RESEARCH UNLOCKS DOORS: BAYESIAN COGNITIVE MODELING OF METAPHOR COMPREHENSION IN INDIVIDUALS WITH AUTISM SPECTRUM DISORDER.

Mackenzie Kephart

Department of Cognitive Sciences, University of California Irvine

Mentors: Dr. Lisa Pearl, Zachary Rosen

April 2, 2021

Abstract

How does an individual with Autism Spectrum Disorder (ASD) comprehend metaphors, such as “The boy is a pig”? Behavioral data from Happé et. al (1993) and others show that a neurotypical (NT) child may select the intended meaning of a metaphor twice as often as a child with ASD in experimentally-controlled settings. In our experiment we used the Kao et al. (2014)’s Bayesian Rational Speech Act model of neurotypical metaphor comprehension. In an attempt to replicate ASD metaphor comprehension behavior, we altered the contrast parameter and the prior probability of the category. Our hypothesis was that we can alter the contrast parameter in order to skew the modeled listener’s representation of the information from their prior experience with the metaphor. Specifically, we wanted to replicate the qualitative pattern seen in the behavioral studies from Happe et al. (1993) and Pastor-Cerezuela et al. (2020), where the intended metaphorical interpretation of a metaphor will become less than half as probable than it is typically. Our study did find values of the contrast parameter and prior probability of the category, alone and in combination, that can replicate ASD behavior for each stimuli used. Future research should focus on the implications of these variables in terms of ASD comprehension behavior, or look further into the interaction of these two variables on the modeled output of ASD metaphor comprehension.

Keywords: metaphor, Autism Spectrum Disorder, contrast parameter, modelling

RESEARCH UNLOCKS DOORS: BAYESIAN COGNITIVE MODELING OF METAPHOR COMPREHENSION IN INDIVIDUALS WITH AUTISM SPECTRUM DISORDER.

Humans have the gift of gab; we are able to communicate in using spoken, written, and body language. However, current research is far from understanding the complexities of human language. Children begin using and understanding figurative language in their first few years of school instruction, suggesting the neurotypical (NT) brain can learn to comprehend non-literal language during childhood (Chabhoun et. al, 2015). On the other hand, individuals with Autism Spectrum Disorder (ASD) have known deficits in comprehending figurative language. The DSM-V characterizes Autism Spectrum Disorder as “persistent deficits in social communication and social interaction across multiple contexts, including deficits in social reciprocity, nonverbal communicative behaviors used for social interaction, and skills in developing, maintaining, and understanding relationships” (DSM V, 2013).

In this study, we were interested in understanding how an individual with ASD comprehends metaphors. Lakeoff and Johnson (2003) define metaphor as “understanding and experiencing one kind of thing in terms of another.” We focused on two types of metaphor: conventional and novel. A conventional metaphor is one where the metaphor has been heard so often that it can lose figurative value (ie: “time is money); in contrast, a novel metaphor is one that is not heard often and retains figurative value (ie: “daddy is a volcano”). Our goal was to understand the cognitive differences between neurotypical (NT) and ASD understanding of novel metaphors, which gives insight to the deficits in mental computations displayed by ASD individuals.

Research on this topic is necessary because less is known about figurative language comprehension as a whole in children with ASD, resulting in conflicting conclusions regarding the true nature of the cognitive deficits in ASD. Figurative language is used in everyday language, even if we don't realize it. If individuals with ASD are not properly understanding figurative language, they are also not able to properly communicate with their peers or understand instruction from their teachers as well as NT individuals. We need to continue studying ASD metaphor comprehension because this could have a huge impact on how to teach metaphor in special education primary school classes for ASD children. With a better understanding of these cognitive deficits in metaphor comprehension that individuals with ASD experience, we can create actionable, impactful solutions to help individuals with ASD recognize and process metaphors.

Here, we used computational cognitive modeling to investigate NT metaphor comprehension, and how it differs for ASD individuals. More specifically, we adapted a current computational cognitive model of NT metaphor comprehension (Kao et. al. (2014)) to capture how an individual with ASD processes and comprehends a novel metaphor. We gathered empirical data highlighting specific observable differences in novel metaphor comprehension for NT vs. ASD individuals (Chouinard (2017), Gold et. al. (2010), Happé (1993), Melogno (2017), Pastor-Cerezuela (2020)) In this way, we manipulated a current working model of NT metaphor comprehension to reflect ASD comprehension. Our objective was to have the model significantly reduce the likelihood of the intended metaphorical interpretation of an utterance.

The process of understanding a metaphor

Theoretical approach

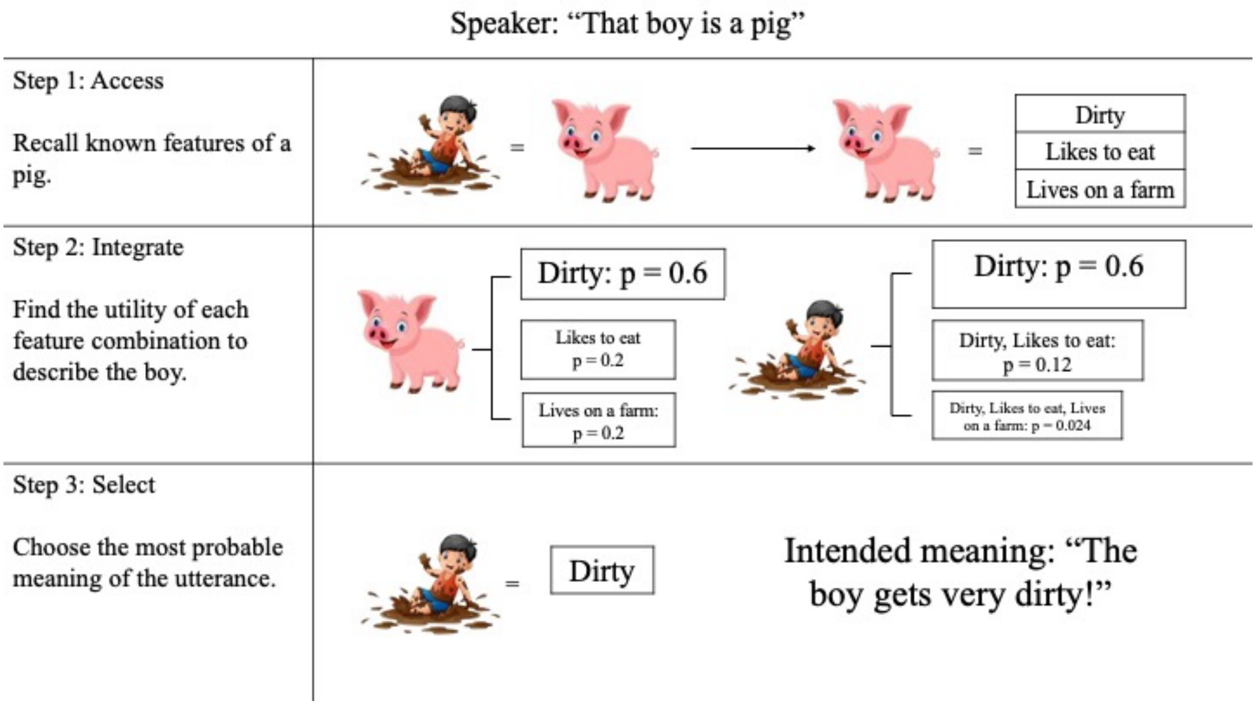
Empirical studies typically follow three steps to metaphor comprehension, best described by Chouinard et. al. (2018): Access, Integrate, and Select (see Figure 1). First upon hearing a metaphoric utterance like "The boy is a pig", the individual must access all relevant information from their semantic representations of the two objects being compared (here, "boy" and "pig").

In the simplified example in Figure 1, the relevant features from our representation of a pig are: dirty, likes to eat, and lives on a farm.

Second, the listener must integrate their semantic representations with external context in order to construct possible interpretations of the utterance. In Figure 1, we rely on visual cues to obtain information about the boy, in this case that he is playing in a mud puddle. Then, we are able to evaluate how valuable each pig feature (or combination of features) is (are) in terms of describing the boy. In our example, dirty was a valuable feature for a pig, so it had the greatest weight at 0.5. Since the boy is playing in mud, dirty is also useful to describe the boy, making pig feature combinations that use dirty more likely. However, knowing a pig is a farm animal typically is not helpful to us as there is no evidence the boy lives on a farm, making any combination with that feature less likely.

Third, the listener must select the appropriate interpretation/meaning. In Figure 1, if listeners had to choose between “the boy is (literally) a pig,” which would mean the boy is a dirty farm animal that likes to eat, versus “the boy is very dirty”, they should select “the boy gets very dirty.”

Figure 1.



Note: A simplified example of metaphor comprehension with "The boy is a pig" following the three steps described by Chouinard et al. (2016).

Empirical data of figurative language comprehension with children and adults, NT vs. ASD

Current behavioral studies find that individuals with ASD have a harder time rejecting the literal interpretation of a metaphoric sentence, and accepting the intended meaning (Kalandadze (2018), McKay & Shaw (2004), Melogno (2017)). Oftentimes, this atypical comprehension behavior is signaled by spontaneous comments or body language the ASD individual displays upon hearing the metaphor (McKay & Shaw (2004), Melogno (2017)).

For example, McKay and Shaw (2004) studied hyperbole estimation differences in children with and without ASD; they found that children with ASD did not comprehend hyperboles (ie. 'thousands of CD's) as exaggerations, and reported the speaker literally meant 'thousands', even though they were not as confident in their answers as NT children. Another study by Melogno et. al. (2017) presented children with novel metaphors (for example: "My daddy is a volcano") and noted that individuals with ASD laughed and accused the speaker of lying. These studies suggest that the child with ASD readily accepted the meaning of the sentence to be daddy is literally a volcano, even though they knew that there were fallacies in the literal statement. Yet, they displayed uncertainty, or misinterpreted the speaker's intent to say something non-literal as the intent to tell a joke. However, NT children understood that daddy can share features with a volcano, such as being burly, tall, or having a propensity to explode.

These studies and more have consistently found that individuals with ASD have deficits in identifying, processing, and correctly identifying figurative meaning across sentence type, including metaphor. Notably, ASD individuals often select literal interpretations for figurative language, see Kalandadze et. al. (2018) for a recent review. However, there is a lot of individual variation within an ASD diagnosis, which affects ASD interpretation behavior regardless of the nature of the task.

Hypotheses on the differences between NT and ASD individuals for metaphor comprehension

Theory of Mind. Theory of Mind (TOM) is typically described as knowing and understanding another person's mental states (Kalandadze (2018)). This includes being able to understand that another person might have a different mental state (ie. beliefs, knowledge, etc.) than your own. Researchers have attempted to measure TOM in ASD vs. NT individuals; however, results on these studies have been inconsistent as TOM can only be tested indirectly.

Hochstein et. al. (2018) tested TOM by asking individuals with ASD to complete a decision-making task that required them to compute scalar implicatures which accounts for a speaker's epistemic states. An epistemic state, in short, is knowing what another person (ie. the speaker) knows, which is an essential TOM function. They found that individuals with ASD were able to compute scalar implicatures, yet oftentimes ignored speaker's epistemic states when computing the implicature. Therefore, they were less accurate than NT individuals in the decision-making task.

For example, imagine a speaker and a listener that can see two red apples and a closed box. Then, the speaker looks inside the box at a third apple and says “Some of the apples are red.” When the listener is asked about the color of the third apple, the correct implicature is to indicate the third apple is not red, orelse the speaker would have said “All the apples are red.” However, an individual with ASD was more likely to ignore a speaker’s epistemic state (ie. the speaker knows the color of the last apple) and were more likely to respond “All the apples are red” because their own epistemic state is: all the apples *they can see* are red.

In terms of metaphor comprehension, these results suggest that an individual with ASD may understand that a speaker is using metaphoric language to send a non-literal message, but is unable to take the speaker’s knowledge into account when evaluating the goal of the speaker’s utterance.

Executive function. Researchers have also attempted to measure executive function to see the role it plays in metaphor comprehension. It’s possible there are fundamental differences in the brain of an individual with ASD that prevents them from identifying relevant information that would help them understand the speaker’s goal. If problematic executive functioning is preventing the listener from identifying relevant features of a target, or preventing integration of environmental context, then this could explain why an individual with ASD is less likely to select the appropriate interpretation of a metaphor.

Brain imaging studies have found that individuals with ASD tend to have more right brain activation and less communication between subcortical and cortical brain regions when hearing a metaphor, indicating a coarse (and uninformative) search for relevant information of the target (Chouinard (2017), Wang (2006)). Behavioral data from Gold et. al. (2010) supports the hypothesis that executive function prevents the identification of relevant features of a target, as they found individuals with ASD are more accurate at identifying two-word pairs that are literally related versus being related by a novel metaphor.

Still, Karsirer et. al. (2011) attempted to find a correlation between metaphor identification, metaphor production and executive functioning tasks; but their only significant result that was the majority of an individual's variance on these tests was accounted for by their score on a vocabulary test. Yet, this opens the door to the possibility that metaphor comprehension for individuals with ASD is affected by measurable aspects of language ability.

Language ability. While TOM and executive functioning are less consistent predictors of metaphor comprehension, current research is finding consistent patterns in measures of language abilities. Studies that compare metaphor comprehension in individuals with ASD matched to NT controls with the same chronological age (CA) or core language abilities (CLA) have found a similar pattern: CA-matched NT individuals perform the best, followed by CLA-matched NT individuals, who still significantly outperformed individuals with ASD (Chahboun 2015, Kalandadze 2018, Pastor-Cerezuela 2020).

This means when a NT child is matched to an individual with ASD of the same age, the NT individuals perform significantly better on a metaphor comprehension test. Furthermore, when a NT individual is matched with an ASD individual who has the same core language abilities (ie. the NT individual may be chronologically younger than the ASD individual or another form of language deficit) the CLA-matched NT individual still outperforms an individual with ASD, though they are both still outperformed by a CA-matched NT individual on a metaphor comprehension test. This pattern holds when tested for accuracy (how many metaphors were correctly interpreted) and response time (how quickly the individual was able to respond during the trial). This indicates there is a serious language deficit in individuals with ASD that impacts metaphor comprehension.

$$S_1\left(u \middle| g, \vec{f}\right) \propto \ e^{\lambda U(u|g,\vec{f})}$$

$$U\left(u \middle| g, \vec{f}\right) = \ \log \sum_{c,\vec{f}} \delta_{g(\vec{f})=g(\vec{f}')}\, L_0\left(c, \vec{f} \middle| u\right)$$

$$L_0\left(c, \vec{f} \middle| u\right) = \begin{cases} P(\vec{f}|c) & if\ c = u \\ 0 & otherwise \end{cases}$$

The cognitive computational model

We adapted Kao et. al. (2014)'s Rational Speech Act (RSA) model of NT metaphor comprehension to generate ASD metaphor comprehension behavior. As noted above, this means the ASD individual will choose the intended interpretation of a metaphoric utterance less than half the time as a NT individual. We utilized stimuli from our studies of interest as inputs to the Kao et. al. (2014) model, attempting to qualitatively replicate these studies' results.

In an RSA model, a *pragmatic listener* (L_1) reasons about what a *speaker* (S) is thinking, who in turn is imagining how a *literal listener* (L_0) would interpret an utterance (see Figure 2). This is sometimes called "recursive social reasoning", which links intuitively to TOM: in this case, a large part of the pragmatic listener's (L_1), reasoning process is about how someone else (S) is thinking about how a third someone (L_0) is thinking (Goodman, 2016).

The model calculates the interpretation of the utterance from the pragmatic listener's (L_1) point of view.

$$L_1(c, \vec{f}\,|u) \propto P(c)P(\vec{f}|c)\sum_{g} P(g)\, S_1(u|g, \vec{f})$$

P(c) is the prior probability that the boy belongs to a possible category. In our example from Figure 1 "The boy is a pig", this corresponds to the probability that the boy is a pig (P(pig)) or that the boy is a human (P(human)). In this case, P(human) will be weighted much more heavily, since most people will refer to a young human male as a "boy", versus any other species. Then, the pragmatic listener evaluates the probability of each feature, and combination of features (ie. feature set) in the, given the category (P(f|c)). To do this, the first step is to determine the feature set from the categories elicited in the utterance. In our example, the feature set for a pig is "F = Dirty, Likes to eat, Lives on a farm". Second, this function determines the probability of each feature in the context of each category. This will allow a pragmatic listener to determine the probability of each feature set in terms of the category. From our example in Figure 1, this is the same as P(dirty, likes to eat, lives on a farm | pig) versus P(dirty, likes to eat, lives on a farm | human). In this case, the feature set in terms of the pig category will be weighted much higher because they were elicited to describe a pig.

This model assumes that TOM is needed for all individuals, as the theoretical speaker needs to understand the mind of the imagined literal listener to produce an informative utterance, and the pragmatic listener must understand the possible epistemic states of the speaker -- like whether or not the speaker thinks the boy is dirty -- in order to properly interpret an utterance.

The pragmatic listener function uses a few steps to predict the probability of the speaker's goal (P(g)). First is to determine all of the possible speaker goals (g). To do this, one must, take the prior probabilities of each feature set, given a category--P(f|c)--and multiply them by the probability of one goal--P(g)--and the listener's mental model of the speaker--S_1(u|g,f). The resulting product leaves us with the probability that the speaker (S_1) intended to use a specific feature set combination (P(f|c)) to achieve their goal (P(g)). This is the equivalent of taking the feature set combination P(dirty, likes to eat | human) multiplied by the goal P(dirty) while considering the mental representation of the speaker (S_1). Then, the model completes this multiplication for each possible goal, and takes the sum of all the resulting products. this multiplication across all of our goals. This final step is necessary because the pragmatic listener does not actually know what the speaker's goal is, so they must consider all the possible goals when interpreting the metaphor.

When considering how the probabilities of features set combinations obtain their probability of being the intended interpretation of an utterance, we have to consider them in terms of the speaker's possible goal. We followed Kao et. al. (2014)'s prior probability on goals (P(F) = 0.6, 0.2, 0.2) to account for conversational context, because this heavily skews the probabilities in favor of the interpretations to ones where the first feature (ie. dirty) is present. Therefore, in Figure 1 the probability that the speaker's goal is only to say whether the boy is dirty P(dirty) is 0.6. However, if the goal was to say whether the boy was dirty and hungry P(dirty, likes to eat), the probability is 0.12 because these are independent goals, so the probability of goal "dirty" (0.6) * goal "likes to eat" (0.2) = 0.12.

By this logic, it would be less likely that the speaker's goal is to convey the boy is dirty, yet elicit the feature set "F = likes to eat, lives on a farm". The pragmatic listener will favor the feature sets that focus on the relevant dimension on the goal (ie. focus on any feature encompassed on the scale from clean to dirty).

In Figure 2, we can see that the pragmatic listener envisions the goal of the speaker when determining the intended meaning of the non-literal sentence. The speaker function (S_1(u|g,f)) describes how a speaker chooses an informative utterance.

$$S_1(u|g,\vec{f}) \propto e^{\lambda U(u|g,\vec{f})}$$

The speaker chooses an informative utterance by determining which utterances provide the most utility (more on this to follow). In other words, the speaker must decide which possible utterances will best communicate the intended feature(s) to a literal listener ($L_0(c,f|u)$). The speaker function does this through a softmax decision rule. Taking the softmax of a set of values describes how valuable one choice is over another, and has been used to correlate with human decision making (Luce, 1959). Softmax is operationalized by taking the product of the utility function ($U(u|g,f)$) and a "contrast parameter" (λ), then exponentiates them. This begins the process of calculating the utility (u) of each feature set (f) in terms of completing a goal (g), taking into account any bias the listener may have (λ) to smoothen or sharpen the true probability contrasts. The next step in softmax is to take the sum of the exponents, and divide each exponent by the total sum--i.e. Divide the exponentiated surprisal for the current utterance and divide it by the sum of all the exponentiated utterances possible. This second step normalizes the probabilities between different interpretations, which makes them far easier to compare. This way, if a given utterance (pig) has a higher utility than other available utterances (whale, shark, etc.) then it will have a higher probability of being selected.

The resulting numbers are probabilities of the utility of each utterance, given how useful their features are to achieve the speaker's goal. In Figure 2, we can see that the speaker utilizes information regarding their goal and feature probabilities in order to produce an informative utterance. This step is important because it allows a NT pragmatic listener to make a logical selection of feature combinations based on the context of the situation.

The contrast parameter is a variable that can create bias by either heightening or reducing the utilities of feature combinations. In Kao et. al. (2014) the contrast parameter is set to 3, which is indicative of no skewing of the prior probabilities. Previous RSA models have kept the contrast parameter equal to 3 to model NT children and adults (Savinelli, Scontras. & Pearl. 2017, 2018, Scontras & Pearl 2020 manuscript). However, changing the value of the contrast parameter would alter how feature combinations are weighted. If the contrast parameter is greater than 3, then the differences between two feature combinations will become drastically greater. On the other hand, if the contrast parameter is less than 3, then the differences between two feature combinations will become significantly less different.

The utility function (U(u|g,*f*)) helps the speaker determine how useful one utterance is, relative to other possible utterances.

$$U(u|g,\vec{f}) = \log \sum_{c,\vec{f}} \delta_{g(\vec{f})=g(\vec{f}')} L_0(c,\vec{f}|u)$$

First, the utility function must examine one utterance from a set of possible utterances and determine if the features from the other categories elicit the relevant feature (g(f) = g(f')). In terms of information processing theory, by taking the log of the probability of any one world, represents its surprisal (Shannon 1948). “Precisely this measure of difficulty was in fact proposed by Hale (2001). Surprisal is minimized (goes to zero) when a word must appear in a given context ... and approaches infinity as a word becomes less and less likely.” However, in context of the Kao et al. (2014) model, minimizing surprisal indicates that an utterance should always be used to indicate a particular feature over other possible utterances. If the category does elicit the desired features, then it is given a higher probability of achieving the speaker’s goal. Then, the sum of these resulting numbers is taken to capture all of the probabilities.

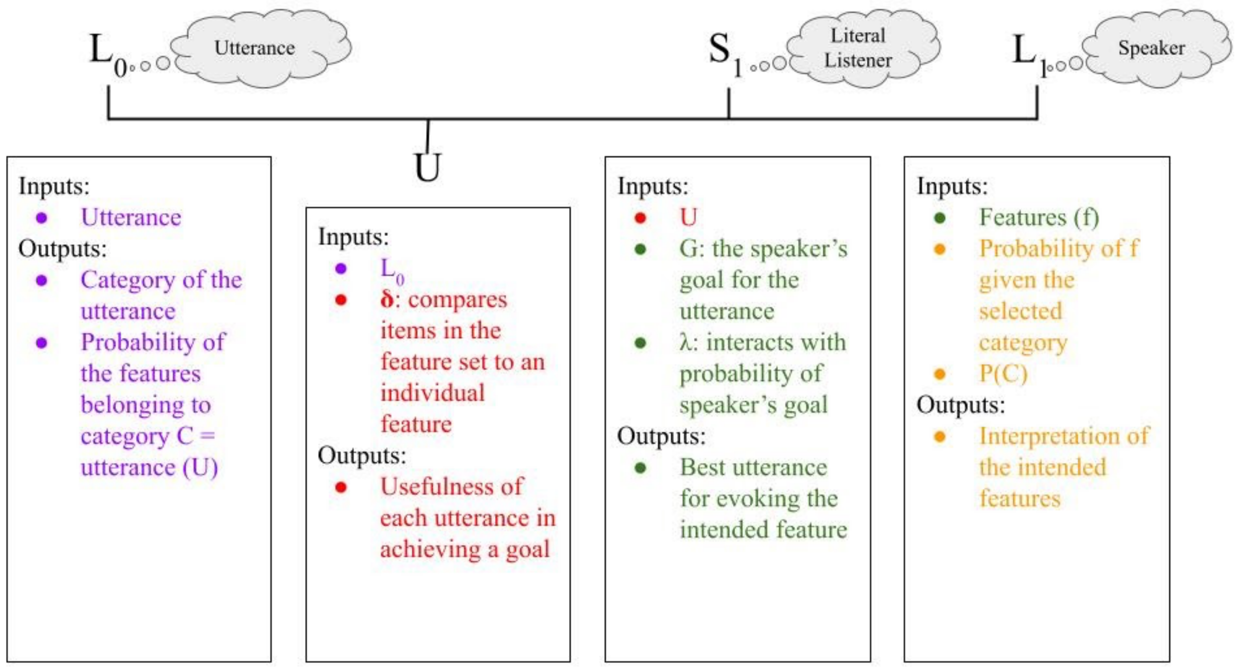
The speaker chooses their utterance by envisioning a literal listener (L_0(c,f|u)), one who will interpret an utterance without pragmatic reasoning , even though the speaker is speaking to a pragmatic listener. By envisioning a literal listener, the speaker chooses the utterance that has the highest utility in order to achieve their goal, while also ending the cycle of pragmatic speakers envisioning the minds of other speakers. Therefore, choosing an utterance with the highest utility will convey the speaker's goal, even in the absence of pragmatic reasoning.

$$L_0\left(c,\vec{f}\middle|u\right) = \begin{cases} P(\vec{f}|c) & if\ c = u \\ 0 & otherwise \end{cases}$$

The comprehension abilities of the literal listener helps the speaker determine the category (c) and feature set (f) of the utterance. The literal listener function implements this process by determining the prior probability that any feature set belongs to the category from the spoken sentence (P(f|c)).

For example, the speaker's goal in Figure 1 is to focus the interpretations of the metaphor to ones including the relevant feature "dirty" to describe the boy, so the speaker will focus on categories that achieve their goal. This means, if the speaker has to choose between a whale (F = majestic, graceful, large), a cat (F = independent, lazy, soft) and a pig (F = dirty, likes to eat, lives on a farm) in order to achieve their goal, the speaker will use a pig because the it has a high probability of a feature relevant to their goal. Other animals would not have yielded these features, therefore the literal listener will be able to understand the sentence even if they do not understand the precise speaker goal.

Figure 2.



Note: Above are the inputs and outputs to all of the equations in Kao et. al.'s RSA model for metaphor comprehension. The literal listener (L_0) will determine the probability of different features given the category from the utterance. These probabilities will determine how useful (U) each feature is in terms of the speaker's goal. The speaker (S_1) takes the utility of each feature into account because they want to choose the most informative utterance for a literal listener. The pragmatic listener (L_1) will select the appropriate interpretation by identifying the relevant features of the intended category, given the speaker's goal.

Adaptations

Our hypothesis was we can alter the probability of the category (P(c), thinking about whether the boy is a pig or human based on knowledge of what people usually say) and contrast parameter (λ, used by the speaker to assess how useful an utterance is based on how the true probability differences are skewed) in order to skew the data in such a manner that the probability of the intended figurative interpretation of a metaphor will become less than half as probable than with the NT settings. In other words, the NT probability of a single intended metaphorical interpretation of an utterance should be twice as high as it would be for an individual with ASD. Kao et. al. (2014) assigned the speaker goal (P(g)) to heavily weigh the first feature in a feature set to achieve the speaker's goal (ie. "dirty" for a pig), to account for conversational context, see Table 2. We defined the intended interpretation of a metaphor, which is the interpretation that only includes the first feature and not the others, based on Kao's prior on goals. For example, in our "The boy is a pig" metaphor we will only focus on the interpretation "The boy is dirty" and not "The boy is dirty and likes to eat" or "The boy is dirty, likes to eat, and lives on a farm". In our analysis, we attempted to decrease the likelihood of this interpretation by 220% to qualitatively replicate ASD behavior.

Methods

Materials

The Kao et. al (2014) model was implemented in Python 3.8 using a mixture of the packages "numpy" and "pandas". We collected a list of animals that can be used in a metaphor, as well as the prior distributions of their feature set combinations[1]. We used the list of utterances, human judgement data for priors over feature set combinations per utterance[1], and the prior on goals collected by the original researchers (Kao 2014).

Task

We replicated the metaphor task using the animals listed in Kao et al. (2014) that overlap with our studies of interest, Happé (1993) and Pastor-Cerezuela (2020). After altering our two identified variables, we hoped the metaphorical interpretation of an utterance will be between 2.23-2.28 times less probable than the NT settings. See Table 1 below for an outline of the data structure used throughout the experiment.

1. The data and model code used will be available at https://github.com/mackephart/asdmetaphorcomp

Table 1.

Data Structure

Paper	Animal	Correct Feature	P(g)
Happé & Pastor-Cerezuela	Fox	F1 = sly F2 = smart F3 = pretty	F1 = 0.6 F2 = 0.2 F3 = 0.2
Kao	Monkey	F1 = funny F2 = smart F3 = playful	F1 = 0.6 F2 = 0.2 F3 = 0.2
Pastor-Cerezuela	Pig	F1 = dirty F2 = fat F3 = smelly	F1 = 0.6 F2 = 0.2 F3 = 0.2
Kao	Shark	F1 = scary F2 = dangerous F3 = mean	F1 = 0.6 F2 = 0.2 F3 = 0.2
Kao	Whale	F1 = large F2 = graceful F3 = majestic	F1 = 0.6 F2 = 0.2 F3 = 0.2

Note: An outline of our five selected utterances. Our model will interpret each utterance in the form “X is a Y”. The intended interpretation is the elicitation of the first feature (F1). We used Kao et. al.’s prior probability on goals (P(g)) when weighing each feature.

Priors

Prior distributions of data tell us how descriptive each feature set combination is in terms of describing the animal itself. This is what we see in Figure 1 during the integration step: the weights of each feature alone and in combination when describing the target. We used the previously collected prior distributions of human judgements to feed into the model to interpret each utterance.

In order to adapt the model to reflect ASD behavior, we wanted the model to decrease the probability of the metaphorical interpretation, where eliciting the first feature completes the speaker's goal, by approximately 220%. We attempted to replicate this behavior using three manipulations of the model, focusing on the contrast parameter (λ) and the prior probability of the category (P(c)).

The neurotypical inputs for P(c) are: P(animal) = 0.01 and P(human) = 0.99, and the NT value of the contrast parameter is 3. We chose P(c) as our prior of focus because previous studies have shown individuals with ASD have tendencies to prefer the literal interpretations of a metaphor (Kalandadze 2018), which could indicate confusion about the literal category (animal) or the intended category (human). In our experiment, we let the model generate values of P(c) alone that returned probable ASD interpretation behavior, as well as tested values of P(c) at 0.5 for animal and human when combined with the contrast parameter. We were interested in the contrast parameter because it can induce bias towards different interpretations of a metaphor. We asked the model to calculate values of the contrast parameter that returned probable ASD interpretation behavior alone and when in combination with P(c). See Table 2 below for an outline of the manipulations we ran with the model.

Manipulation	NT - Test 0	ASD - Test 1	ASD - Test 2	ASD - Test 3
P(c)	P(A) = 0.01 P(H) = 0.99	0 < P(A) <1 0 < P(H) <1 Increment: 0.01	P(A) = 0.01 P(H) = 0.99	P(A) = 0.5 P(H) = 0.5
Contrast	3	3	0 < λ < 500 Increment: 0.1	0 < λ < 500 Increment: 0.1

Note: An outline of our model manipulations that were completed for every animal in Table 1.

Analysis

In order to compare the NT behavior to ASD behavior, we ran three analyses. The first step for all three manipulations was to collect the NT probability of the intended metaphorical interpretation for each animal, see Test 0 in Table 2. Then, we altered the model in three different ways in an attempt to qualitatively replicate ASD. Finally, we took the quotient of the NT output of the intended metaphorical interpretation and the ASD output of the intended metaphorical interpretation.

The first adaptation checked if changing the probability of the category (P(c)) alone was sufficient to return probable ASD behavior (see Test 1 in Table 2). To test this, we ran an ad-hoc grid search algorithm to test values of the probability of the category (P(c)) for both animal and human, from 0 to 1 at an interval of 0.01. In this manipulation, we left the contrast parameter at the neurotypical constant value of 3. Then, we divided the NT output by the ASD output per each stimulus. We collected all values of P(animal) and P(human) per stimuli, and isolated the values that yielded our target behavior.

The second adaptation checked if changing the probability of the contrast parameter alone was sufficient to return probable ASD behavior (see Test 2 in Table 2). Similar to the first manipulation, we ran an ad-hoc grid search algorithm to test values of the contrast parameter from 0 to 500, at every 0.1 interval. In this manipulation, we used the neurotypical inputs for the probability of the category, where P(animal) = 0.01 and P(human) = 0.99. Then, we divided the NT output by the ASD output per each stimulus. We collected all values of the contrast parameter per stimuli, and isolated the values that yielded our target behavior.

The third adaptation involved changing the weight of the probability of the category (P(c)) to 0.5 for both animal and human. Simultaneously, we ran an ad-hoc grid search algorithm to test values of the contrast parameter from 0 to 500, at every 0.1 interval (see Test 3 in Table 2). Then, we divided the NT output by the ASD output per each stimulus. We collected all values of the contrast parameter per stimuli, and isolated the values that yielded our target behavior.

Results

For each manipulation of the model, we generated results for five animal stimuli. The intended metaphorical interpretation for each stimulus are as follows: the interpretation of pig was “dirty”, the interpretation of fox was “sly”, the interpretation of shark was “scary”, the interpretation of monkey was “funny”, and the interpretation of whale was “large”.

In our first adaptation, where the contrast parameter was set to 3, we found all metaphors required a higher probability for the category animal versus human (P(animal) > P(human)) in order to yield probable ASD behavior. First, for the utterance “pig”, the values of the probability of the category that fit the criteria were: P(a) = 0.55 and P(h) = 0.45. For the utterance “fox”, the values of the probability of the category that fit the criteria were: P(a) = 0.56 and P(h) = 0.44, as well as P(a) = 0.57 and P(h) = 0.43. The utterance “shark”, the values of the probability of the category that fit the criteria were: P(a) = 0.56 and P(h) = 0.44. In the utterance "monkey", the values of the probability of the category that fit the criteria were: P(a) = 0.54 and P(h) = 0.46. Finally, for the utterance "whale", the values of the probability of the category that fit the criteria were: P(a) = 0.57 and P(h) = 0.43. See Table 3 below.

Table 3.
Values of P(c) that replicates ASD behavior per each utterance

Utterance	Speaker's Goal	P(c)	Contrast Parameter
Pig	Dirty	P(A) = 0.55 P(H) = 0.45	3
Fox	Sly	P(A) = 0.56 P(H) = 0.44 and P(A) = 0.57 P(H) = 0.43	3
Shark	Scary	P(A) = 0.54 P(H) = 0.46	3
Monkey	Funny	P(A) = 0.57 P(H) = 0.43	3
Whale	Large	P(A) = 0.56 P(H) = 0.44	3

In our second adaptation, where the probability of the category was kept at the neurotypical inputs (P(animal) = 0.01 and P(human) = 0.99), we found all metaphors required a contrast parameter value higher than the neurotypical value of 3 in order to yield probable ASD behavior. First, for the utterance "pig", the values of the contrast parameter that fit the criteria were: $28.8 \leq \lambda \leq 29.6$. For the utterance "fox", the values of the contrast parameter that fit the criteria were: $76.2 \leq \lambda \leq 78.2$. The utterance "shark", the values of the contrast parameter that fit the criteria were: $108.7 \leq \lambda \leq 110.5$. In the utterance "monkey", the values of the contrast parameter that fit the criteria were: $64 \leq \lambda \leq 65.6$. Finally, for the utterance "whale", the values of the contrast parameter that fit the criteria were: $378.8 \leq \lambda \leq 381$. See Table 4 below.

Table 4.
Values of the contrast parameter that replicates ASD behavior per each utterance

Utterance	Speaker's Goal	P(c)	Contrast Parameter
Pig	Dirty	P(A) = 0.01 P(H) = 0.99	$28.8 \leq \lambda \leq 29.6$
Fox	Sly	P(A) = 0.01 P(H) = 0.99	$76.2 \leq \lambda \leq 78.2$
Shark	Scary	P(A) = 0.01 P(H) = 0.99	$108.7 \leq \lambda \leq 110.5$
Monkey	Funny	P(A) = 0.01 P(H) = 0.99	$63.9 \leq \lambda \leq 65.6$
Whale	Large	P(A) = 0.01 P(H) = 0.99	$378.8 \leq \lambda \leq 381$

In our third adaptation, where the probability of the category was set at 0.5 for both animal and human, we found all metaphors required a contrast parameter value higher than the neurotypical value of 3 in order to yield probable ASD behavior. First, for the utterance "pig", the values of the contrast parameter that fit the criteria were: $6 \leq \lambda \leq 6.8$. For the utterance "fox", the values of the contrast parameter that fit the criteria were: $28.7 \leq \lambda \leq 33.4$. The utterance "shark", the values of the contrast parameter that fit the criteria were: $6.7 \leq \lambda \leq 8$. In the utterance "monkey", the values of the contrast parameter that fit the criteria were: $23.7 \leq \lambda \leq 26.7$. Finally, for the utterance "whale", the values of the contrast parameter that fit the criteria were: $346.2 \leq \lambda \leq 348.5$. See Table 5 below.

Table 5.
Values P(c) and contrast value that replicates ASD behavior per each utterance

Utterance	Speaker's Goal	P(c)	Contrast Parameter
Pig	Dirty	P(A) = 0.50 P(H) = 0.50	$6 \leq \lambda \leq 6.8$
Fox	Sly	P(A) = 0.50 P(H) = 0.50	$28.7 \leq \lambda \leq 33.4$
Shark	Scary	P(A) = 0.50 P(H) = 0.50	$6.7 \leq \lambda \leq 8$
Monkey	Funny	P(A) = 0.50 P(H) = 0.50	$23.7 \leq \lambda \leq 26.7$
Whale	Large	P(A) = 0.50 P(H) = 0.50	$346.2 \leq \lambda \leq 348.5$

Discussion

Our results support our hypothesis that altering the contrast parameter and probability of the category, alone and in combination, can replicate ASD behavior.

One qualitative pattern we noticed across stimuli is that when the prior probability of the category was changed independently of the contrast parameter, the probability of the literal (animal) category was higher than the intended category (human). This is consistent with the literature we discussed previously, where individuals with ASD tend to prefer literal interpretations of figurative utterances (Kalandadze (2018), Happé (1993), McKay & Shaw (2004), Melogno (2017)). Kao includes an operational definition for the literal interpretation of an utterance, so future studies could see if P(c) reliably approximates how likely the literal interpretation of a metaphor will be chosen, using a different study to replicate.

Another qualitative pattern we noticed in the data is that the contrast parameter required a value higher than 3 for all of the stimuli, whether the probability of the category was also altered or not. This indicates that individuals with ASD engage in an abnormally high amount of skewing by drastically increasing the differences in possible feature sets. However, the large differences between the average contrast values across stimuli indicates there is not a single contrast value that can replicate ASD metaphor comprehension behavior. This may have happened for a number of reasons.

One possible explanation is that the contrast parameter simply does not have enough power to make large differences in the calculations of the model. As we noticed with P(c) alone, there was a very small difference between the prior values, indicating that this prior has a strong influence on the resulting probabilities of the interpretations. It is possible that the contrast parameter would have a more reliable effect on capturing smaller interpretation differences, such as interpreting “The boy is a pig” to mean “The boy is dirty” or “The boy likes to eat”.

Another possibility as to why the contrast parameter did not replicate ASD comprehension behavior may be due to the differences in the content of the metaphors. Our studies of interest tested far more metaphors than just animals, see Appendix 1 and 2 below. However, from Kao et al. 2014 we only had priors for neurotypical individuals’ comprehension of metaphors containing animals. Future researchers could follow the original Kao et. al. (2014) methods to finding neurotypical priors for objects and other animals (such as a swan and a volcano) to see if this model produces more reliable results.

References

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders. Autism Spectrum Disorder.* (Fifth Edition). American Psychiatric Association. https://doi.org/10.1176/appi.books.9780890425596

Chouinard, B., Volden, J., Cribben, I., & Cummine, J. (2017). Neurological evaluation of the selection stage of metaphor comprehension in individuals with and without autism spectrum disorder. *Neuroscience*, *361*, 19–33. https://doi.org/10.1016/j.neuroscience.2017.08.001

Gold, R., Faust, M., & Goldstein, A. (2010). Semantic integration during metaphor comprehension in Asperger syndrome. *Brain and Language*, *113*(3), 124–134. https://doi.org/10.1016/j.bandl.2010.03.002

Goodman, N. D., & Frank, M. C. (2016). Pragmatic Language Interpretation as Probabilistic Inference. *Trends in Cognitive Sciences*, *20*(11), 818–829. https://doi.org/10.1016/j.tics.2016.08.005

Happé, F. G. E. (1993a). Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition*, *48*(2), 101–119. https://doi.org/10.1016/0010-0277(93)90026-R

Hochstein, L., Bale, A., & Barner, D. (2018). Scalar Implicature in Absence of Epistemic Reasoning? The Case of Autism Spectrum Disorder. *Language Learning and Development*, *14*(3), 224–240. https://doi.org/10.1080/15475441.2017.1343670

Kana, R. K., & Wadsworth, H. M. (2012). “The archeologist’s career ended in ruins”: Hemispheric differences in pun comprehension in autism. *NeuroImage*, *62*(1), 77–86. https://doi.org/10.1016/j.neuroimage.2012.04.034

Kalandadze, T., Norbury, C., Nærland, T., & Næss, K.-A. B. (2018). Figurative language comprehension in individuals with autism spectrum disorder: A meta-analytic review. *Autism*, *22*(2), 99–117. https://doi.org/10.1177/1362361316668652

Kao, J., Bergen, L., & Goodman, N. (2014). Formalizing the Pragmatics of Metaphor Understanding. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *36*, 6.

Lakoff, G., & Johnson, M. (2003). *Metaphors we live by*. University of Chicago Press.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

Luce, R. D. (1959). *Individual Choice Behavior: A theoretical analysis*. John Wiley & Sons.

MacKay, G., & Shaw, A. (2004). A comparative study of figurative language in children with autistic spectrum disorders. *Child Language Teaching and Therapy*, *20*(1), 13–32. https://doi.org/10.1191/0265659004ct261oa

Mashal, N., & Kasirer, A. (2011). Thinking maps enhance metaphoric competence in children with autism and learning disabilities. *Research in Developmental Disabilities*, *32*(6), 2045–2054. https://doi.org/10.1016/j.ridd.2011.08.012

Melogno, S., Pinto, M., & Di Filippo, G. (2017). Sensory and Physico-Psychological Metaphor Comprehension in Children with ASD: A Preliminary Study on the Outcomes of a Treatment. *Brain Sciences*, *7*(12), 85. https://doi.org/10.3390/brainsci7070085

Pastor-Cerezuela, G., Fernández-Andrés, M.-I., Tordera-Yllescas, J. C., & González-Sala, F. (2020). Metaphor comprehension in children with and without autism spectrum disorder. *Research in Autism Spectrum Disorders*, *76*, 101588. https://doi.org/10.1016/j.rasd.2020.101588

Savinelli, K. J., Scontras, G., & Pearl, L. (2017). Modeling scope ambiguity resolution as pragmatic inference: Formalizing differences in child and adult behavior. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 6.

Savinelli, K. J., Scontras, G., & Pearl, L. (2018). Exactly two things to learn from modeling scope ambiguity resolution: Developmental continuity and numeral semantics. *Proceedings of the 8th Workshop on Cognitive Modeling and Computational Linguistics (CMCL 2018)*, 67–75. https://doi.org/10.18653/v1/W18-0108

Scontras, G., & Pearl, L. (2020). *When pragmatics matters more for truth-value judgments: An investigation of quantifier scope ambiguity* [Manuscript]. https://www.socsci.uci.edu/~lpearl/papers/ScontrasPearl2020Manu_ChildScope.pdf

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Sobh Chahboun, Valentin Vulchanov, David Saldaña, Hendrik Eshuis, & Mila Vulchanova. (2016). Predictors of metaphorical understanding in high functioning autism. *Lingue e Linguaggio*, *1*, 29–58. https://doi.org/10.1418/83653

Wang, A. T., Lee, S. S., Sigman, M., & Dapretto, M. (2006). Neural basis of irony comprehension in children with autism: The role of prosody and context. *Brain*, *129*(4), 932–943. https://doi.org/10.1093/brain/awl032

Williams, D. L., Cherkassky, V. L., Mason, R. A., Keller, T. A., Minshew, N. J., & Just, M. A. (2013). Brain Function Differences in Language Processing in Children and Adults with Autism: Differences in language processing in autism. *Autism Research*, *6*(4), 288–302. https://doi.org/10.1002/aur.1291

Appendix 1.

Synonym

1. The oak tree was so knarled and crooked. It really was . . .
2. Jane was so pale and quiet. She really was . . .
3. Sarah was so beautiful. She really was . . .
4. Steve was always rushing around, leaving everything in a mess. He really was . . .
5. Everyone found it hard to make friends with Penny. She really was . . .

Choose one item from the following list to complete each sentence:

lovely   unwell   energetic   ancient   generous   unapproachable

Simile

1. The dog was so wet. It was like . . .
2. Carol glared at Nicola. She was so cross. Her eyes were like . . .
3. The night sky was so clear. The stars were like . . .
4. Simon just couldn't make Lucy understand. She was like . . .
5. Caroline was so embarrassed. Her face was like . . .

Choose one item from the following list to complete each sentence:

A brick wall   dresses   daggers   a beetroot   a walking puddle   diamonds

Metaphor

1. The dancer was so graceful. She really was . . .
2. Father was very very cross. He really was . . .
3. Michael was so cold. His nose really was . . .
4. Ian was very clever and tricky. He really was . . .
5. Ann always felt safe with Tom. He really was . . .

Choose one item from the following list to complete each sentence:

an icicle   a fox   a safe harbour   a hat   a swan   a volcano

Note: A full list of the stimuli used by Happé et. al. (1993).

Appendix 2.

What does it mean? Choose the correct response in each case.

1. <u>This girl is a fox.</u>
    a. This girl has a fox in her house.
    b. This girl is very sly and alert.
    c. This girl likes to play chess.
2. This boy is a pig.
    a. This boy has pigs in his house.
    b. This boy likes to eat ham.
    c. This boy gets very dirty.
3. <u>My brother is brilliant.</u>
    a. My brother is very smart.
    b. My brother has a shiny face.
    c. My brother talks a lot.
4. <u>I'm up to my eyeballs these days.</u>
    a. I'm very busy.
    b. My eyes hurt.
    c. My eyes are blue.
5. This school is a prison.
    a. I have a hard time in this school.
    b. There are books in this school.
    c. I have a good time in this school.
6. The news lifted me up.
    a. The news made me happy.
    b. The news made me raise my arms.
    c. The news made me sad.
7. Pepe has a high position in the company.
    a. Pepe lives on a high floor.
    b. Pepe's job is very important.
    c. Pepe is very fat.
8. <u>Juan is very dense today.</u>
    a. Juan isn't hungry.
    b. Juan has irritated skin.
    c. Juan doesn't understand things well.
9. I love Maria's velvety voice.
    a. Maria is wearing a scarf on her neck.
    b. Maria is hoarse from talking so much.
    c. Maria has a pleasant voice.
10. My friend has a screw loose.
    a. When they made my friend, they forgot to put a screw in.
    b. My friend does a lot of silly and stupid things.
    c. My friend's computer is missing a screw.
11. <u>The classroom was a zoo.</u>
    a. The children were working quietly.
    b. The teacher was a monkey.
    c. The children were running and playing.
12. <u>The child is down in the dumps.</u>
    a. The child fell down.
    b. The child is sad.
    c. The child is a little clumsy.
13. Luis doesn't grasp the idea.
    a. Luis can't hold it in his hand.
    b. Luis doesn't understand the explanation.
    c. Luis dropped something.
14. <u>Quique was feeling low.</u>
    a. Quique was sitting on the floor.
    b. Quique was touching the ground.
    c. Quique was sad.
15. <u>During the argument, Pedro really attacked Maria.</u>
    a. Pedro was very critical of Maria.
    b. Pedro hit Maria.
    c. Pedro didn't want to argue with Maria.
16. Juan wasted his time.
    a. Juan lost a valuable watch.
    b. Juan used his time for something useless.
    c. Juan couldn't find something valuable he had lost.
17. <u>The conversation with Rachel flowed.</u>
    a. Rachel drools a lot when she talks.
    b. Rachel is a good swimmer.

   c. Rachel expresses herself well when she talks.
18. Time is golden.
   a. Time goes by very slowly.
   b. Time is important.
   c. The weather is nice today.
19. Antonio lost the thread of the conversation.
   a. Antonio doesn’t know what the conversation was about.
   b. Antonio could not sew a type of clothing.
   c. Antonio spoke in a very soft voice.
20. Ana had high hopes.
   a. Ana had great hopes and expectations.
   b. Ana shot an arrow high in the air.
   c. Ana doesn’t know how to hold a shotgun.

Note: A full list of the stimuli used by Pastor-Cerezuela (2020). If the question is underlined, that indicates the sentence is a conventional metaphor, and not underlined sentences are novel metaphors.