RUNNING HEAD: INDIVIDUAL DIFFERENCES IN METAPHOR USE

Investigating Individual Differences in Metaphor Use and Its Outcomes:

Research Questions, Measurement, and Findings

Adam K. Fetterman[1]

Katherine French[2]

Brian P. Meier[3]

[1]University of Houston

[2]University of Texas at El Paso

[3]Gettysburg College

Suppose that three people described Jane's sadness. Person 1 wrote, "Jane was feeling extremely sad." Person 2 wrote, "Jane was in a dark place." Person 3 wrote, "Jane's sadness weighed her down." Person 1's statement is literal. Persons 2 and 3's statements are metaphorical. Jane is not literally in a dark place, nor does sadness have a physical weight. Metaphors such as these are common in language (Gibbs, 1994). So common, in fact, that people often do not realize they are using a metaphor (Thibodeau et al., 2017). There is a reason for their ubiquity. According to scholars (e.g., Gentner et al., 2001; Glucksberg & Keysar, 1990; Holyoak & Stamenković, 2018; Lakoff & Johnson, 1980; Landau, 2017), metaphors reflect the way that people process abstract concepts and complex ideas. In other words, metaphors are the trace evidence of how people think about and understand abstract concepts (Lakoff & Johnson, 1980). Therefore, the use of metaphors and the use of different metaphors to describe Jane's sadness reveals something about how these people understand her sadness.

Structurally, a metaphor is the mapping between a "target" and a "source" (Ortony et al., 1978). The target is an abstract concept (e.g., "sadness") that is not inherently physical or concrete. The source is a concrete object or physical referent (e.g., "darkness" or "weight") that represents the target. This mapping is likely useful. For instance, if someone wants to explain what a rock is, they can just show an actual rock to others. A rock is a concrete object. However, one cannot hold up sadness and show it to others. Sadness is an abstract concept, made up of complex intrapsychic and physiological qualities. Therefore, to understand what someone is feeling, we might liken sadness (the target) to darkness (the source). Doing so makes the idea, or qualia, of feeling sadness more concrete and understandable.

There are entire literatures, journals, and subfields whose primary focus is on metaphors. Some focus on the ubiquity of metaphor in language (Kövecses, 2005) and its use in rhetoric

(Cameron & Low, 1999). Others focus on the comprehension of metaphors

(Holyoak & Stamenković, 2018), the detection of metaphors in natural language (Shutova,

2010), and the use of metaphors to make inferences about target domains (Gentner & Gentner,

1983). Others still look at the cognitive architecture and impact of metaphoric mappings (Landau

et al., 2010). However, a relatively neglected area of metaphor research is on the psychological

functions and outcomes of metaphor use (Fetterman et al., 2016).

Investigating the functions and outcomes of metaphor use is important. Thinking back to

the three people describing Jane's sadness, we note that one person used a literal statement and

the other two used different metaphors. On one level, the fact that one person used a literal

statement may mean that they have a more surface-level understanding of sadness compared to

the metaphor users. On another level, the two metaphor users may understand sadness in a

fundamentally different way. For example, understanding sadness as "darkness" versus "weight"

may lead to different solutions (e.g., changing life circumstances vs. finding one's burdening

thoughts in therapy). These may have important implications for psychology.

That metaphors allow us to peak into the black box of the processing mind suggests great

promise for the impact of their use on cognition, emotion, and behavior. Yet, much of the extant

work on metaphor use has not measured metaphor use for testing such psychological outcomes.

This chapter hopes to inspire interdisciplinary work in this direction.

There are seemingly infinite approaches to metaphor research, attempting to answer

seemingly infinite questions. Metaphor research transcends research fields. Therefore, we cannot

hope to cover this topic in a comprehensive way. Given our backgrounds, we will primarily

approach the topic from a personality and social psychological point of view. In the following

sections, we will identify some broad research questions and describe what types of metaphor

use measurements will help answer them. Then, we will provide examples of how to measure metaphor use for these purposes and examples of how researchers have done so in the past.

The Correlates & Consequences of Metaphor Use

We breakdown the research questions into two goals. First, we will describe research questions and measurement regarding general metaphor use – i.e., individual differences in the relative frequency of metaphors use, rather than the types of metaphors used. Second, we will describe research questions and measurement regarding specific metaphor use – i.e., individual differences in the metaphoric sources used to describe specific target concepts. Again, we cannot hope to cover all of the research questions each discipline has regarding metaphor use, but we hope those we cover here will inspire new lines of research across many fields.

Individual Differences in General Metaphor Use

The basic logic of most metaphor theories is that metaphors help people understand abstract concepts. However, like most other psychological constructs (e.g., personality traits), there are individual differences in people's tendencies to use metaphors in general (Fetterman et al., 2016). Therefore, there are likely quantifiable psychological differences between those who use metaphors more versus less. These differences then provide insight into the functions of metaphor use. Recent work supports this assumption. For example, Fetterman et al. (2016) found that those who use metaphors more often scored higher in emotional understanding and that using metaphors versus literal statements led to reductions in negative affect. Further, Fetterman et al. (in press) found that the use of metaphors in daily life was positively associated with empathy and perspective taking. These investigations revealed emotional and interpersonal functions of metaphor use, respectively.

Recent research also suggests an existential function of metaphor use. Baldwin et al. (2018) had participants use metaphors versus literal language to describe their lives. The metaphor users scored higher in meaning in life than the non-metaphor users. However, we consider this indirect because it used a manipulation of metaphor use, rather than measuring metaphor use as it naturally occurs. Doing so may have forced participants to think about life in a novel way, which would question whether the effect was really about metaphor use. Instead, it would be preferable to complement this work with measures of metaphor usage tendencies to assess to what extent these metaphoric processes are functional.

While we have evidence for various functions of metaphor use, this area is in its infancy. Given the link between emotional understanding and metaphor use, future work should test clinical theories implicating the efficacy of metaphor use in practice (Bucci, 1997). In the interpersonal realm, if metaphor users are more empathetic, they may be good relationship partners. Further, since people are motivated to find communal relationship partners (Abele & Brack, 2013), a person's tendency to use metaphors might serve as a signal to others that they are empathetic and caring (Fetterman et al., in press). Additional research questions might include whether metaphor use correlates with meaning in life in varying situations or whether frequent metaphor users solve problems more creatively.

To investigate each of these general metaphor use questions, researchers need a measure of the quantity of metaphors each individual uses in general (i.e., in a trait sense) or in a specific instance (e.g., to talk about life). For example, if participants write about their day for eight minutes, the researcher needs a "metaphor use score" for each participant. Analyses would then indicate whether these scores predict scores on the measured psychological outcome of interest.

Measuring general metaphor use is more complicated than it seems and different approaches bring about different challenges. We describe three approaches, here. The gold-standard approach is counting by hand. However, this is the most time-consuming. Therefore, we will describe two alternatives that allow for relatively easier, but less precise measurement: self-report questionnaires and natural language processing (NLP). Starting with the gold standard approach, we lay down the gauntlet for what the alternative methods must achieve. In doing so, we describe prior research using each approach, as well as their promise and limitations.

Counting by hand approach. The "counting by hand" approach has received the most attention and is the most widely used approach (for thorough treatments of the history of this approach, we suggest starting with Cameron & Low, 1999 and Crisp et al., 2002). Counting by hand involves human coders going through language samples word-by-word and identifying the existence of metaphors. The design for every method of coding starts with the research question (Crisp et al., 2002; Pragglejaz Group, 2007; Kimmel, 2012). For example, the research question will determine if the coders are counting all metaphors or only those related to specific targets. In the former case, there might be neutral writing samples and the coders are looking for participants' tendency to use any metaphors at all. In the latter case, one might be interested in how many metaphors participants use referring to a specific topic (e.g., alcohol or anger).

Let's say that the research question is, "Do those who tend to use more metaphors score higher in emotional understanding?" In this case, one could use a simple counting procedure. However, counting metaphors is anything but simple. There are a number of issues to deal with. Of most importance, the coders must understand what counts as a metaphor. This is the biggest challenge to any type of metaphor measurement. To make this distinction, the coder must identify the context. For example, suppose someone wrote, "he is a shark." If the writer was

describing a fish, this is a literal statement. If, however, the writer was describing a lawyer, this is a metaphoric statement. This is a simple example, but metaphors are often complex (e.g., "that lawyer acts as if he was a shark") with new versions of the same metaphor uttered all the time (e.g., "he's an apex predator in an ocean of guppies").

Further complicating things, people sometimes use metonymy, in which they do not mention the target with the source (e.g., "I'm going to see the shark") and dead metaphors, which are metaphors that are so common that the source and target are synonymous (Ortony et al., 1978) and have lost their imagery (e.g., table leg; Paivio & Walsh, 1993). Further, the broader context of the words and phrases can be more or less ambiguous. If the task is to describe the lawyer, the context is easy. If it is to "describe your day," the person could be talking about a fish or a lawyer. Without proper training on all of these issues, the coders will be unreliable.

There are many ways to count metaphors. The method that we have found useful is similar to Metaphor Identification Procedure (Pragglejaz Group, 2007). Here, the coders get a sense of the context and then go word-by-word to identify metaphoric uses of words – in difficult cases coders compare word use to the definition in an agreed upon dictionary. Once the rater has gone through a sentence, they must decide if there is a metaphor present. To deal with dead metaphors, when we train raters, we suggest that they make a judgement call about whether the metaphor creates an image in their mind or if they can trace the source back to its original meaning. Once the rater has decided how many metaphors are present in the sample, they assign a frequency count score for that sample. As an additional measure, we often have the raters indicate how metaphoric they feel the sample was.

As an example, let's take the sentence, "Jane was in a dark place." The first word is "Jane." We can assume that "Jane" is a human, so any non-human actions will indicate a

metaphor. Then, we look at "was." The word "was" could be part of a metaphor, so we need to consider the context. Jane was "in a…" Humans can be in things literally or metaphorically. Jane was in a "dark place." Humans can also be inside dark places, but the context, in this case, suggests the person is describing sadness and emotions are not light or dark or containers. As such, the coders mark this sentence as having one metaphor.

Let's say that the sentence was "Jane's sadness weighed her down." Here, we seemingly have two metaphors: "weight" and "down." One might argue that these are two separate metaphors or one complex metaphor. The team must decide what to do in this case. One option is give a score of two. Another is to give a score of one. A third option is to give a "complexity" score. Later, then, when data are analyzed one can have a frequency of complex metaphors score, in addition to the general frequency score. However, this has already changed the research question. Therefore, the research team must make these decisions ahead of time.

A final part of our typical procedure deals with interrater agreement. There will never be 1-to-1 interrater agreement, but it is important to make sure they are mostly on the same page. Therefore, we often start the counting procedure with the coders scoring the first five-to-ten samples and then meet to discuss difficulties and disagreements. We then hold regular coder meetings throughout the process to discuss and adapt to any further issues and difficult cases.

At the end of the counting procedure, each participant or sample will have a metaphor usage score. Higher scores indicate that the participant or sample uses metaphors at a greater frequency and vice versa. Then, to answer the research question, one would correlate such a score with scores on an emotional understanding measure.

Prior work using this approach. Fetterman et al. (2016) investigated whether individual differences in metaphor use predicted emotional understanding. The researchers created a

measure of individual differences in metaphor use (more below). As a validation, they collected writing samples from participants and counted the metaphors that each person had used. Participants wrote about their lives for eight minutes. The coders scored each essay using the counting method described above. They also gave a subjective rating of "how metaphoric" the participant was. Fetterman et al. (2016) predicted and found that those who used more metaphors and were rated as more metaphorical scored higher on the metaphor use measure. This counting procedure provided convergent validity to the metaphor use measure.

Another line of forthcoming research is investigating metaphor use in motivated interviewing (French et al., 2020). Here, the researchers used the counting by hand approach to investigate the impact of metaphor use on the outcomes of alcohol abuse interventions. Coders counted the number of metaphors used by both the patients and interviewers in motivated interviews. The researchers then used these scores to see if the amount of metaphors used by the patient, the interviewer, and their combination, predicted positive alcohol use outcomes three months later. When both used more metaphors, there were significant decreases in alcohol use. Alcohol use stayed the same when one or both used metaphors less.

The general idea in this example is that people struggling with addiction issues might have trouble understanding their addictions, because addiction is abstract (Shinebourne & Smith, 2010). If metaphors help people understand the abstract and complex, then we should see that using metaphors in these settings is associated with better outcomes.

Final thoughts on counting by hand. The most notable benefit of the counting by hand approach is that it is the most comprehensive way to measure metaphor use. Any other method will suffer in that not they will not capture all natural metaphors. Therefore, we suggest that if the researcher has the resources and time, this is the optimal strategy.

The limitations of the counting by hand approach are also clear. It is time-intensive, especially when there are many samples. In fact, it is time-intensive on two fronts. First, one must train coders to identify metaphors. Second, the coders need to code the samples and each coder needs to be on the same page. Therefore, multiple meetings are required.

The self-report questionnaire approach. The challenges of the counting by hand approach can make research on metaphor use unrealistic for researchers who lack the resources or time to go through such a rigorous process. Therefore, some have looked for alternative ways to measure metaphor use. One notable option is the use of self-report questionnaires.

Let's say that the research question is, again, "Do those who tend to use more metaphors score higher in emotional understanding?" Here, for the metaphor side, the researcher only needs a measure of relative metaphor use. One could create such a measure in at least two ways. First, one could create a measure that compares metaphor and literal phrases, and have participants endorse one or the other. Second, one could create explanatory items that help participants report their level of metaphor use. In both cases, we assume that those scoring high on these measures would also receive a high count of metaphors in the counting approach. However, creating questionnaires of metaphor use is more difficult than standard personality measures.

Much like the issues with training coders for hand counting, most people do not readily understand what a metaphor is, in a scientific sense. As such, researchers might create a metaphor use measure that avoids confusion. For example, one can create a list of items that force participants to endorse metaphoric versus literal phrases. The phrases may not be exactly as the participant would use them, but they are endorsing those "types" of statements. Each participant ends up with a sum or average score of metaphoric phrases endorsed.

Alternatively, one could create a measure that describes metaphors and that uses numerous terms for metaphors. Here, participants respond to items in the context of a clear and plain description of metaphors, alongside an example. Assuming that the description and example give participants a general sense of what a metaphor is, one can ask them how metaphoric they are on a Likert type scale. Using numerous items also helps with comprehension. Reworded items using different terms (e.g., "figurative" language) and different uses (e.g., thinking versus speaking) can provide further context for the participant.

Both types of self-report measures approximate people's general metaphor use. The researcher then investigates correlations between scores on these measures and measures of emotional understanding (or the psychological variable of interest).

Prior work using this approach. Fetterman et al. (2016) investigated whether individual differences in metaphor use was predictive of emotional understanding. They developed the Metaphor Usage Measure (MUM). Initially, they generated 60 paired metaphoric (e.g., "She uses her head") and literal (e.g., "She is a rational person") phrases. Then, they pared these items down to a 30-item scale that was internally reliable and valid. For each item, participants endorsed which statement they would be more likely to use. While the participant may not have recalled a time in which they used the exact statement, they should have a sense of being more likely to use the metaphoric or literal statement. Scores on this measure significantly correlated with frequency scores from a counting by hand procedure.

Fetterman et al. (2016) found that those scoring higher on the MUM also scored higher on two measures of emotional understanding. These findings support the idea that metaphor use is associated with the understanding of abstract concepts and that metaphor use serves an emotional function. Interestingly, they also found that those scoring higher (vs. lower) on MUM

also reported being nicer on days in which they ate sweet food. This finding supports the idea that the MUM is measuring something akin to a "metaphoric mind." It does so because the link between sweet tastes and friendliness is a common and impactful metaphoric mapping (Fetterman et al., 2017; Meier et al., 2012; Ren et al., 2015). High versus low scorers on the MUM appear to hold this metaphoric mapping to a stronger degree (Fetterman et al., 2016).

Another line of research investigated whether daily metaphor use was associated with daily empathy and perspective taking (Fetterman et al., in press). To measure daily metaphor use, the researchers provided participants with a description of metaphors and an example of a metaphoric statement, contrasted with a literal statement. Participants then indicated how metaphoric they were for that day on three items. Using similar measures of empathy and perspective taking, they assessed day-to-day associations between daily metaphor use and these interpersonal variables. Indeed, they found positive associations. These findings support the idea that metaphors are operative in daily life and that metaphor use serves an interpersonal function.

Final thoughts on self-reported general metaphor use. The most notable benefit for the self-report questionnaire approach is the ease at which a researcher can collect data. Further, researchers can tailor questionnaires around their needs (e.g., metaphor use in certain contexts). As for general metaphor use, the MUM already exists and is openly available. An improved, revised measure using the explanatory approach is forthcoming (Fetterman et al., 2021).

While it is easier to use self-report questionnaires, they contain issues that are common amongst all self-report measures: An endless stream of biases hampers validity and reliability (Paulhus & Vazire, 2007). However, self-reporting metaphor use has its unique issues. First, we have already mentioned that people do not know much about metaphors, as we study them. Therefore, they may not understand what they are responding to, even with extensive

explanation. Second, if one has participants endorse metaphoric versus literal phrases, it is not always clear that they understand the metaphors. This is particularly problematic in cross-cultural samples. Some metaphors are either not present or have different meanings in other languages or cultures. To "feel blue" is to be drunk in Germany and sad in the United States.

The latter issue brings about another overall problem with self-report questionnaires. In the counting by hand approach, the researcher is measuring metaphors as they naturally occur. Self-report questionnaires force metaphors upon the participant. Therefore, even if they select only literal statements, they may still use metaphors frequently. The metaphor they prefer might not be an option. One can alleviate this issue to a certain degree, but the counting by hand approach is still the gold standard for studying individual differences in metaphor use.

The Natural Language Processing Approach (NLP)

A goal of this chapter is to inspire interdisciplinary research on metaphor use. This is particularly so when it comes to incorporating cutting-edge NLP approaches to quantitative social sciences. However, the NLP approach simultaneously holds the most promise, but also faces the highest hurdles for studying individual differences in quantitative social sciences.

NLP is rooted in linguistics and computer science (Kennedy et al., this volume). In a broad sense, the goals of NLP are to analyze natural language data for any features of interest. Most social scientists will be familiar with the Linguistic Inquiry and Word Count (LIWC; Pennebaker et al., 2015), which is a basic tool for NLP. LIWC runs through text and identifies words or phrases fitting preset language categories or "dictionaries" (e.g., "pronouns," "positive emotions," and "religion"), counts them, and provides the percentage for each category, as a score. However, NLP goes far beyond this basic tool, as is apparent in other chapters of this volume (see Kennedy et al., for example). In many cases, NLP uses machine learning and neural

networks to process large amounts of text for anything from detecting parts of speech and speech recognition to sentiment and word meaning. There has also been decades of work in NLP in the detection of metaphor (for review see, Shutova, 2010).

The general challenges to NLP are the same as the hand counting approach, but with the added challenge of training algorithms, which arguably do not have the contextual knowledge of humans (e.g., the "lawyer-fish is a shark" problem x infinity)(Welch, Azab, & Mihalcea, this volume). Even with all of these challenges, NLP research has trained some relatively accurate algorithms for detecting metaphors in large corpora (Shutova, 2010). For example, Neuman et al. (2013) created three algorithms with an average precision detection rate of 71% based on hand-coded norms. Even so, we will not review these methods for three reasons. First, we do not have the expertise to evaluate these algorithms. Second, the goals of NLP research on metaphors are beyond the scope of the current chapter, which is about individual differences and psychological outcomes. Third, none of these NLP techniques provide what social scientists need to answer the type of research questions we have discussed thus far. Specifically, no work that we are aware of has created a program or algorithm that can process text from numerous samples and return analyzable scores for each participant for purposes of quantitative social science research. This also means that this area is wide open for exploration and collaboration.

With these complications in mind, we will introduce two methods with potential to serve our purposes, albeit using less sophisticated or precise NLP methods. Each one will have strengths and weakness, but they both generally provide an approximation of metaphor use. However, we should reiterate that there is little evidence for the validity or reliability of any of these methods. As such, what follows is mostly speculative.

Let's say that our research question is, once again, "Do those who tend to use more metaphors score higher in emotional understanding?" We have writing samples and we have measures of emotional understanding. To score the writing sample for metaphors, we want to use an NLP tool to return a score for "metaphor use." One could either get a score for a proxy language feature or find as many basic metaphor fragments as possible and ask the program to find how many times those fragments have been used.

The first option is what we call the "proxy method." The proxy method uses language features as serviceable proxies for metaphors. That is, metaphors have several properties and uses. For example, the simplest property is the necessary component of concrete words (the source) to describe the target. One might conclude, then, that a program returning a "concreteness" score is an indirect way to measure metaphor use. However, the precision of a concreteness score depends on the writing sample. If the participants in our study wrote about their emotions, then the concreteness of the writing sample may be more likely to suggest the presence of metaphors than if they were writing about any topic. The proxy method, in theory, offers an indirect way to get metaphor frequency score for the samples.

The second option is what we call the "fishing method." In this case, the researcher collects as many basic metaphor fragments as possible and uses them as a "category" for the program to detect. What we mean by "basic" metaphor fragment is something like "destroyed him", or "used her head." Even more precise would be fragments that include direct metaphoric description, such as "like/as a rock" or "is a beast." The general idea with the fishing method is that, while a program will not detect every metaphor, if it has enough metaphor fragments, it should "catch" at least some when processing large samples. And, the fishing method *should* "catch" more metaphors in a high metaphor user than a low metaphor user.

Prior work using this approach. We are unaware of any work leveraging the proxy method, meaning its utility is yet to be tested. However, two investigations have employed the fishing method. Nelson and Thorne (2012) created a dictionary based on their own coding. They then used the LIWC to detect source categories for friendship formation. The LIWC did detect all the source categories in the text. However, it also incorrectly detected 10% of words that were not metaphorical. As such, they focused on their hand-coded data, which we describe later.

Akpinar and Berger (2015) investigated whether sensory metaphors were more culturally successful than literal statements. To do so, they employed the fishing method. They first compiled a list of sensory metaphors and their literal counterparts, comprised of 2-grams – a 2-gram is a two-word phrase (e.g., "bright future"). They focused on sensory adjectives (e.g., "bright") paired with nouns (e.g., "future") for the metaphors and literal counterpart adjectives (e.g., "promising"). They then used Google Books and measured the frequency of sensory metaphors and literal counterparts for each year from 1800-2000. Sensory metaphors were present at a higher rate across those 200 years than literal phrases and non-sensory metaphors. Akpinar and Berger concluded that our senses shape language and culture and metaphors are the trace evidence of this. The glaring limitation here is that the small set of precise metaphoric and literal 2-grams in their dictionary means that they did not detect a majority of metaphors.

Final thoughts on NLP for detecting general metaphor use. The largest benefit of the NLP approach is that it is quick, and it can process large amounts of data. Indeed, counting metaphors by hand in millions of books over 200 years is an insurmountable task. In this light, even if Akpinar and Berger (2015) were unable to detect even a quarter of the naturally used metaphors, NLP methods allowed them to test a hypothesis that would otherwise not have been

tested. An additional benefit of NLP is that it takes some of the human error out of the coding process and self-report questionnaires (Boyd et al., in press).

There are also many challenges to using the NLP approach. Most importantly, it is difficult for NLP to detect all the instantiations of metaphor use. As such, they will miss some metaphors and over-detect others. Using the methods we mentioned, researchers may have to create their own dictionaries, which is itself time-consuming and, given the imperfect results, might not be worth the trouble (Nelson & Thorne, 2012). Finally, whether using the proxy method or the fishing method, NLP is currently at best imprecise.

Individual Differences in Specific Metaphor Use

A large portion of the extant literature on the psychological impact of metaphor use has focused on the use of specific source domains and its impact on decision-making or behavior. For example, Hauser and Schwartz (2015) found that using war-metaphors to describe cancer treatment leads to poorer preventative measures. Additionally, Thibodeau and Boroditsky (2011) found that the specific source used to describe crime (beast vs. virus) leads to differential support for law enforcement policies (see Jordan, this volume, regarding political messages). Additional research questions might include whether using different metaphors for God relates daily behavior, if using different metaphors for relationships (e.g., a journey vs. a battle) leads to differing relationship outcomes, or whether using different metaphors for alcoholism leads to different outcomes in recovery.

Prior research that has attempted to answer such questions has mostly relied on experimental manipulations of source domain use, such as priming (Landau et al., 2010). While doing so has its benefits, it does not allow researchers to identify the metaphors that participants naturally use, which can be limiting in two ways. First, it forces participants to use metaphors

they would not typically use, which is good for developing interventions, but may not represent realistic views of how people process concepts. Second, this method limits the effects to specific metaphors and misses others that may have stronger or more important outcomes.

To answer these types of research questions from an individual differences perspective, researchers need to be able to identify the metaphoric sources that an individual is using to describe a target concept. For example, if the researcher has participants describe God, they would need a way to get a score for each source category. Analyses would then indicate whether using specific metaphors predicts scores on the psychological outcome of interest.

Much like the process of measuring the frequency of metaphor use in general, measuring specific metaphoric source use is challenging. Again, we will describe three approaches below and start with coding by hand (i.e., the gold standard). We will then describe self-report and NLP methods, as alternatives. Since many of the methodological intricacies, limitations, and promise for these approaches are the same as the general metaphor use approaches, we simply will refer back to prior sections for clarity, rather than repeat ourselves.

Coding by hand. In contrast to the *counting* by hand approach, the coding by hand approach involves the coder identifying metaphors and then noting the specific sources used to represent specific targets. Again, the design starts with the research question (Crisp et al., 2002; Pragglejaz Group, 2007; Kimmel, 2012).

Let's say that our research question is, "Does using a particular source to represent a life relate to differences in meaning?" In this case, we must go beyond simply counting metaphors. Further, in coding, the researcher is more often interested in a single target domain (e.g., "life", "cancer", or "God"). The target domain of interest, then, will determine which metaphors are coded for their sources (e.g., metaphors unrelated to "life" are not coded).

There are seemingly endless ways to code metaphors, but a method we have used is similar to Kimmel (2012), minus the qualitative coding software. Let's say that there is a large sample of participants who have described life. Coders can code these responses for the metaphoric sources used to describe life. The researcher can take a bottom-up, top-down, or combined approach. In the bottom-up approach, the coders create categories of sources as they appear in the samples: If the first description of life involves "a path," they will create a code called "journey" and then give a point each time a journey metaphor is used in the samples. In the top-down approach, the research team creates the likely source categories based on their knowledge of that area: If the questions is about life metaphors, there might be "journey," "balance," and "battle," source categories. The coders then decide which source categories, if any, are present in the sample. When using the top-down approach, one typically allows for bottom-up additions to the source categories (a combined approach): If raters find that "sea" metaphors are common, the researchers add a "sea" category. In the end, each participant has score for each source category (e.g., a "1" for present and "0" for absent). The researcher can then investigate correlations between these scores and psychological variables of interest.

In the end, the researchers should have a good idea of what types of sources people use for a specific target. Of course, the more the coders code, the richer the dataset will become. That is, instead of focusing on a single target, the coders could code all the targets present in the text. Kimmel (2012) provides tips for making the most of coding, with relatively little extra cost.

Prior work using this approach. There has been much work in various fields using the coding approach, particularly in studies of discourse (e.g., Deignan et al., 2017). However, less work has focused on connecting the use of specific metaphors with psychological variables and outcomes. We provide a couple notable exceptions.

First, Moser (2007) was interested in the metaphors that people use to conceptualize their "self" as an entity in the world. Participants took part in a semi-structured interview and described themselves. The researchers coded the transcripts for the sources used for different targets (e.g., ought, ideal, and social self). In a second study, participants completed the Big 5 personality inventory. What Moser found was that participants used 22 source domains when describing their selves. However, she noted that eight metaphors represented around 70% of all the metaphors used. Those who discussed their actual-self tended to use balance metaphors (e.g., "stable", "unbalanced"). She also found that when talking about the ideal-self, participants tended to use more "path" metaphors. Further, those scoring high on openness tended to use more self-as-a-container metaphors. Moser suggests that this makes sense, as containers can be "open" or "closed" and this imagery comports to an open-minded person.

Second, Nelson and Thorne (2012) were interested in how the metaphors used to describe one's relationship predicted the personality of the relationship partners. Participants brought a friend to the lab and discussed the beginning of their relationship. The researchers coded the transcripts for the types of source domains the partners used. Participants who had extraverted friends used more force-impact metaphors (e.g., "we hit it off", "right off the bat"). Nelson and Thorne suggest that this is likely because an extraverted partner quickly engages compared to introverts and this "feels" forceful. They also found that participants with a more introverted partner used more journey metaphors. They reasoned that people with an introverted partner see the relationship as growing more gradually: like a journey. These findings suggest that metaphors are informative for the perception of relationship development.

Finally, in one of our own projects, we were interested in the metaphors used to describe God. In one study, we (Fetterman et al., 2020) examined data from 3,000 participants who

described God. We hand coded the responses on 25 predefined source categories and found 16 commonly used source categories. Of those, four sources accounted for 70% of all sources used (human, male, power, & concrete entity). Interestingly, spiritual folks were also more likely to use "journey" metaphors. Those from Abrahamic religions were more likely to use male and power than those from other religions (e.g., Buddhism). These findings suggest that people have fundamentally different ways of understanding God and we predict these differences will be associated with important psychological outcomes in daily life, including meaning in life.

Final thoughts on coding by hand. The main benefits and challenges for counting by hand are the same for coding by hand. It is time intensive, not only to perform the coding, but also to train the coders. However, done systematically and thoroughly, coding a dataset by hand can lead to a rich dataset that a researcher can use for numerous exploratory purposes.

The self-report questionnaire approach. Let's say that our research question is, again, "Does using a particular source to represent life relate to differences in meaning?" In this approach, we need participants to self-report which source (e.g., "journey" vs. "battle") they use to understand a specific target domain of interest (e.g., life). We suggest two ways of doing so.

If the researcher knows of the most prevalent sources used to describe the target and these sources appear to contrast each other, they may just need a single item. For example, people have used the head and the heart metaphorically (and sometimes literally) for hundreds of years as the location of the self (Swan, 2009). As such, one could create a measure where participants have to choose between the head and heart as the location of their sense of self. Responses to this item suggest that the participants use a specific metaphor to understand their sense of self, which should be predictive of metaphorically consistent psychological outcomes.

The researcher may not be aware of the most prevalent sources used to describe a target, or know many. In this case, they could create a measure in which participants indicate their level of endorsement for multiple source domains. For example, if we want to know how people conceptualize life, we could come up with the most common sources used to describe life and create items for each one (e.g., "journey", "battle", and "balance"). Then, participants indicate the extent that they use each source. The researcher can then use the relative endorsement of each source category to predict metaphorically consistent outcomes (e.g., meaning in life).

Prior work using this method. Fetterman and Robinson (2013) investigated whether endorsing one metaphor for the self over another is predictive of psychological outcomes. The researchers noted that the common distinction between the self as being located in the head or in the heart mapped on to common personality traits and psychological tendencies. In a series of studies, participants indicated whether their sense of self is located in their head or in their heart. Head-locators scored higher on measures of rationality, logic, interpersonal coldness and had higher grade point averages, than heart-locators. Heart-locators scored higher on measures of emotionality, intuition, and interpersonal warmth, than head-locators. Fetterman et al. (2020) also found that heart-locaters scored higher in God-belief than head-locators, while Adam et al. (2015) found the head-locators are more independent than heart-locators.

Final thoughts on self-report questionnaires. As opposed to self-reporting metaphor use in general, self-reporting metaphoric source preference is an easier task for participants. Therefore, there is some deal of certainty in accurately measuring how someone metaphorically conceptualizes a target domain. Even so, it may be that one endorses a specific metaphoric source, but never uses it in language. This could mean that they do not process the target in the way that they indicated. Further, if the person's source category of preference is not a response

option, they may select an option that they do not use. Even with these limitations in mind, the self-report questionnaire approach is a viable alternative when coding by hand is not plausible.

The natural language processing approach (NLP). Let's say that our research question is, one last time, "Does using a particular source to represent life relate to differences in meaning?" Here, we will want our NLP tool to detect source categories. Again, decades of work in NLP has provided sophisticated solutions for these purposes (for review see Shutova, 2010), but there is still a gap between this work and the needs of quantitative social scientists. However, less sophisticated NLP methods, that involve heavy human involvement, might provide a good approximation. We suggest one such solution here.

The purpose of a dictionary in NLP tools is to tell the program to which categories words belong (Kennedy et al., this volume). So, if the program detects the word "she," it knows that "she" belongs in the pronoun and female categories, among others, and scores it as such. Therefore, if a participant is describing the target domain, "life," and the program returns a relatively high score on a "motion" category for a participant, one might conclude that this person uses "journey" as a source to describe life. The design of the study should take care that the participants' only task is to describe or talk about the target domain of interest (e.g., life). If "motion" is a category that a participant used often in describing life, it would make sense that the use of the "motion" category is in reference to life, specifically. The researcher could then use this score to predict scores on the psychological outcome of interest (e.g., meaning in life).

Prior work using this method. Renzi et al. (2017) investigated the types of metaphors used in newspapers to describe nuclear power and the emotional tone surrounding them. Inspired by NLP methods, in which algorithms detect when a word is not being used with its literal definition, the researchers extracted a word vector and manually selected words in the list that

would not be used literally in describing nuclear power (e.g., "addiction"). They also used a program to identify words that followed the phrase "nuclear power is like…", assuming that the word that followed was being used metaphorically. Using both methods, they found 18 sources, seven of which were common. Interestingly, some were used to denote the positive aspects (e.g., "renaissance") and the negative aspects (e.g., "sickness") of nuclear power.

In the same study mentioned in the coding by hand section (Fetterman et al., 2020) we used LIWC (Pennebaker et al., 2015) to further support our hand coding results. To do so, we put all of the God descriptions into a single text file and analyzed it using LIWC's 2015 built-in dictionary. The most used categories in the LIWC results were similar to our hand-coded results (e.g., power, male, and human). It is important to point out, though, that the task was for participants to describe God. As such, we could ignore irrelevant LIWC categories and focus only relevant categories. Further, we could be somewhat assured that since the task was to describe a specific target domain, any descriptive LIWC categories were used metaphorically to describe the target in question (i.e., "God").

Final thoughts on NLP. Again, the benefit of using the NLP approach for these purposes is speed and ease. Further, we can eliminate some amount of human idiosyncrasies in the coding process and self-report questionnaires (Boyd et al., in press). However, the current limits of the NLP approach mean that it is best used in tandem with one of the other approaches.

The Future is "Looking Bright"

We provided a taste of recent trends in the study of metaphor use. The measurement of metaphor use will continue to expand as technology advances. We are particularly excited about the development of NLP techniques. Given the rise of machine learning and AI in computer and computational social sciences, more sophisticated tools for returning metaphor use scores are

becoming more likely. We imagine that NLP techniques will not only advance to the point that they do the counting and coding for us, but that they will lead to more sophisticated analyses of how and when metaphors are used. For example, imagine an algorithm that detects the conditions that lead to metaphor use (e.g., signals of ambiguity); or an algorithm that detects psychological outcomes after a metaphor is used (e.g., reduction in negative emotion). These may be far off, but this area is rapidly advancing. Until NLP programs come along, the counting and coding by hand approaches remain the gold standard. In addition to these advances, we are excited about the important scientific discoveries that will come from a variety of approaches to studying individual differences in metaphor use. We discuss two such directions below.

While we have provided evidence that metaphor use is associated with numerous positive outcomes, there are likely drawbacks. For example, as metaphors and mental imagery are inextricably linked (Ortony et al., 1978; Paivio & Walsh, 1993), we may see shared outcomes. Although, mental imagery is associated with benefits to memory storage (Paivio, 1969) and recall (Marks, 1973), it has also been implicated in false memories (Hyman & Pentland, 1996). Metaphors, then, might contribute to the creation of false memories. Indeed, the interruption caused by metaphors in social cognitive processes is well-established (Landau et al., 2010). It could be that metaphor use contributes to racial prejudice through linking black with bad (Meier et al, 2004; see also Caliskan & Lewis, this volume; Charlesworth & Banaji, this volume). Future research should investigate whether those who score higher in metaphor use are more prone to such potential negative outcomes.

Metaphor use may have drawbacks, but it is certainly adaptive. Akpinar and Berger (2015) argued for metaphor's evolutionary success in language. It may even be the case that metaphor use is genetic. Therefore, it may be of interest to look at the heritability of metaphor

use. For example, one could have monozygotic and dizygotic twins complete a measure of metaphor use. Should the former twins be more similar in their level of metaphor use than the latter, it would support the idea of a genetic component of metaphor use (Plomin et al., 1990). The exploration of the evolutionary basis of metaphor use has exciting prospects.

Conclusion

Metaphors in language are the trace evidence of how people process abstract and complex processes. As such, metaphors, like many language features, provide an important window into the black box of the human mind (Pennebaker, this volume). Much research has been dedicated to the study of metaphors and their uses. Less work, however, has focused on the psychological functions and outcomes of metaphor use. In service to increasing the latter output, we reviewed different approaches for measuring individual differences in metaphor use and discussed prior work that has adopted these approaches. We also acknowledged the limitations of these approaches. Overall, it appears that metaphor users tend to have a better grasp of their emotions and are empathetic. Further, the specific sources that people naturally use to describe target domains may have important psychological implications. We "look forward" to the "bright horizons" of future work on the topic of metaphor use and its measurement.

References

Abele, A. E., & Brack, S. (2013). Preference for other persons' traits is dependent on the kind of social relationship. *Social Psychology, 44*, 94-94.

Adam, H., Obodaru, O., & Galinsky, A. D. (2015). Who you are is where you are: Antecedents and consequences of locating the self in the brain or the heart. *Organizational Behavior and Human Decision Processes, 128*, 74–83.

Akpinar, E., & Berger, J. (2015). Drivers of cultural success: The case of sensory metaphors. *Journal of Personality and Social Psychology, 109*, 20–34.

Baldwin, M., Landau, M. J., & Swanson, T. J. (2018). Metaphors can give life meaning. *Self and Identity, 17*, 163–193.

Boyd, R. L., Pasca, P., & Lanning, K. (in press). The personality panorama: Conceptualizing personality through big behavioural data. *European Journal of Personality*.

Bucci, W. (1997). *Psychoanalysis and cognitive science: A multiple code theory*. New York: Guilford Press.

Cameron, L. & Low, G. (Eds.) (1999). *Researching and applying metaphor.* Cambridge: Cambridge University Press.

Crisp, P., Heywood, J. & Steen, G. (2002). Metaphor identification and analysis, classification and quantification. *Language and Literature, 11*, 55-69.

Deignan, A., Semino, E, & Paul, S. A. (2017). Metaphors of climate science in three genres: Research articles, educational texts, and secondary school student talk. *Applied Linguistics, 40*, 379-403.

Fetterman, A. K., Evans, N. D., & Covarrubias, J. J. (in press). On the interpersonal function of metaphor use: Daily metaphor use with empathy and perspective taking. *Social Psychology*.

Fetterman, A. K., Evans, N. D., Exline, J., & Meier, B. P. (2021). What shall we call God? An exploration of common metaphors for God. *Manuscript submitted for publication.*

Fetterman, A. K., Evans, N. D., & Meier, B. P. (2021). Individual differences in metaphor use: A revised measure, replications, and extensions. *Manuscript submitted for publication*

Fetterman, A. K., Juhl, J., Meier, B. P., Abeyta, A., Routledge, C. & Robinson, M. D. (2020). The path to God is through the heart: Metaphoric self-location as a predictor of religiosity. *Self and Identity, 6*, 650-672.

Fetterman, A. K., Meier, B. P., & Robinson, M. D. (2017). Dispositional properties of metaphor: The predictive power of the sweet taste metaphor for trait and daily prosociality. *Journal of Individual Differences, 38*, 175–188.

Fetterman, A. K. & Robinson, M. D. (2013). Do you use your head or follow your heart? Self-location predicts personality, emotion, decision making, and performance. *Journal of Personality and Social Psychology, 105*, 316-334

Fetterman, A. K., Bair, J. L., Werth, M., Landkammer, F., & Robinson, M. D. (2016). The scope and consequences of metaphoric thinking: Using individual differences in metaphor usage to understand how metaphor functions. *Journal of Personality & Social Psychology*, *110*, 458–476.

French, K., Fetterman, A. K., Najera, S., Evans, N. D., & Field, C. (2020). The use of metaphors in therapeutic interviews. *Unpublished data.*

Gentner, D., & Gentner, D. R. (1983). Flowing waters or teeming crowds: Mental models of electricity. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 99-129). Hillsdale, NJ: Erlbaum.

Gentner, D., Holyoak, K. J., & Kokinov, B. N. (Eds.). (2001). *The analogical mind: Perspectives from cognitive science.* Cambridge, MA: MIT Press.

Gibbs, R. W., Jr. (1994). *The poetics of mind: Figurative thought, language, and understanding.* New York: Cambridge University Press.

Glucksberg, S., & Keysar, B. (1990). Understanding metaphorical comparisons: Beyond similarity. *Psychological Review*, *97*, 3–18.

Hauser, D. J., & Schwarz, N. (2015). The war on prevention: Bellicose cancer metaphors hurt (some) prevention intentions. *Personality and Social Psychology Bulletin, 41*, 66–77.

Holyoak, K. J. & Stamenković, D. (2018). Metaphor comprehension: A critical review of theories and evidence. *Psychological Bulletin, 144*, 641-671.

Hyman, I. E., Jr., & Pentland, J. (1996). The role of mental imagery in the creation of false childhood memories. *Journal of Memory and Language, 35*, 101–117.

Kimmel, M. (2012). Optimizing the analysis of metaphor in discourse: How to make the most of qualitative software and find a good research design. *Review of Cognitive Linguistics, 10*, 1-48

Kövecses, Z. (2005). *Metaphor in culture: Universality and variation*. New York, NY: Cambridge University Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago, IL: The University of Chicago Press.

Landau, M. J. (2017). *Conceptual metaphor in social psychology: The poetics of everyday life.* New York: Routledge.

Landau, M. J., Meier, B. P., & Keefer, L. A. (2010). A metaphor-enriched social cognition. *Psychological Bulletin, 136*, 1045-1067.

Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology, 64*, 17–24.

Meier, B. P., Moeller, S. K., Riemer-Peltz, M., & Robinson, M. D. (2012). Sweet taste preferences and experiences predict prosocial inferences, personalities, and behaviors. *Journal of Personality and Social Psychology, 102*, 163–174.

Meier, B. P., Robinson, M. D., & Clore, G. L. (2004). Why good guys wear white: Automatic inferences about stimulus valence based on brightness. *Psychological science, 15*, 82-87.

Moser, K. S. (2007). Metaphors as symbolic environment of the self: How self-knowledge is expressed verbally. *Current Research in Social Psychology, 12*, 151–178.

Nelson, P. A., & Thorne, A. (2012). Personality and metaphor use: How extraverted and introverted young adults experience becoming friends. *European Journal of Personality, 26*, 600–612.

Neuman, Y., Assaf, D., Cohen, Y., Last, M., Argamon, S., Howard, N., & Frieder, O. (2013). Metaphor identification in large texts corpora. *PloS One, 8*, e62343.

Ortony, A., Reynolds, R. E., & Arter, J. A. (1978). Metaphor: Theoretical and empirical research. *Psychological Bulletin, 85*, 919–943.

Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review, 76*, 241–263.

Paivio, A., & Walsh, M. (1993). Psychological processes in metaphor comprehension and memory. In A. Ortony (Ed.), *Metaphor and thought. 2nd ed.* (pp. 307–328). New York, NY: Cambridge University Press.

Paulhus, D. L., & Vazire, S. (2007). The self-report method. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology.* (pp. 224–239). New York, NY: The Guilford Press.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015.*

Plomin, R., DeFries, J. C., & McClearn, G. E. (1990). *Behavioral genetics: A primer, 2nd ed.* New York, NY: W H Freeman/Times Books/ Henry Holt & Co.

Pragglejaz Group. (2007). MIP: A method for identifying metaphorically used words in discourse. *Metaphor and Symbol, 22*, 1–39.

Ren, D., Tan, K., Arriaga, X. B., & Chan, K. Q. (2015). Sweet love: The effects of sweet taste experience on romantic perceptions. *Journal of Social and Personal Relationships, 32*, 905-921.

Renzi, B. G., Cotton, M., Napolitano, G., & Barkemeyer, R. (2017). Rebirth, devastation and sickness: Analyzing the role of metaphor in media discourses of nuclear power. *Environmental Communication, 11*, 624-640.

Shinebourne, P., & Smith, J. A. (2010). The communicative power of metaphors: An analysis and interpretation of metaphors in accounts of the experience of addiction. *Psychology and Psychotherapy: Theory, Research and Practice, 83*, 59–73.

Shutova, E. (2010, July). Models of metaphor in NLP. In *Proceedings of the 48th annual meeting of the association for computational linguistics* (pp. 688-697).

Swan, T. (2009). Metaphors of body and mind in the history of English. *English Studies, 90*, 460-475.

Thibodeau, P. H., & Boroditsky, L. (2011). Metaphors we think with: The role of metaphor in

  reasoning. *PLoS ONE, 6*: e16782.

Thibodeau, P. H., Hendricks, R. K., & Boroditsky, L. (2017). How linguistic metaphor scaffolds

  reasoning. *Trends in Cognitive Sciences, 21*, 852–863.