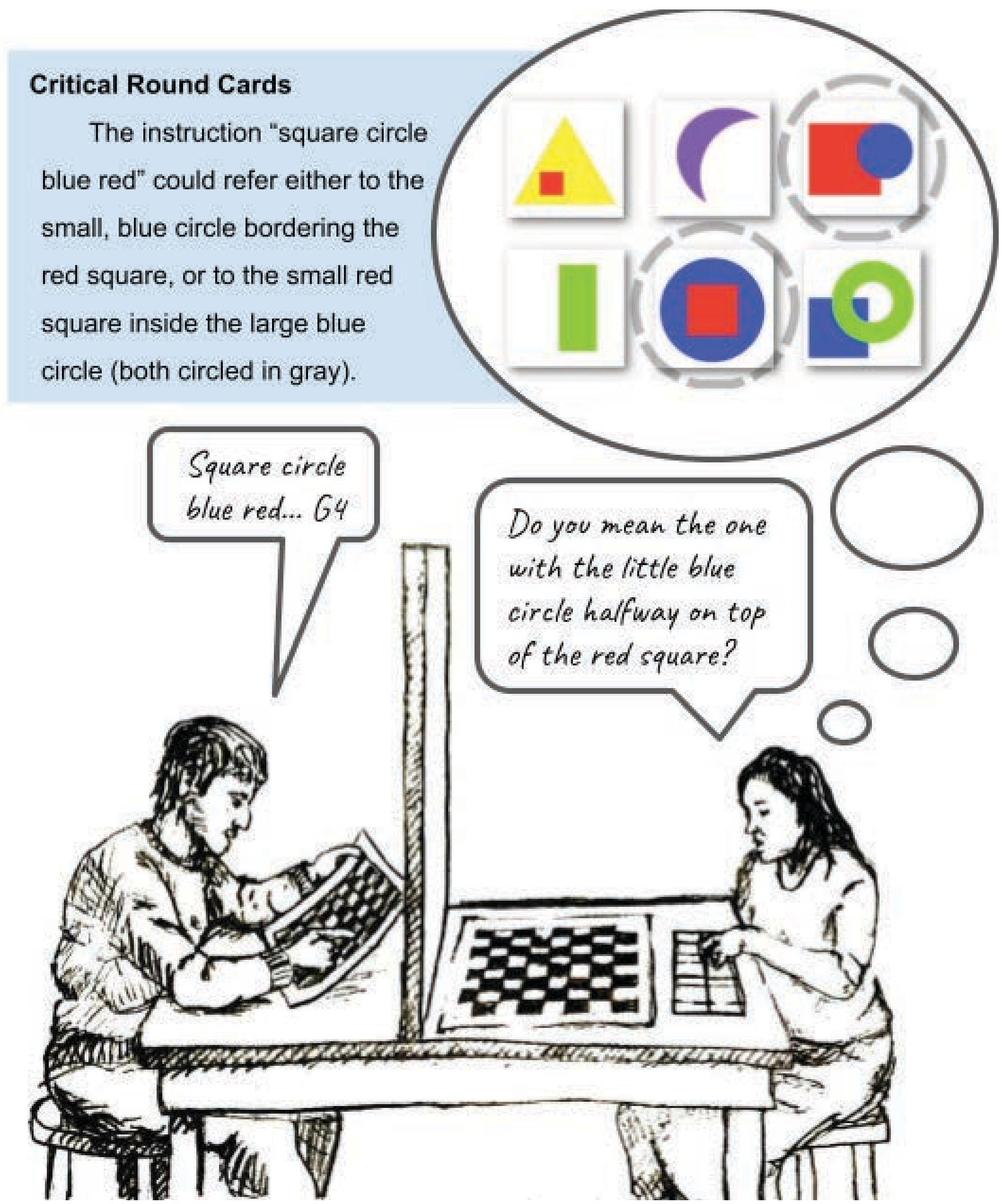


*Note*. The first four instructions for this round (given by the confederate) will resemble the following: 1 .“Black heart… G3,” 2. “Circle triangle… C5,” 3. “Little square red… F7,” 4. “Big lightning... F3.”

The third and fifth rounds will be different: for two cards in the tray, the lack of grammar will result in referential ambiguity. To minimize the chance of participants not noticing the ambiguity, both of these cards will have appeared separately in earlier, mundane rounds. At this point, the participant must choose whether to ask a clarifying question, as shown in Figure 9 below.

**Figure 9**

*Proposed Critical Trial Procedure*



If the participant initiates repair in the critical round, the confederate will indicate the small red square inside the blue circle. This step will remedy a possible problem with Galantucci et al. (2020) resulting from the lack of such a procedure. In that study, the confederate floundered in response to repair initiation in the critical round, given that there was no correct answer to "which one is the skask?" Three out of four participants who reported that their partner was not acting genuinely (and were therefore excluded from analyses) had initiated repair in the critical trial, suggesting that the confederate's inability to complete repair contributed to feelings that he wasn't acting genuinely. If this is the case, the risk of a type I error would be inflated (since those who avoid repair are disproportionately kept in the analysis). A scripted response to repair initiation may therefore minimize the need to exclude data points and improve the validity of the followup study.

**Questionnaire 2.** After the completion of the Critical Task, the experimenter will escort the participant and the confederate to separate rooms to answer a questionnaire. The questionnaire, administered only to the participant, will record standard demographic information, big five personality measures (BFI-2-S; Soto & John, 2017), a social intelligence test (the Mind in the Eye Test; Baron-Cohen, 2001), multilingualism (the LEAP-Q inventory; Marian et al., 2007), exclusionary measures (what do you think was the purpose of the study; do you think that your partner was acting genuinely), the three items from questionnaire 1, and a series of questions organized in a funnel approach to determine participants' awareness and perceptions of their repair or repair avoidance in the critical trial.

***Operationalization of Key Constructs***

**Repair Avoidance.** The absence of any explicit attempt to clarify the ambiguity of the critical round (Figure 9) through communication with the director (i.e. the confederate) will be considered repair avoidance.

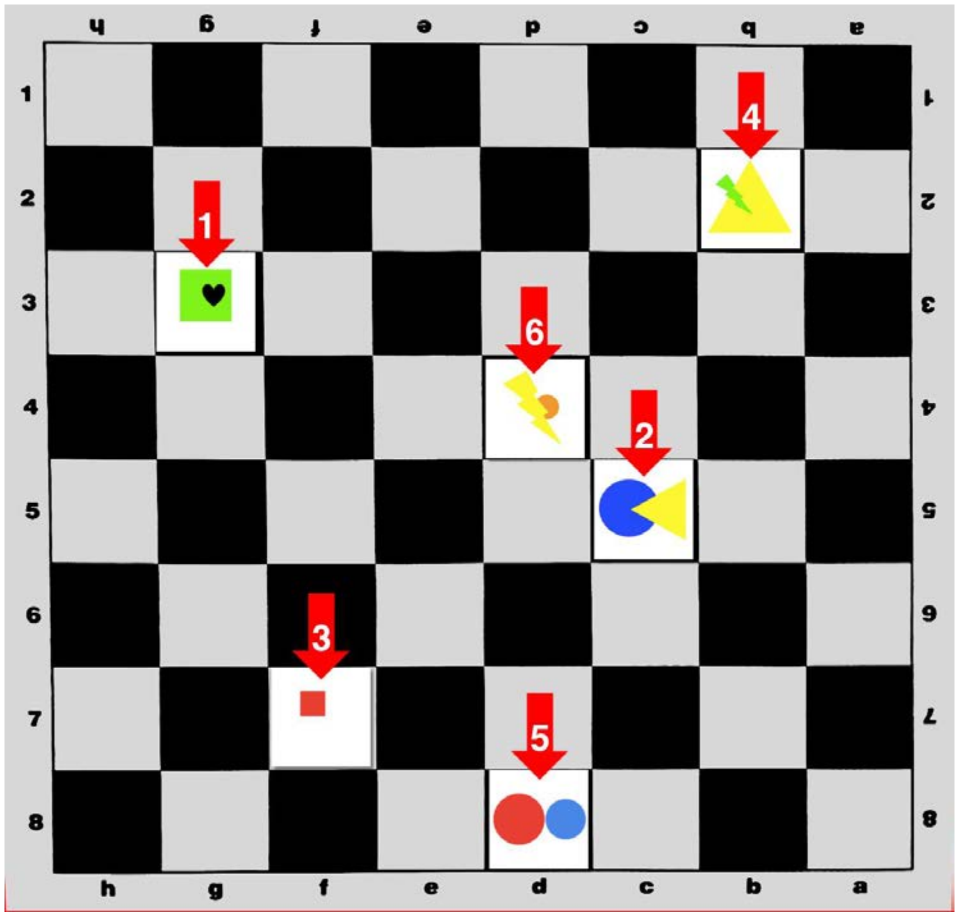
**Reformulation Effort.** Critical trial repair in the Galantucci et al. (2020) study will be considered to have required low reformulation effort, while that in the current study will be considered to require high reformulation effort. This manipulation will be verified in terms of orthographic length of transcribed repair initiations between the two studies, and in terms of similarity between repair initiations and the trouble source turn (both as rated by coders and computed using TF-IDF cosine similarity scores).

***Prediction***

Galantucci et al. (2020) found that repair avoidance was very common. Nevertheless, we predict that the greater cognitive cost required to formulate an effective repair in this proposed study will drive repair avoidance rates even higher. If the rates of repair avoidance are higher when reformulation costs are high, we may reasonably conclude that the frequent repair avoidance observed in conversation is at least partially a product of cognitive effort-saving.

**Figure 10**

*Example Director's Sheet*




*Note*. To be shown to the mover (i.e. participant) before being handed to the director (i.e. confederate). This will convince the mover that the director is independently formulating instructions, and that visual referents (the shape cards) are shared by both parties during the game. With this grounding, the participant will be able to reformulate instructions through reference to visual features of the cards.

## Conclusion

In this thesis I have investigated the intersection between two fields of inquiry: that concerned with the reasons for repair avoidance, and that concerned with the influence of cognitive effort-saving on human behavior. In particular, I have argued that the avoidance of repair is due in part to avoidance of the cognitive effort expenditure required to formulate a repair initiation. I have found significant support for a wider implication of this hypothesis: that individuals who tend in general to avoid formulating new utterances are also more likely to avoid repair initiation (the Individual Effort Hypothesis). In two analyses of the HCRC Map Task Corpus (Anderson et al., 1991), I have found that various operational definitions of individual average reformulation effort predict repair avoidance by participants when they are acting as Follower in the map task. This is true when reformulation is measured in repair initiations alone (Study 1) and when it is measured in all utterances (Study 2). Despite the importance of reformulation, some evidence suggests that a more comprehensive measure of effort expenditure in conversation might produce an even more precise prediction of repair avoidance (Study 2, Post hoc Analysis and Discussion).

Besides the above correlational research, I have proposed an experiment to confirm the causal link between reformulation effort and repair avoidance on the level of an individual trouble source.

Because this thesis is concerned with the intersection between two fields of inquiry—repair avoidance and cognitive effort—the findings summarized above suggest future directions for both.

**Investigating Repair Avoidance**

A primary question I have posed in this thesis is: Why do people avoid asking for clarification? This thesis argues for one possible answer to this question—they are deterred by the cognitive effort involved. When I explain the topic of my thesis to friends, however, their most common response is, “I know why people don’t ask when they should! They’re afraid they’ll look stupid!” Indeed, cognitive effort-saving is likely to be one of many reasons for repair avoidance. Conversely, there are likely to be many situationally-dependent benefits that cause people to initiate repair more often. Future research should investigate these costs and benefits systematically to arrive at a comprehensive explanation. In the following pages, therefore, I will outline a general framework within which this research may proceed, and suggest specific research questions for investigation.

Figure 11 outlines theoretical costs and benefits associated with initiating various types of repairs. Costs are in red boxes, while benefits are in green. Looking down the leftmost column, for example, one can see what considerations might play on a speaker who would like to verify his understanding of what his partner has just said, and is considering initiating a reformulated offer, as in the following dialogue from the Map Task:

G *head southwest. that's to the left and down past indian country.*

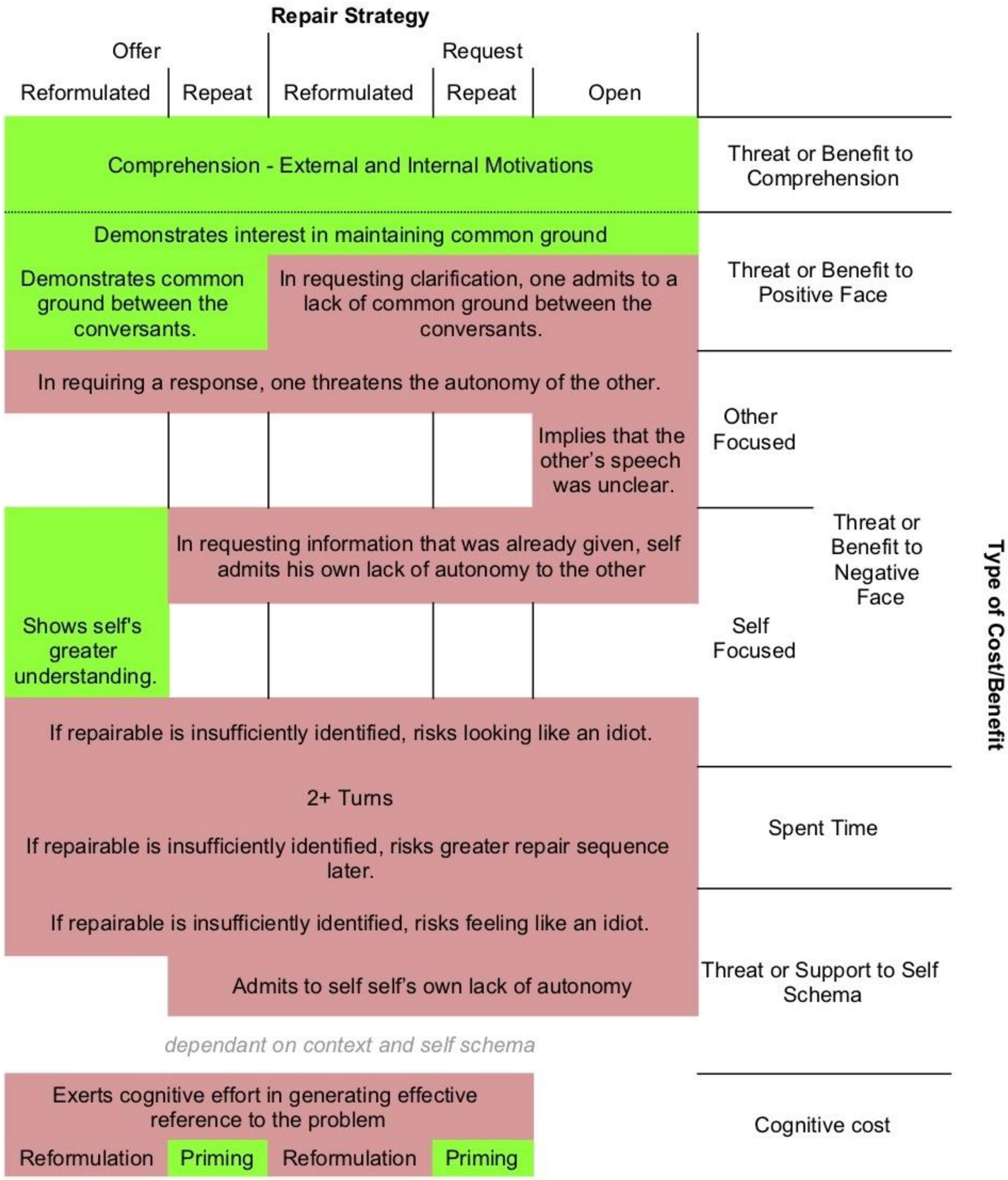
F ***move west followed by south down to***

G *no*

G *s-- straight southwest.*

**Figure 11**

*Theoretical Cost/Benefit Factors in Other-Initiated Repair*



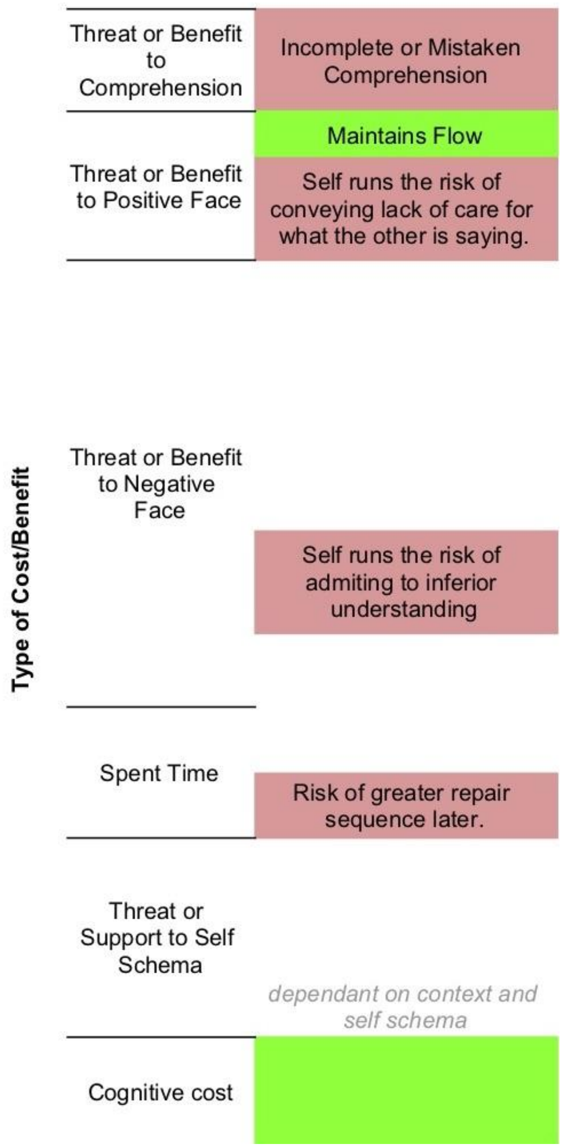

*Note*. The potential repair initiator here is referred to as "self", while the speaker of the trouble source is "other".

In uttering the repair initiation, “move west followed by south down to…”, F ‘benefits’ his comprehension of G’s instruction. In fact, it seems likely that F would have entirely misunderstood the instruction had he not offered his understanding for confirmation. In initiating repair, F also demonstrates his interest in maintaining common ground, which may be important for his friendship with G. On the other hand, F is requiring G to respond to his offer, which may slightly threaten G’s autonomy or “negative face” (see Holtgraves, 2002, ch. 2; cf. Yoon et al., 2020). Furthermore, F has taken the risk of embarrassing himself by showing clearly how he misunderstood the instruction. F may think to himself afterward, “How could I not have realized that southwest means straight southwest!?”. And so on down the column.

If we can draw up a chart of costs and benefits of repair initiation, we can likewise chart costs and benefits of repair avoidance. This is shown in Figure 12. Since avoidance is the absence of any kind of repair, there is only one column.

**Figure 12**

*Theoretical Cost/Benefit Factors in Repair Avoidance*




Each colored box in Figures 9 and 10 represents a possible predictor of repair avoidance. Many of them might be experimentally manipulated, as was proposed for cognitive cost in Study 3.

To complicate matters further, it is possible that some of the cost-benefit factors outlined in Figure 12 apply to the perception of communicative problems as well as conscious repair avoidance. Just as it is much more costly to misperceive a poisonous mushroom as edible than to confuse two edible mushrooms, it is much more costly to miss the fact that you did not understand certain pieces of information than others. This cost, as well as the corresponding benefits, may impact the probability that certain misunderstandings are perceived. Recent research has already applied similar cost-benefit models to visual perception and to cognitive biases in general (Haselton & Nettle, 2006; Sims, 2018).

A complete understanding of the mechanisms at work in problem perception, repair initiation, and repair avoidance would be valuable. This is because factors at work in repair initiation are, in all likelihood, closely related to those that govern conversation in general, and grounding mechanisms in particular. Repair is a method of maintaining clear communication between participants (i.e. faithfulness). Of course, repair is only one method of maintaining faithfulness, and is sometimes difficult to distinguish clearly from other such methods. As anyone who is capable of speaking a language knows well, conversation is a complex phenomenon, involving constant decisions about what to say and what not to say, how direct or indirect to be, when to offer backchannels (e.g. uh huh; yeah; you don't say!), when to take the floor, etc. All of these decisions are likely to be influenced by situationally-dependent cost-benefit concerns.

Developing a precise understanding of the social-psychological forces that influence communicative faithfulness is essential, first of all, for the development of protocols that maximize faithfulness. Mis- or non-understandings of critical information can cause fatal accidents in some fields (Cushing, 1994). In all businesses, they can waste valuable time and resources. Certain communicative roles, such as those of airplane pilots and air traffic control operators, are governed by strict communication protocols designed to reduce miscommunication. Such protocols are becoming increasingly automated (see e.g. Lemoine et al., 1996), opening the door to more precise control of cost-benefit factors (see Cahn & Brennan, 1999, for a similar approach). For example, the findings presented in this thesis indicate that communication protocols should find ways to reduce the cognitive costs associated with asking for clarification (e.g. by installing a button that automatically replays the last few seconds of conversation, or by implementing regular reminders that repair can be initiated through some simple shorthand; cf. Bohus & Rudnicky, 2005). Of course, many automated and non-automated communication protocols must optimize for multiple goals (communication should be efficient, precise, polite, etc.). Further research into the mechanisms which drive the management of these various goals in conversation is therefore necessary.

**Investigating Cognitive Effort in Dyadic Interaction**

In the process of investigating the reasons for repair avoidance in this thesis, I have made some steps toward developing a method for measuring an individual's propensity to expend effort in conversation (and doing so computationally using only a few transcripts). In my view, this is the most promising contribution of this thesis. The analyses above suggest that the propensity to expend effort is likely to be a stable individual trait. If this could be shown with a measure based on relatively crude language processing techniques, it is quite possible that a carefully developed and validated iteration of the algorithm developed here could become a clinically useful tool.

First, the propensity to expend cognitive effort in conversational utterances is likely to reflect both a general need for cognition (cf. Cacioppo et al., 1996) and some component of general cognitive ability. In support of this theory, Engelhardt et al. (2013) found that incidence of self-initiated repairs was closely tied to a particular cognitive ability (inhibitory control). They also showed significant correlations between repair disfluencies and two intelligence subtests, suggesting a direct link between intelligence and utterance formulation. Verbal intelligence testing is useful for a variety of clinical applications (see e.g. de Oliveira et al., 2020), but the administration of verbal intelligence tests requires trained professionals (Wechsler, 1958). In contrast, voice to text transcription can now be performed nearly flawlessly by a computer—the administration of an automated test of cognition in dialogue (such as that developed for this thesis) might require no human intervention beyond the construction of a controlled environment in which to have a conversation (or more precisely, a few with different partners). Indeed, I was not the first to conceive of such a test: Zablotskaya et al. (2012) were able to predict verbal IQ scores using TF-IDF similarity between the text of a video and participants' summaries of that video. This technique is less developed than the one presented in this thesis, but it does demonstrate the potential of natural language processing algorithms for measuring underlying psychological traits.

Of course, the dream of an automated verbal IQ test is a long way off. The techniques developed in this thesis need refinement (see "Opportunities for Improvement of the Algorithm" above), test-retest reliability must be demonstrated, and the construct should be compared to known measures such as that of verbal IQ or inhibitory control. I believe that this thesis presents viable first steps towards an effective test of communication intelligence. But even if a procedure like that used in this thesis does not prove to be clinically useful, it could have important implications for the study of language and communication. In particular, it has the potential to demystify the link between cognitive psychology and dyadic interaction. Psycholinguistics has had a long history of focus on the individual, treating language production and comprehension as if they occur only alone in a laboratory setting (O'Connell & Kowal, 2003). In recent decades, there has been increasing recognition that language is a dynamic process that occurs naturally between multiple people, and that it must be treated as such (e.g. Pickering & Garrod, 2004; Mills, 2014a; Stephens et al., 2010). The ability to computationally measure such complex processes as lexical alignment (e.g. Brennan, 1996; Dideriksen et al., 2020) has contributed to many important advances in recent years. With further research, the measurement of cognitive effort expenditure may prove similarly useful.

## References

Achananuparp, P., Hu, X., Shen, X. (2008). The evaluation of sentence similarity measures. In: Song, I. Y., Eder, J., Nguyen, T. M. (eds). *International Conference on Data Warehousing and Knowledge Discovery - Lecture Notes in Computer Science, 5182*, 305–316. doi: 10.1007/978-3-540-85836-2_29

Albert, S., & de Ruiter, J. P. (2018). Repair: The interface between interaction and cognition. *Topics in Cognitive Science, 10*(2), 279–313. doi: 10.1111/tops.12339

Alexander, D., Wetherby, A., & Prizant, B. (1997). The emergence of repair strategies in infants and toddlers. *Seminars in speech and language, 18*(3), 197–212. doi: 10.1055/s-2008-1064073

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S., & Weiniert, R. (1991). The Hcrc Map Task Corpus. *Language & Speech, 34*(4), 351–366. doi: 10.1177/002383099103400404

Apperly, I. A., Riggs, K. J., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is Belief Reasoning Automatic? *Psychological Science, 17*(10), 841–844. doi: 10.1111/j.1467-9280.2006.01791.x

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G.H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). New York: Academic Press.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 42*(2), 241–251.

Bock, K., & Levelt, W. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945–984). Academic Press.

Bohus, D., Rudnicky, A. I. (2005) Sorry, I didn't catch that! - An investigation of non-understanding errors and recovery strategies. *Proceedings of the 6th SIGdial Workshop on Discourse and Dialogue*, 128–143.

Boyle, E. A., Anderson, A. H., & Newlands, A. (1994). The Effects of Visibility on Dialogue and Performance in a Cooperative Problem Solving Task. Language and Speech, 37(1), 1–20. doi: 10.1177/002383099403700101

Brennan, S. E. (1996). Lexical entrainment in spontaneous dialog. *Proceedings of the International Symposium on Spoken Dialogue*, 41–44.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(6), 1482–1493. doi: 10.1037/0278-7393.22.6.1482

Brown, G. , Anderson, A. , Yule, G. , and Shillcock, R. (1983). *Teaching Talk*. Cambridge, U.K.: Cambridge University Press.

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin, 119*(2), 197–253. doi: 10.1037/0033-2909.119.2.197

Cahn, J. E., & Brennan, S. E. (1999). A psychological model of grounding and repair in dialogue. *AAAI Technical Report FS-99-03*, Association for the Advancement of Artificial Intelligence.

Ciapuscio, G. E. (2003). Formulation and reformulation procedures in verbal interactions between experts and (semi-)laypersons. *Discourse Studies, 5*(2), 207. doi: 10.1177/1461445603005002310

Clark, H. H. & Brennan, S. E. (1991). Grounding in Communication. In Resnick, L. B. & Levine, J. M. (eds.), *Perspectives on Socially Shared Cognition*, American Psychological Association, 127-149.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*(1), 1–39. doi: 10.1016/0010-0277(86)90010-7

Clift, R. (2016). *Conversation analysis*. Cambridge, UK: Cambridge University Press.

Coane, J. H., & Balota, D. A. (2010). Repetition priming across distinct contexts: effects of lexical status, word frequency, and retrieval test. *Quarterly Journal of Experimental Psychology 63*(12), 2376–2398. doi: 10.1080/17470211003687546

Colman, M., & Healey, P. (2011). The Distribution of Repair in Dialogue. *Cognitive Science, 33*.

Cushing, S. (1994). *Fatal words: Communication clashes and aircraft crashes*. University of Chicago Press.

de Oliveira, L. R., Brian, J., Kelley, E., Beal, D., Nicolson, R., Georgiades, S., Iaboni, A., Fragiadakis, S. D., Ristic, L., Anagnostou, E., & Sanjeevan, T. (2020). Exploring the use of the verbal intelligence quotient as a proxy for language ability in autism spectrum disorder. *Research in Autism Spectrum Disorders, 73*.

Dideriksen, C., Christiansen, M. H., Tylén, K., Dingemanse, M., & Fusaroli, R. (2020, October 12). *Building common ground: Quantifying the interplay of mechanisms that promote understanding in conversations*. doi: 10.31234/osf.io/a5r74

Dingemanse, M., Roberts, S. G., Baranova, J., Blythe, J., Drew, P., Floyd, S., Gisladottir, R. S., Kendrick, K. H., Levinson, S. C., Manrique, E., Rossi, G. & Enfield, N. J. (2015). Universal principles in the repair of communication problems. *PLOS One 10*(9): e0136100. doi:10.1371/journal.pone.0136100.

Dupoux, E., & Green, K. (1997). Perceptual adjustment to highly compressed speech: Effects of talker and rate changes. *Journal of Experimental Psychology: Human Perception and Performance, 23*(3), 914–927. doi: 10.1037/0096-1523.23.3.914

Engelhardt, P. E., Nigg, J. T., & Ferreira, F. (2013). Is the fluency of language outputs related to individual differences in intelligence and executive function?. *Acta psychologica, 144*(2), 424–432. doi: 10.1016/j.actpsy.2013.08.00

Fenn, K. M., Shintel, H., Atkins, A. S., Skipper, J. I., Bond, V. C., & Nusbaum, H. C. (2011). When less is heard than meets the ear: Change deafness in a telephone conversation. *Quarterly Journal of Experimental Psychology, 64*(7), 1442-1456.

Ferreira, F., and Engelhardt, P. E. (2006). Syntax and Production. In M. J. Traxler and M. A. Gernsbacher (Eds.), *Handbook of Psycholinguistics, Second Edition* (pp. 61-91). Academic Press. doi: 10.1016/B978-012369374-7/50004-3

Frings, C., Schneider, K. K., & Fox, E. (2015). The negative priming paradigm: An update and implications for selective attention. *Psychonomic Bulletin & Review, 22*(6), 1577–1597. doi: 10.3758/s13423-015-0841-4

Gabaix, X. (2017) Behavioral inattention. *NBER Working Paper No. 24096*. National Bureau of Economic Research.

Galantucci, B., Roberts, G., & Langstein, B. (2018). Content deafness: When coherent talk just doesn't matter. *Language and Communication 61*, 29-34. doi: 10.1016/j.langcom.2018.01.001

Galantucci, B., Langstein, B., Spivack, E., & Paley, N. (2020). Repair avoidance: When faithful informational exchanges don't matter that much. *Cognitive Science, 44*(10). doi: 10.17605/OSF.IO/3X9C6

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A Study in conceptual and semantic co-ordination. *Cognition, 27*, 181–218.

Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences, 8*, 8–11. http://dx.doi.org/10.1016/j.tics.2003.10 .016

Grice, H. P. (1975). Logic and conversation (from the William James lectures, Harvard University, 1967). In P. Cole, & J. Morgan (Eds.), *Syntax and semantics 3: Speech acts*, Academic Press, New York, 41–58.

Hawkins, J. A. (2004) *Efficiency and complexity in grammars*. Oxford University Press.

Haselton, M. G., & Nettle, D. (2006). The paranoid optimist: an integrative evolutionary model of cognitive biases. *Personality and Social Psychology Review, 10*(1), 47–66. doi: 10.1207/s15327957pspr1001_3

Healey, P. G. T. (2008). Interactive misalignment: The Role of Repair in the Development of Group Sub-languages. In R. Cooper & R. Kempson (Eds.), *Language in Flux*. College Publications.

Healey, P. G., Mills, G., Eshghi, A., & Howes, C. (2018). Running Repairs: Coordinating Meaning in Dialogue. *Topics in Cognitive Science, 10*(2), 367–388. doi: 10.1111/tops.12336

Healey, P. G, Purver, M., King, J., Ginzburg, J., & Mills, G. J. (2003). Experimenting with Clarification in Dialogue. *Proceedings of the Annual Meeting of the Cognitive Science Society, 25.*

Holtgraves, T. M. (2002). *Language as social action: Social psychology and language use*. Lawrence Erlbaum Associates Publishers.

Kemp, C. & Regier, T. (2012) Kinship categories across languages reflect general communicative principles. *Science 336*(6084):1049–54. doi:10.1126/science.1218811.

Kendrick, K. H. (2015). Other-initiated repair in English. *Open Linguistics*, *1*. doi: doi: 10.2478/opli-2014-0009

Kendrick K. H. (2015). The intersection of turn-taking and repair: the timing of other-initiations of repair in conversation. *Frontiers in Psychology, 6*, 250. doi: 10.3389/fpsyg.2015.00250

Keysar, B. (2007). Communication and miscommunication: The role of egocentric processes, *Intercultural Pragmatics, 4*(1), 71-84. doi: 10.1515/IP.2007.004

Krauss, R. M., & Glucksberg, S. (1977). Social and nonsocial speech. *Scientific American, 236*(2), 100–105. doi: 10.1038/scientificamerican0277-100

Lemoine, M.P., Debernard, S., Crevits, I. et al. Cooperation between humans and machines: First results of an experiment with a multi-level cooperative organisation in air traffic control. *Computer Supported Cooperative Work 5*, 299–321 (1996). doi: 10.1007/BF00133661

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levelt W. J. (1999). Models of word production. *Trends in Cognitive Sciences, 3*(6), 223–232. doi: 10.1016/s1364-6613(99)01319-4

Levinson S. C. (2016). Turn-taking in Human Communication--Origins and Implications for Language Processing. *Trends in Cognitive Sciences, 20*(1), 6–14. doi: 10.1016/j.tics.2015.10.010

Lieder, F., & Griffiths, T. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences, 43*, E1. doi:10.1017/S0140525X1900061X

Lin, S., Keysar, B., & Epley, N. (2010). Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention. *Journal of Experimental Social Psychology, 46*(3), 551–556. doi: 10.1016/j.jesp.2009.12.019

Long, J. S. & Ervin, L. H. (2000). Using heteroscedasticity consistent standard errors in the linear regression model. *The American Statistician, 54*(3), 217–224. doi: 10.2307/2685594

Malinowski, B. (1923). The Problem of Meaning in Primitive Languages. In C. K. Ogden, & I. A. Richards (Eds.), *The Meaning of Meaning* (pp. 296–336). London: K. Paul, Trend, Trubner.

Manrique, E., & Enfield, N. J. (2015). Suspending the next turn as a form of repair initiation: Evidence from Argentine sign language. *Frontiers in Psychology, 6*, 1–21. doi: 10.3389/fpsyg.2015.01326

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50*(4), 940–967. doi: 10.1044/1092-4388(2007/067)

Micklos, A., Walker, B., & Fay, N. (2020). Are People Sensitive to Problems in Communication?. *Cognitive Science, 44*(2), e12816.

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

Mills, G. (2014). Establishing a communication system: Miscommunication drives abstraction. *Proceedings of the Cognitive Science Society, 36*(36).

Müller, R., Paul, D., & Li, Y. (2021). Reformulation of symptom descriptions in dialogue systems for fault diagnosis: How to ask for clarification? *International Journal of Human - Computer Studies*, 145. doi: 10.1016/j.ijhcs.2020.102516

Murre, J. M., & Dros, J. (2015). Replication and Analysis of Ebbinghaus' Forgetting Curve. *PLOS One, 10*(7), e0120644. doi: 10.1371/journal.pone.0120644

Murray, W. S., & Forster, K. I. (2004). Serial Mechanisms in Lexical Access: The Rank Hypothesis. *Psychological Review, 111*(3), 721–756. doi: 10.1037/0033-295X.111.3.721

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*(3), 231–259. doi: 10.1037/0033-295X.84.3.231

O'Connell, D. C., & Kowal, S. (2003). Psycholinguistics: A half century of monologism. The *American Journal of Psychology, 116*(2), 191–212.

Pace-Sigge, M., & Patterson, K. J. (2017). *Lexical priming : Applications and advances*. John Benjamins Publishing Company.

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences of the United States of America, 108*(9), 3526–3529. doi: 10.1073/pnas.1012551108

Pickering, M. J., & Garrod, S. (2004). The interactive-alignment model: Developments and refinements. *The Behavioral and Brain Sciences, 27*(2), 212–225. doi: 10.1017/S0140525X04450055

Reitter, D., & Moore, J. D. (2007). Predicting success in dialogue. *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics (ACL)*, 808–815.

Regier, T., Kay, P. & Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences 104*(4):1436–41. doi:10.1073/pnas.0610341104.

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

Schober, M. F., & Clark, H. H. (1989). Understanding by Addressees and Overhearers. *Cognitive Psychology, 2*, 211.

Schegloff, E.A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language 53*(2), 361-382. doi:10.1353/lan.1977.0041.

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. London: Cambridge University Press.

Si, L., & Callan, J. (2001). A statistical model for scientific readability. CIKM '01.

Sidnell, J. (2011). *Conversation analysis: An introduction*. London: Wiley-Blackwell.

Simon, H. A. (1956) Rational choice and the structure of the environment. *Psychological Review 63*(2):129–38. doi:10.1037/h0042769.

Simons, D. J., & Levin, D. T. (1998). Failure to detect changes to people during a real-world interaction. *Psychonomic Bulletin & Review, 5*(4), 644-649. doi: 10.3758/bf03208840

Sims, C. A. (2006) Rational inattention: Beyond the linear-quadratic case. *American Economic Review 96*(2):158–63. doi:10.1257/000282806777212431.

Sims C. R. (2018). Efficient coding explains the universal law of generalization in human perception. *Science* 360(6389), 652–656. doi: 10.1126/science.aaq1118

Sinclair, J.M., Jones, S. & Daley, R. (2004). *English Collocation Studies: The OSTI Report*. London: Continuum.

Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology, 113*(1), 117–143. doi: 10.1037/pspp0000096

Spärck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation. 28*, 11–21. doi: 10.1108/eb026526.

Stephens, G. J., Silbert, L. J., & Hasson, U. (2010). Speaker-listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences of the United States of America, 107*(32), 14425–14430. doi: 10.1073/pnas.1008662107

Tobar-Henríquez, A., Rabagliati, H., & Branigan, H. P. (2020). Lexical entrainment reflects a stable individual trait: Implications for individual differences in language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(6), 1091–1105. doi: 10.1037/xlm0000774

Vul, E., Goodman, N. D., Griffiths, T. L. & Tenenbaum, J. B. (2014) One and done? Optimal decisions from very few samples. *Cognitive Science 38*(4):599–637. doi:10.1111/cogs.12101.

Wechsler, D. (1958). *The Measurement and Appraisal of Adult Intelligence (4th ed.)*. Baltimore (MD): Williams & Witkins.

Westbrook, A., & Braver, T. S. (2015). Cognitive effort: A neuroeconomic approach. *Cognitive, Affective & Behavioral Neuroscience, 15*(2), 395–415. doi: 10.3758/s13415-015-0334-y

Williams, S. A., & Korko, M. (2019). Pause behavior within reformulations and the proficiency level of second language learners of English. *Applied Psycholinguistics, 40*(3), 723–742. doi: 10.1017/S0142716418000802

Yoon, E. J., Tessler, M. H., Goodman, N. D., & Frank, M. C. (2020). Polite Speech Emerges From Competing Social Goals. Open mind : discoveries in cognitive science, 4, 71–87. doi: 10.1162/opmi_a_00035

Zablotskaya, K., Martínez, F., & Minker, W. (2012). Investigating Verbal Intelligence Using the TF-IDF Approach. *Proceedings of LREC 2012, Eighth International Conference on Language Resources and Evaluation*. pp. 1573–1576.

Zaslavsky, N., Kemp, C., Regier, T. & Tishby, N. (2018) Efficient compression in color naming and its evolution. *Proceedings of the National Academy of Sciences 115*(31):7937–42. doi:10.1073/pnas.1800521115.

Zipf, G. K. (1949) *Human behavior and the principle of least effort: An introduction to human ecology.* Addison-Wesley Press. [aFL, MD]

# Appendix A: Utterance Similarity Algorithm

```
import os
import numpy as np
from sklearn.feature_extraction.text import TfidfVectorizer

# Change this directory depending on where you store your conversation files
# https://groups.inf.ed.ac.uk/maptask/transcripts/
directory = r'/Users/louisteitelbaum/Downloads/Transcripts'

corpus = []
ConversationSizes = []
conversationList = []

for entry in sorted(os.scandir(directory), key=lambda e: e.name):
	labeledList = []
	if entry.path.endswith(".txt"):
		with open(entry.path) as file:
			for line in file:
				labeledList.append([line[:1], line[2:(len(line) - 3)]])  # Make a list of
				lists, separating each utterance from its label

	# Remove the file header
	labeledList.pop(0)
	labeledList.pop(0)
	labeledList.pop(0)

	# This is useful later when we have to chop up the giant corpus matrix into conversations
	ConversationSizes.append(len(labeledList))

	# Appends every line of the current conversation to the corpus
	for utterance in labeledList:
		corpus.append(utterance[1])

	# Saves this list with all of the labeled utterances so we can look at it later
	conversationList.append(labeledList)

# Compute TF-IDF similarity scores
#
https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html?h
ighlight=tf%20idf#sklearn.feature_extraction.text.TfidfVectorizer
vect = TfidfVectorizer(min_df=1)
tfidf = vect.fit_transform(corpus)
pairwise_similarity = tfidf * tfidf.T

# Giant array of similarity numbers for every utterance in every conversation
bigArray = np.array(pairwise_similarity.toarray())
```

```
ConvoIndex = 0
for ConversationSize in ConversationSizes:

	# Look only at each conversation at a time, store all the similarity data for each utterance in
currSimilarity
	currSimilarity = bigArray[:ConversationSize]  # Remove unnecessary rows
	currSimilarity = np.delete(currSimilarity, np.s_[ConversationSize:], 1)  # Remove unnecessary
	columns

	# Find the highest similarity for each utterance, save it and append it to our labeled list
	for utteranceIndex in range(5, ConversationSize):
		maxSimilarity = 0
		for index in range(utteranceIndex):
			if currSimilarity[utteranceIndex][index] > maxSimilarity:
				maxSimilarity = currSimilarity[utteranceIndex][index]
		conversationList[ConvoIndex][utteranceIndex].append(maxSimilarity)

	# Calculate number of utterances per person and the average similarity for each person in the
conversation
	Fsum = Gsum = Fnum = Gnum = 0
		for utteranceIndex in range(5, ConversationSize):
			if conversationList[ConvoIndex][utteranceIndex][0] == "f":
				Fsum += conversationList[ConvoIndex][utteranceIndex][2]
				Fnum += 1
			if conversationList[ConvoIndex][utteranceIndex][0] == "g":
				Gsum += conversationList[ConvoIndex][utteranceIndex][2]
				Gnum += 1
		Favg = Fsum / Fnum
		Gavg = Gsum / Gnum

	# Print some useful information
	print("File: ", ConvoIndex+1)
	print("Number of utterances of person g: ", Gnum)
	print("Number of utterances of person f: ", Fnum)
	print("Average Similarity for person g: ", Gavg)
	print("Average Similarity for person f: ", Favg)
	print("")

	# Remove the conversation we just looked at from the giant array
	bigArray = bigArray[ConversationSize:]  # Remove rows
	bigArray = np.delete(bigArray, np.s_[:ConversationSize], 1)  # Remove columns

	# Iterate to next conversation
	ConvoIndex += 1
```

## Appendix B: Integrated Utterance Effort Algorithm

```
import os
```

```
import sys
import numpy as np
from sklearn.feature_extraction.text import TfidfVectorizer
import csv
from csv import writer

# Change this directory depending on where you store your conversation files
# https://groups.inf.ed.ac.uk/maptask/transcripts/
directory = r'/Users/louisteitelbaum/Downloads/Transcripts'

# Make a new spreadsheet
with open('NewSimilarityResults2.csv', 'w+') as csvfile:
    filewriter = csv.writer(csvfile, delimiter=',',
                            quotechar='|', quoting=csv.QUOTE_MINIMAL)

def append_list_as_row(file_name, list_of_elem):
    # Open file in append mode
    with open(file_name, 'a+', newline='') as write_obj:
        # Create a writer object from csv module
        csv_writer = writer(write_obj)
        # Add contents of list as last row in the csv file
        csv_writer.writerow(list_of_elem)
fields = ['Conversation', 'UtterancesG', 'UtterancesF', 'AvgSimilarityG', 'AvgSimilarityF', 'AvgEffortG',
'AvgEffortF', 'AvgOrthoF', 'AvgOrthoG']
append_list_as_row('NewSimilarityResults2.csv', fields)

deletions = ["aye", "erm no", "ehm no", "eh oh right", "nope", "er right okay", "'kay", "yep okay", "yep
yep", "ehm well", "erm well", "uh-huh uh-huh", "alright", "and then", "oh right", "um mmhmm", "uh-huh
right", "right so", "so", "okay so", "right um", "mm oh aye", "okay erm", "so ehm", "um", "right er",
"mm", "ehm", "no", "no no", "no no no", "erm", "but", "now", "well", "oh right okay","mmhmm", "yeah",
"okay", "okay yeah", "right okay", "right", "uh-huh", "right-fine", "uh-huh yeah", "yep", "okay then"]
backchannels = ["mmhmm", "yeah", "okay", "okay yeah", "right okay", "right", "uh-huh", "right-fine",
"uh-huh yeah", "yep", "okay then"]

corpus = []
ConversationSizes = []
conversationList = []

ConvoIndex = 0
for entry in sorted(os.scandir(directory), key=lambda e: e.name):
    labeledList = []
    if entry.path.endswith(".txt"):
        with open(entry.path) as file:
            for line in file:
                labeledList.append([line[:1], line[2:(len(line) - 3)]])  # Make a list of lists, separating each
utterance from its label

        # Remove the file header
```

```
    labeledList.pop(0)
    labeledList.pop(0)
    labeledList.pop(0)

    # Remove specific utterances from the list
    for utterance in labeledList:
      for word in deletions:
        if utterance[1] == word:
          labeledList.remove(utterance)

    # This is useful later when we have to chop up the giant corpus matrix into conversations
    ConversationSizes.append(len(labeledList))

    # Appends every line of the current conversation to the corpus
    for utterance in labeledList:
      corpus.append(utterance[1])

    # Saves this list with all of the labeled utterances so we can look at it later
    conversationList.append(labeledList)

    ConvoIndex += 1
    #print("File: ", ConvoIndex, "     ", entry.path[45:])
    #print("Number of Backchannels: ", NumBackchannels)
    #print(" ")

# Compute TF-IDF similarity scores
vect = TfidfVectorizer(min_df=1)
tfidf = vect.fit_transform(corpus)
pairwise_similarity = tfidf * tfidf.T

# Giant array of similarity numbers for every utterance in every conversation
bigArray = np.array(pairwise_similarity.toarray())

ConvoIndex = 0
for ConversationSize in ConversationSizes:
  # Look only at each conversation at a time, store all the similarity data for each utterance in
currSimilarity
  currSimilarity = bigArray[:ConversationSize]  # Remove unnecessary rows
  currSimilarity = np.delete(currSimilarity, np.s_[ConversationSize:], 1)  # Remove unnecessary
columns
  # Find the highest similarity for each utterance, save it and append it to our labeled list
  for utteranceIndex in range(5, ConversationSize):
    maxSimilarity = 0
    distance = 0
    for index in range(utteranceIndex):
      if currSimilarity[utteranceIndex][index] >= maxSimilarity:
        maxSimilarity = currSimilarity[utteranceIndex][index]
        distance = utteranceIndex - index
```

```
        conversationList[ConvoIndex][utteranceIndex].append(maxSimilarity)
        currReformulation = 1 - maxSimilarity
        currEffort = currReformulation * np.log(distance) *
np.log(len(conversationList[ConvoIndex][utteranceIndex][1]))
        conversationList[ConvoIndex][utteranceIndex].append(currEffort)

    # Calculate number of utterances per person and the average similarity for each person in the
conversation
    FsumS = GsumS = FsumE = GsumE = FsumO = GsumO = Fnum = Gnum = 0
    for utteranceIndex in range(5, ConversationSize):
        if conversationList[ConvoIndex][utteranceIndex][0] == "f":
            FsumS += conversationList[ConvoIndex][utteranceIndex][2]
            FsumE += conversationList[ConvoIndex][utteranceIndex][3]
            FsumO += np.log(len(conversationList[ConvoIndex][utteranceIndex][1]))
            Fnum += 1
        if conversationList[ConvoIndex][utteranceIndex][0] == "g":
            GsumS += conversationList[ConvoIndex][utteranceIndex][2]
            GsumE += conversationList[ConvoIndex][utteranceIndex][3]
            GsumO += np.log(len(conversationList[ConvoIndex][utteranceIndex][1]))
            Gnum += 1
    FavgS = FsumS / Fnum
    GavgS = GsumS / Gnum
    FavgE = FsumE / Fnum
    GavgE = GsumE / Gnum
    FavgO = FsumO / Fnum
    GavgO = GsumO / Gnum

    # Print some useful information to the spreadsheet
    row = [ConvoIndex+1, Gnum, Fnum, GavgS, FavgS, GavgE, FavgE, GavgO, FavgO]
    append_list_as_row('NewSimilarityResults2.csv', row)

    # Remove the conversation we just looked at from the giant array
    bigArray = bigArray[ConversationSize:]  # Remove rows
    bigArray = np.delete(bigArray, np.s_[:ConversationSize], 1)  # Remove columns

    # Iterate to next conversation
    ConvoIndex += 1
```

## Appendix C: Operational Definition of Repair Initiation

For the purposes of this thesis, a repair initiation is a turn in talk that identifies trouble (i.e. a need for clarification) in a preceding turn or turns uttered by an interlocutor. The trouble in question need not be an error or mistake in any sense other than that it is identified as such by the repair initiation. Conversely, not all identifiable errors in conversation are followed by repairs (Schegloff et. al, 1977). The repair initiation may do nothing more than identify the presence of trouble (e.g. “what?”) or go so far as to correct the trouble entirely (i.e. require only confirmation by the interlocutor), as in the following example.

G *do you have white water*

F ***er well that'll be the rapids***

G *yeah that's true*

**How to Distinguish a Repair Initiation from a Request for Additional Information**

The trouble to which a repair initiation refers may concern information that was explicitly uttered by the interlocutor, or information that was not yet uttered at all, so long as the repair initiation refers to an utterance of the interlocutor. For example, the following utterance by F continues (i.e. implicitly refers to) the previous utterance by G and is therefore called a repair initiation even though it does not concern information explicitly uttered by G.

G *i want you to take your line from the bottom of the lagoon*

F ***to there?***

On the other hand, requests for additional information without reference to a previous utterance, such as the following, are not considered repair initiations.

G *just go a couple of steps*

F *mmhmm*

F ***have i to pass pass by the waterhole***

Some repair initiations do not explicitly identify trouble with the interlocutor’s utterance, but do so only implicitly by stating difficulty with its content. Take the following for example.

G *down about an inch so you come you've come off the rope bridge and you drop down about an inch*

F *right*

F ***erm i don't know if i can do that without going in the river actually***

In the above example, F's repair initiation does not explicitly indicate a need for clarification of G's instruction. Nevertheless, the fact that F cannot go through with the instruction without going in the river is clearly meant to indicate that there must be something wrong with the instruction, and G should either try again or confirm that F is supposed to go through the river.

**How to Distinguish a Repair Initiation from a Relevant Next Turn**

Some turns inadvertently identify trouble by virtue of their incongruity with a previous utterance. Clark & Brennan (1991, p. 225) refer to this phenomenon as a "relevant next turn" and give the following example.

A *Did you know mother had been drinking -*

B *I don't think mother had been drinking at all.*

A's utterance is formulated as a question and therefore anticipates an answer by B. The fact that B's next turn is evidently not an answer to the question indicates to A that B has not understood his last utterance. Though many repair initiations are also relevant next turns in that they identify trouble beyond that to which they refer, not all relevant next turns are repair initiations. The key difference is that repair initiations refer to trouble in a previous turn in a way that can be reasonably construed as intentional, whereas relevant next turns are not designed to identify trouble either explicitly or by implicature—they do so automatically by virtue of their incongruity.

**How Complete Must a Turn be in Order to be Called One Repair Initiation?**

Repair initiations that are rescinded by their utterer before the completion of the turn are not considered repair initiations, as in the following example.

G *and then come along sort of sweeping up to right underneath the outlaws' hideout so you're crossing the whole page*

F *right okay*

F ***that's ab--*** *oh you don't have the fort*

In the above example, F utters a false-start of what seems to be a repair initiation (bold) before rescinding it when he realizes that the way in which he planned to refer to the trouble would be ineffective (because G does not have a fort). Compare this with the following example, in which F does not complete her turn, but also does not rescind it. Note that despite the incomplete repair initiation, F has successfully signalled the existence of trouble, as evidenced by G's clarification in the next turn.

G *it's going to the right of the parked van underneath the lef-- the very left-hand point of the east lake*

F *right*

F ***so it's between… ehm***

G *it's on a level with the picket fence*

Because F's utterance was not rescinded and therefore successfully signalled trouble, the above example is considered one repair initiation.

**How to Differentiate Adjacent Repair Initiations**

Multiple repair *initiations* may occur within a single repair *sequence*, as in the following example.

G *ehm the next one you've got to go to... can you see springboks on yours*

F *no*

**F** ***where's that in relation to highest viewpoint***

G *it's right next to it virtually*

**F** ***to its right or left***

G *to its left*

**F** ***on the same level as highest viewpoint yeah***

G *just slightly below it*

F *okay right*

Although all three bolded turns refer to the same trouble source (i.e. concerning the location of the springboks), they are formulated independently and are therefore counted as three separate repair initiations. Multiple repair initiations may even be nested (i.e. a second initiation appears before the interlocutor's response to the first), as in the following example.

G *we've come down from a caravan park*

F *mmhmm*

G *and we're doing a "u" shape*

F *mmhmm*

G *okay on the and then you're going to proceed north right*

F ***hang on where's the "u" shape… just underneath the mill?***

F *just underneath the caravan park?*

G *no no*

G *yeah*

In the above example, F utters two specific offers in a row without waiting for a response in between. Because one is not a correction or rescission of the other, these are counted as two separate repair initiations. Conversely, a self-repair of a repair initiation turn is not considered to be an independent repair initiation, as in the following example.

G *right have you got have you got a fallen you've got a fallen cairn haven't you over to the right-hand side of the page*

F ***right at the very bottom***

F ***eh sorry in the middle i mean***

G *no*

G *in e-- up sort of to the right-hand northeast of the granite quarry*

Because the second bolded turn is a clarification (or in this case, correction) of the first, rather than a separate offer, the two are counted as one repair initiation. The following is a similar example.

G *you'll see a diamond mine on your map*

F *mmhmm*

g *avoid the diamond mine*

f *due south*

g *going due south followed by east*

f ***to the left of the diamond mine***

f *__to the west__*

g *east the right*

f *so you want me to go east then south*

g *no*

g *south then east*

Because the second bolded turn is a clarification of the first, the two bolded turns are counted as a single repair initiation.