Human Errors in Interpreting Visual Metaphor

Savvas Petridis
Columbia University
New York, USA
savvas@cs.columbia.edu

Lydia B. Chilton
Columbia University
New York, USA
chilton@cs.columbia.edu

ABSTRACT

Visual metaphors are a creative technique used in print media to convey a message through images. This message is not said directly, but implied through symbols and how those symbols are juxtaposed in the image. The messages we see affect our thoughts and lives, and it is an open research challenge to get machines to automatically understand the implied messages in images. However, it is unclear how people process these images or to what degree they understand the meaning. We test several theories about how people interpret visual metaphors and find people can interpret the visual metaphor correctly without explanatory text with 41.3% accuracy. We provide evidence for four distinct types of errors people make in their interpretation, which speaks to the cognitive processes people use to infer the meaning. We also show that people's ability to interpret a visual message is not simply a function of image content but also of message familiarity. This implies that efforts to automatically understand visual images should take into account message familiarity.

Author Keywords

Visual Design; Visual Thinking; Crowdsourcing

CCS Concepts

•**Human-centered computing** → **Visualization;** *Empirical studies in visualization;*

INTRODUCTION

Visual metaphors are a creative technique used in print media to convey a message. They lead viewers to form an association between two objects like face cream and night or fast food and dangerous. Visual metaphors are intriguing because they do not convey messages directly, they convey meaning by implying it through the symbolism and juxtaposition of the symbols. However, when communicating a message implicitly there is a risk that a) people are not fully aware of the messages or b) that viewers will misinterpret them.

Visual messaging is everywhere. Message creators want to know if their images can be understood. Designers of public service announcements about healthy eating or the importance
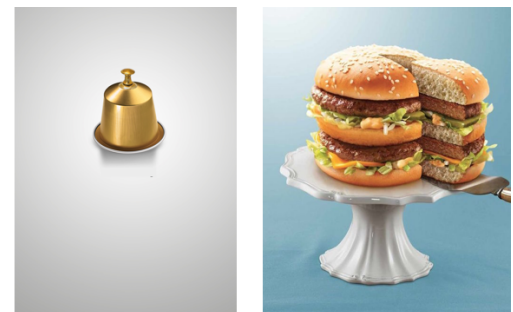
Figure 1. The images participants interpreted with the highest and lowest accuracy

of recycling want to know if their messages are comprehensible. Also, people are worried about what implicit messages are being directed at them, and how this may influence their thoughts and behavior [24]. Many people are worried about being tricked into buying something based on the implication that it will help them, rather than being swayed by actual facts. To better the creation and understanding of visual communication, it is necessary to discern how these images are interpreted and where individuals err in their interpretations.

Efforts to get machines to automatically understand these messages have proven to be difficult; it is unclear whether automatic understanding is even possible. Perhaps there is social knowledge that is hard for machines to learn, preventing them from understanding implicit messages. Moreover, linguist theories of visual metaphors suggest it is difficult even for people to interpret their meanings with the image alone. To improve automatic understanding, it is necessary to understand how people err in interpreting visual metaphors.

We seek to provide understanding as to how meaning is conveyed in visual metaphors. Although visual metaphors often

Figure 2. Four visual metaphors from three domains: product advertisements, public service announcements, and journalism.

appear with supporting text, we test linguist theories of visual metaphor interpretation to see how well people can interpret the image alone. Additionally, we look at where people fail. Our findings have implications for helping designers create more interpretable visual metaphors, setting expectations on where machines might reasonably fail, and helping machines predict whether people will fail or succeed.

We also examine the degree knowledge of the world people use in interpreting visual metaphors. If outside knowledge is key, then machine interpretations will have to be augmented to incorporate this world knowledge, rather than solely rely on annotated training examples.

This paper makes the following contributions:

- A study of 48 visual metaphors showing that people can correctly interpret visual metaphors, and they do so 41.3% of the time. Figure 1 shows the two metaphors with the most correct interpretations and the two with the least.

- Empirical evidence of four types of errors people make when interpreting visual metaphors that points to insights designers should keep in mind and how machines should interpret visual metaphors.

- An analysis showing that contrary to theory, visual metaphors in advertisements and non-advertisements (PSA's and book covers) have the same rate of error in direction of property transfer.

- An analysis showing that familiarity with the message is correlated with people's ability to correctly interpret a visual metaphor, indicating that world knowledge is helpful for interpretation and automated approaches might need to incorporate this world knowledge.

We discuss implications of these findings for how to help machines automatically interpret the messages of visual metaphors, and to help machines and people create visual metaphors to convey a meaning.

BACKGROUND ON VISUAL METAPHORS
Visual metaphors are studied in linguistics, specifically in the area of pragmatics, which has to do with how context is used to convey meaning. They visually combine objects in an image in order to compare one to the other. For example, in Figure 2, the first, leftmost image is an advertisement to "Use Nivea

cream at night", that blends Nivea cream with the moon. Nivea cream is like the moon in that it is associated with night. The second image is a public service announcement (PSA) that conveys a very different meaning: "Fries are deadly". In this image, bullets have been visually blended with french fries. Fries are like bullets in that they are deadly. The third image is the cover of a book about how ideas spread, depicting a blend of a dandelion and light bulb. In this example, the light bulb represents ideas and is like a dandelion in that it spreads. The fourth image is an advertisement conveying: "Absolut Vodka is fashionable". In this case Absolut is like a dress form in that it is fashionable. In each case, an implicit meaning is being conveyed by the juxtaposition of these objects in the image. Visual metaphors are a powerful means for conveying a variety of ideas.

Linguists such as Charles Forceville research visual metaphors and pose theories of how these images convey meaning. Forceville contends that visual metaphors, like verbal metaphors, must explain one object in terms of another [5]. More specifically, he claims that in both verbal and visual metaphors, one object, the 'target', receives a property from the other object, the 'source'. Consider the verbal metaphor: "the classroom is a zoo". In this case, the classroom is the target, receiving a property *wildness* from zoo, the source. Visual metaphors follow these same rules [6]. In the "fries are deadly" image in Figure 2, the french fries are the target, receiving the property *deadly* from the source, bullets. Thus both visual and verbal metaphors convey a relationship between two objects, in which one receives a property from the other.

However, visual metaphors are harder to interpret than verbal metaphors because there are more things the reader must infer. Source and target are clear in verbal metaphors but ambiguous in visual metaphors. Consider once again the metaphor: "the classroom is a zoo". Because classroom comes before zoo in the sentence we know classroom is the target and zoo is the source [5]. There are grammatical cues that indicate source and target. Meanwhile, visual metaphors are more ambiguous. In the french fry and bullet blend from Figure 2, how can we be sure that french fries is receiving a property from bullets? Perhaps the image means bullets are cheap like french fries. And beyond source and target, how can we ascertain the property being transferred? Are french fries deadly like bullets, or are they metallic like bullets? One can err in multiple ways

when determining the meaning of a visual metaphor, since the source, target, and property being transferred must all be inferred.
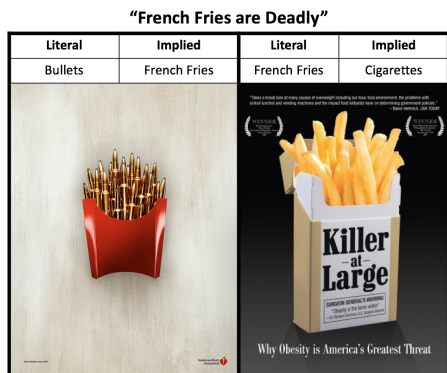

Figure 3. Two public service announcements conveying the same meaning: "French fries are deadly". In the left image, the french fries are implied, and in the right image, the french fries are literal.

Theory of Visual Metaphor Interpretation

Forceville has theorized where viewers could potentially falter in their interpretations of visual metaphors [7]. The most fundamental error one could make is not recognizing the two objects compared in the image. The viewer could potentially see only one object and miss the the second entirely, like the moon in the Nivea advertisement in Figure 2. In this case, the viewer's interpretation will entirely miss a metaphorical comparison. One could also misidentify one or both of the objects and interpret a comparison between objects that are not in the image. If the viewer correctly identifies the two objects, he or she must then infer the source and target, which are ambiguous as well.

Forceville contends that there are no visual cues that definitively identify source and target [5]. For example, visual metaphors often contain one term that appears literally and one that is implied. In the "fries are dangerous" PSA, the bullets appear literally and the fries are implied by the container and silhouette of the bullets. So in this case, the implied object (fries) is the target, receiving the property *deadly*, from the literal object (bullets). In Figure 3 we have another "fries are deadly" PSA, where once again, fries are the target receiving the property *deadly*. But the key difference is that now the fries are literally present, and the cigarettes (source) are implied. Crucially, it is the literal term (french fries) receiving a property from the implied term (cigarettes). One cannot rely on visual properties to determine source and target in visual metaphors. Is there any way to infer the meanings of these images?

Forceville believes that one clue for identifying source and target is discerning that the visual metaphor is an advertisement. For example, if we know an image is an advertisement and we see a product in it, like Nivea cream in Figure 2, then we can safely assume Nivea cream is the target and that the advertisement is making a statement about Nivea cream, not the moon. This is an interesting theory to test.

Even if viewers identify the source and the target, they still must determine the property being transferred. Forceville claims that it is only through the image's explanatory text that we can be sure what property is transferred. For example, on the bottom-left of the "French fry is deadly" PSA in Figure 2, there is text that asks: "How deadly is your diet?" From this we can now be sure that the property is *deadly* and that the target is french fry. Similarly, in the Nivea advertisement from the same Figure, the property *night* is written on the cream container. In this paper, we test this claim as well. We assess whether or not people can correctly determine the meaning of visual metaphors without their explanatory text. We also seek to provide empirical evidence of errors people make when interpreting visual metaphors.

RELATED WORK

Visual Metaphor Comprehension

The visual structure of the metaphor affects its comprehensibility. Phillips and McQuarrie (2004) developed a typology of visual metaphors in advertisements, identified three types of visual structures, and ordered them by complexity [17]. The first and simplest structure is 'juxtaposition', where the two objects are both completely present and shown side-by-side. The second structure is 'fusion', where the two objects are both partially present and have been blended into a single object, like the dandelion-light bulb metaphor in Figure 2. The third and most complex structure is 'replacement', where only one object appears and the other is implied, exemplified in the Nivea Night Cream ad in Figure 2. This work provides a useful typology of visual metaphor in advertising and argues that some are more complex than others, implying that certain visual structures of metaphors are more comprehensible than others.

More recent work involves experiments in which visual metaphor comprehension is tested. In the following studies, comprehension is assessed in order to make a comparison, often between visual metaphors and plain advertisements. For example, McQuarrie and Mick (1999) measure comprehension in order to compare metaphor and non-metaphor advertisement understanding [13]. They measured comprehension by asking participants whether they found the advertisement's meaning straightforward or not. This simple measure was used in other studies as well, including one comparing visual metaphor comprehension across European cultures [12] and another where comprehension is measured for metaphor in television advertisements [20]. A different measure of comprehension was used by Van Mulken et. al. (2014), in which they asked participants to choose the correct meaning in a multiple choice question [22]. The most open-ended test of comprehension was conducted by Van Mulken et. al. (2010), in which participants were asked if they perceived a comparison in the image and if so, to label the source object (the target was always a car) [21]. No work has of yet had participants provide free response interpretations of visual metaphors in order to identify the errors in their interpretations. By learning how people fail to interpret their meanings, we can better tools that seek to automatically understand and generate persuasive visual messages.

Visual Metaphors and Persuasion

Prior work has shown that people are more likely to form positive associations with a product when viewing advertisements that contain visual metaphors [14]. This is the case because viewers can map multiple positive associations with the product and the object it is being compared to. Because of their ambiguity, visual metaphors are like puzzles and encourage viewers to spend time with the image to form these associations. Furthermore, this puzzle-like quality of visual metaphors leads to a positive reaction once the viewer unlocks the meaning, leading to more positive views on the brand and willingness to purchase the product [10] [1] [23]. Because visual metaphors have proven to be an effective means of persuading in advertisements, they have been applied and analyzed in various other contexts such as public service announcements for mental illness [11] and environmental protection [15]. Visual metaphors are a much more persuasive means of conveying a message than plain advertisements or text alone.

Automatically Understanding Visual Messages

Given the prominence of visual advertisements in our daily lives, researchers have begun creating intelligent systems to automatically understand advertisement content [9]. Hussain et. al. (2017) collected a data set of 64,832 visual advertisements and paid Mechanical Turk workers to annotate them for sentiment, symbolism, and the action the advertisement wants the viewer to do ("buy a dress") and why ("it will make me pretty"). The ads in their data set use many different techniques to convey messages. Very few of them are visual metaphors. Instead, most of the ads use a "straightforward" strategy where the message can be inferred from the objects or the text in the image. They trained a model to predict the meaning of an advertisement with 48.45% accuracy.

In fact, much of the success of these algorithms comes not from looking at the images, but from reading the accompanying text. The text often contains a fairly explicit statement of what the advertisement wants you to do and why. It is an open challenge to try to interpret the meaning just from the images. One step made in this direction is including object identification for decoding visual symbols in advertisements [25]. We touch upon challenges related to object identification as well as the role of culture in the discussion.

Creating Visual Blends

Even more challenging than understanding visual advertisements is generating them. One aspect of visual metaphors is that they combine two objects together in a visual blend. Three computational systems have been built that combine two figures. One is in the domain of emojis [4]. The second is in the domain of hand-drawn figures [3]. The results of these two systems show they produce blends that people considered interesting and surprising. The third system is VisiBlends [2], which involves computational steps and human microtasks to create blends of images for advertisements, news, and public service announcements. These images convey messages like "football is dangerous" and "washing your hands is smart." The evaluation shows that by using the workflow, novices' ability to create visual blends increased by a factor of 10. This system relies on human ability to brainstorm visual symbols for concepts and computational approaches for searching for symbols that meet the constraints required for blending.

A limitation of these systems is that they do not evaluate how well viewers can interpret the meaning of the images. This is a difficult problem because we thus far do not have a baseline of how well people interpret professional visual metaphors. This paper aims to provide that professional baseline. Additionally, it would help computational systems if computers could automatically assess whether a blend conveyed the intended meaning and whether viewers would be able to understand it. By classifying the errors people make when interpreting visual metaphors, we can start to break down the process of understanding them and identify the common pitfalls of misunderstanding.

EXPERIMENTAL SETUP

To test if people can correctly interpret visual metaphors, we selected 48 visual metaphors, removed their text, and asked participants what they meant. The data set includes a variety of metaphors from advertisements, news articles, and public service announcements in order to diversify the set of meanings. In the following experiment section, we specifically address the following research questions:

1. Can people correctly interpret the meaning of visual metaphors without their explanatory text?

2. When people misinterpret visual metaphors, where are the errors in their interpretations?

3. Are people more likely to correctly identify the direction of property transfer in ads (vs non-ads) because viewers can infer the product is the target?

4. Does the familiarity of the message have an effect on how likely people are to interpret it correctly?

Participants

We recruited 20 participants on Amazon Mechanical Turk to each interpret all 48 metaphors. Participants were restricted to having at least 95% approval on at least 500 tasks. Workers were also restricted to being located in the US. This restriction was made because relevant cultural knowledge is important for interpreting metaphors, and the metaphors in the data set were taken from American advertisements, news articles, and public service announcements. Each worker was paid $10 dollars and took on average 38 minutes and 34 seconds to complete the task, which averages to an $15.56 hourly wage.

Data Preparation

The images used in this study were found by searching for "creative ads" on image databases like Google Images and Pinterest. Images were included if they met the criteria for a visual metaphor. Specifically, each image needed to compare two objects. One object had to be the source and the other the target, receiving a clear property. The authors used each image's explanatory text to identify its two objects, source, target, and property. Experts also condensed each meaning to the target object and property it receives. For example, consider the popsicle and iceberg blend in Table 1. The iceberg
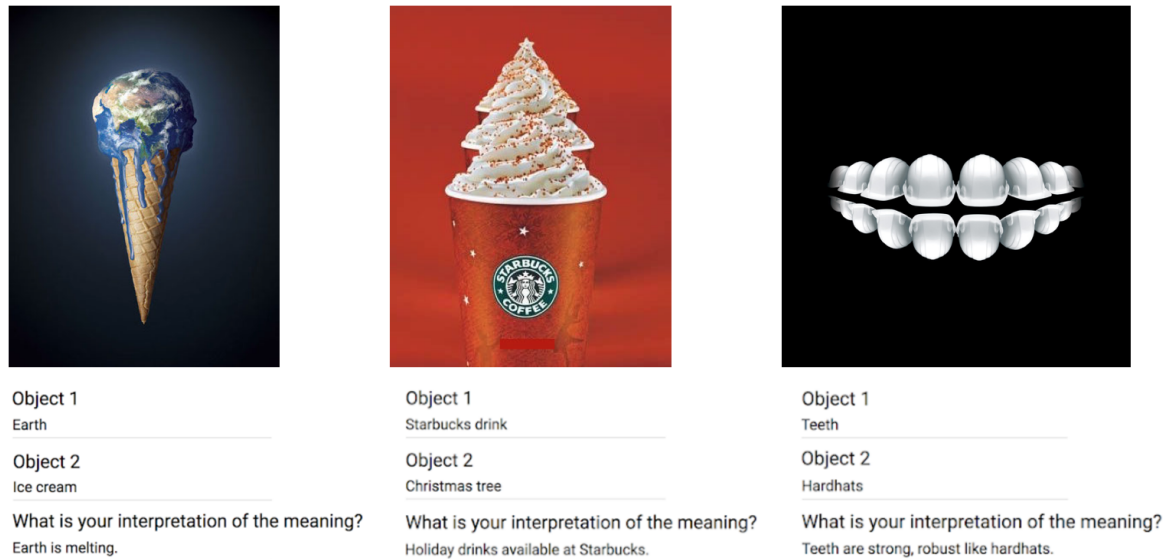
Figure 4. The three example metaphors in the experiment

receives the property *melt* from popsicle, conveying the meaning: "Iceberg (target) is melting (property)". These phrases were used as the gold labels for meaning. Finally, the authors removed each image's explanatory text by cropping it out or covering it.

Task

The task was implemented in a Google Form. Figure 4 shows three examples of the task. Consider the first example, consisting of a blend of the earth and ice cream. Each visual metaphor was displayed on exactly one page, scaled to at least 500 pixels in width. On each page, the workers were asked to do three things. They first identified the two objects that were blended in the image, such as the "earth" and "ice cream". Then they gave an interpretation of the meaning, like "Earth is melting." We had participants identify the two objects explicitly, as doing so was crucial for interpreting the meaning correctly. Finally, participants were required to give an answer for all three inputs and were not allowed to go back to previous metaphors and change answers.

We formulated the task in this way to elicit participants' natural interpretations of the meaning. We did not have participants explicitly label source and target domains, because it is normally clear from their interpretation of the meaning what they are. Additionally, introducing new terminology like source and target requires training and can be confusing. We wanted participants' natural reactions, rather than over analyzed meanings. For the same reason, we did not allow users to change answers during the study, as people normally do not spend time analyzing advertisements in the real world.

In the instructions, participants were first shown three visual metaphors with their two objects labeled and their intended meanings. This introduced participants to the task and provided examples for how detailed their interpretations should be. We also selected the images to cover the types of visual

metaphors in the data set: they are taken from different domains (advertisements, PSA's, book/magazine covers), transfer a variety of properties in different parts of speech, and contain both literal and implied targets. Figure 4 shows the three example images with their objects and meanings. The first example is a PSA on global warming, which compares the Earth to ice cream. In this ad, the literal term Earth, receives the verb: *melt*, from the implied term: ice cream. The second example is an advertisement, which compares a Starbucks coffee to a Christmas tree. Starbucks coffee, the literal term, receives the noun: *Christmas*, from the implied term: Christmas tree. The third example is an advertisement for toothpaste and compares teeth to hardhats. This time, teeth, the implied term, receives the adjective: *robust*, from hardhats, the literal term. By providing these very different examples, we hoped to prepare participants for the wide variety of meanings and visual combinations they would encounter in the 48-image data set.

ANALYSIS

RQ1: Can people interpret visual metaphors?

To address research question 1: if users could correctly determine the meaning of visual metaphors without text, we first labeled the intended meanings of each image based on its explanatory text. Each meaning could be condensed to the target term and the property it receives. For example, consider the popsicle and iceberg blend in Table 1. Here the iceberg receives the property *melt* from popsicle, conveying the meaning: "Iceberg (target) is melting (property)." To assess correctness we then compared the gold target and property with the target and property implied in each interpreted meaning.

Participants wrote their meaning as free text, and we judged the target and property from their interpretations. For example, one meaning provided for the French fry and bullet image in Figure 2 is "French fries are dangerous like bullets". The word

Meaning: "Iceberg is melting"

	User Input			Error Analysis			
	Object 1	Object 2	Meaning	Type of Error	Source, Target	Property	Target Symbol
1	Ice cream	Mountains	"Organic ice cream"	(1) Incorrect Objects	Mountains (S), Popsicle (T)	Organic	(none)
2	Wooden spoon	Iceberg	"Natural ice cream"	(2) Incorrect Direction	Iceberg (S), Popsicle (T)	Natural	(none)
3	Popsicle	Iceberg	"Icebergs are cold like popsicles"	(3) Incorrect Property	Popsicle (S), Iceberg (T)	Cold	(none)
4	Popsicle	Iceberg	"Winter is melting. Springs is here"	(4) Incorrect Target Symbol	Popsicle (S), Iceberg (T)	Melt	Winter
5	Ice cream	Glaciers	"Perception is everything"	No relationship to meaning	(none)	(none)	(none)
6	Popsicle stick	Iceberg	"The ice caps are melting"	None	Popsicle (S), Iceberg (T)	Melt	Iceberg

Table 1. Examples of four common types of errors, as well as a correct interpretation, and one that had no relation to the intended meaning.

dangerous is modifying one of the objects, french fries, so *dangerous* is the property. Also, the word *like* indicates that french fries are similar to bullets, in that they are dangerous. Therefore, french fries are the target, receiving the property *dangerous* from bullets. Although free text can be hard to interpret, we had three grammatical structures that helped us interpret them: "Object is [verb]", "Object is [adjective]", and "[adjective] object". We exemplify each formulation with the Iceberg popsicle image in Table 1. The first formulation is the form: "Object is [verb]", such as "The ice caps are melting". The verb, *melting*, is modifying the noun, ice caps, so ice caps is the target receiving the property *melting*. Since both target and property are correct, the meaning is correct. The second formulation uses an adjective instead of a verb: "Object is [adjective]", such as "Icebergs are cold like popsicles". The adjective *cold* is modifying the noun, icebergs, so icebergs is the target receiving the property *cold*. In this case, cold is the interpreted property, making the meaning incorrect. The third formulation uses an adjective, but places it before the object: "[adjective] object", such as "Natural ice cream". *Natural* is the property transferred to ice cream, the incorrect target. The fourth formulation involved phrases highly associated with the intended target and property. For example, the interpretation "Global warming" is highly associated with "Icebergs are melting". This is the only case that is difficult to interpret, and happened rarely. In these cases interpretation is abstracted away from the actual objects. We judged if these were correct if the interpretation was consistent with the metaphor. In this case, "Global warming" related to the message "icebergs are melting", so the interpretation is correct. By utilizing these rules, we were able to easily judge target and property from participants' free-text interpretations.

Results

After evaluating the 960 responses (20 participants * 48 images), we found participants correctly interpreted the meaning 41.32% of the time. Thus to answer research question 1, we find that while it is difficult, users can interpret the meaning of visual metaphors without their explanatory text.

There was significant variation in the percentage of correct answers across the 48 images. The average percentage of correct answers was 41% and the standard deviation was 31.2%. No image was interpreted correctly by all of the participants, the highest percentage of correct answers being 90%. About 8.33% of images were this successful, two of which are shown in Figure 1. The first conveys the meaning: "Tabasco is hot" and second conveys the meaning: "Don't gamble with your teeth". Meanwhile 12.5% of images had 0 correct interpretations, also shown in Figure 1. The first is "Nespresso wakes you up (like a bell)", and the second is "McDonald's anniversary". We assess how exactly participants misinterpreted these images in the following section.

RQ2: Types of errors

Now that we have determined people can interpret visual metaphors without their text some of the time, we address our second research question: when people misinterpret the meaning, what kind of mistakes do they make? In looking at the mistakes people make, we provide evidence for four distinct types of errors.

1. **Incorrect Objects.** This the most basic error. When the participant misidentifies one or both of the objects in the image, they will incorrectly interpret the meaning because either the source or target is an object that is not in the image. In Table 1, the two objects are popsicle and iceberg. Consider interpretation 1 in Table 1 for the same image, "Iceberg is melting". This participant saw mountains instead of icebergs, so now the perceived source object is actually not in the image. This error continues to propagate, as the interpreted property, *organic*, is coming from an erroneous source object: mountains.

2. **Incorrect Direction of Property Transfer.** If participants identify the correct objects, they then must correctly interpret the direction of the property transfer. In other words, they must correctly interpret which object is the target and which is the source transferring a property. In Table 1, the direction is popsicle (source) to iceberg (target). In interpretation 2 of Table 1, iceberg is mislabeled as the source and
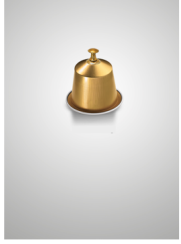
Meaning	"Nespresso wakes you up"	"Listen to your body"	"Botox is a cure-all"	"Brazil takes off"
Most Frequent Error	Incorrect Object Identification	Incorrect Direction	Incorrect Property	Incorrect Target Symbol
Image				
Examples of main error	"Time to eat"	"Headphones for women"	"Botox is easily accessible like a pocket knife"	"Christianity is blasting off"
	"Dinner is ready"	"Music is life."	"Botox is dangerous"	"Christianity is taking off in the hearts and minds of people"

Table 2. Examples of images that had many errors of one type.

ice cream is mislabeled as the target. Since the direction has been reversed, the property, *natural*, now erroneously comes from the intended target.

3. **Incorrect Property.** If a participant correctly identifies the objects, source and target, they must still correctly interpret the property transferred from source to target. In Table 1, the property transferred from Popsicle to Iceberg is *melt*, conveying the meaning: "Iceberg is melting". The third interpretation in Table 1 correctly identifies iceberg as the target and popsicle as the source, but misinterprets the property to be *cold*, conveying the wrong meaning: "Icebergs are cold like popsicles."

4. **Incorrect Target Symbol.** Even if a participant correctly identifies the objects, source, target, and property, they can still misinterpret what the target represents. In the "Iceberg is melting" image of 1, the iceberg is meant to be interpreted literally. Meanwhile, the fourth line in Table 1 interprets the iceberg as winter, and that the image represents winter melting away for spring. This participant erred crucially in his interpretation of the iceberg, while understanding every other component of the metaphor.

Before we could assess the origin of error in each interpretation, we needed to extract and evaluate the objects, source, target, and property of each interpretation. Evaluating object identification was straightforward, since we had participants explicitly label the objects they thought were in the image. For each image we compared the objects identified by the participants with those actually blended in the metaphor. We allowed responses directly related to the actual objects in the image, as there was an understandable variety in the objects identified. For example, in the "Iceberg is melting" PSA in Table 1 (object 1: popsicle, object 2: iceberg), we accepted answers like "ice cream" and "wooden spoon" for popsicle, as these are both highly related objects. Similarly, we accepted responses like "glaciers" for icebergs. This enabled us to handle the variety of responses for each object.

Identifying source object stemmed directly from how we identified target and property, when we addressed research ques-

tion 1. We judged target and property using the grammatical cues: "Object is [verb]", "Object is [adjective]", and "[Adjective] object". After we judged the target and property of a participant's interpretation, we determined if the property was reasonably associated with the other object. If it was, we would assume that object to be the source. For example, consider the interpretation "natural ice cream" for the "Icebergs are melting" image in Table 1. As we have seen before, this meaning falls into the "[Adjective] object" category and thus ice cream is the target and *natural* is the property. *Natural* is reasonably associated with iceberg, so we label iceberg as the source. Evaluating meanings in this way let us identify the source object even when it was not explicitly stated in the interpretation.

We found a number of participant's meanings unrelated to the objects and meaning of the image. Consider interpretation 5 in Table 1: "Perception is everything". This interpretation does not include either of the objects, and thus we cannot identify a target or source. In these cases, the meaning is labeled as incorrect as it is completely removed from the meaning of the image, and we leave target, source, and property unlabeled. We handle all interpretations unrelated to the meaning in this manner.

Error type (%)				
Incorrect Objects	Incorrect Direction of Property Transfer	Incorrect Property	Incorrect Target Symbol	No relationship to meaning
33.7%	13.5%	34.9%	15.2%	17%

Error type (%), for interpretations that correctly identified the objects			
Incorrect Direction of Property Transfer	Incorrect Property	Incorrect Target Symbol	No relationship to meaning
19.5%	51.8%	21.2%	27.6%

Table 3. Percentage of each error.

Results

With these four errors defined, we were able to assess how often they occurred across the data set. In Table 3, we summarize these results. Of the 960 interpretations: 33.7% identified

the objects incorrectly, 13.5% reversed the direction of property transfer, 34.9% identified an incorrect property, 15.2% misidentified the target symbol, and 17% had no relationship to the meaning.

About a third of the errors were due to participants misunderstanding the objects being compared in the image. Consider the "Nespresso wakes you up" image in Table 2. Nespresso receives the property *wakes you up* from bell. Participants all identified the bell correctly, but not one identified the Nespresso capsule. Instead, most users saw a dinner plate and cover blended with the bell. Their interpretation then became: "time to eat". Perhaps these participants are not familiar with Nespresso (see Discussion).

When we removed interpretations that identified the objects incorrectly, we found that of the remaining errors: 19.5% inferred the incorrect direction, 51.8% identified an incorrect property, 21.2% identified an incorrect target symbol, and 27.6% had no relationship to the meaning. From this we can see that users generally had trouble with each error, with incorrect property being the most common.

Some images suffered mostly from one particular error. For example, the "listen to your body" image in Table 2 was mostly interpreted in the incorrect direction. The two objects blended are uterus and headphones. Uterus receives the property *listen* from headphones. However, people interpreted the headphones as the target, receiving the property *woman* or *life* from uterus, obtaining meanings like: "headphones for women" and "music is life". The "Botox is a cure-all" image in Table 2 mainly suffered from incorrect property identification. In this metaphor, Botox receives the property *cure-all* or *many uses* from the Swiss Army Knife. Even though participants all understood the direction, they associated the Swiss Army Knife more with properties like *dangerous* and *accessible*, obtaining meanings like: "Botox is easily accessible like a pocket knife" and "Botox is dangerous". Finally the "Brazil takes off" image in Table 2, predominantly suffered from incorrect target symbol. This visual metaphor blends Christ the Redeemer, the iconic Brazilian statue, with a rocket. Here Brazil receives the property *takes off* from the rocket. Participants understood that the statue was the target, the rocket was the source, and *takes off* was the property, but they interpreted the statue either literally or representative of Christ, leading to interpretations like: "Christianity is blasting off". Perhaps Christ the Redeemer is not a good symbol for Brazil, as it is too associated with Christianity. We touch on this more in the discussion. Overall, certain images had significant difficulties with a specific error.

To answer research question 2, all errors featured prominently across the 960 interpretations collected in this study, with incorrect object and property identification featured most frequently. We noted that some images had many errors of a specific type. That being said, does the domain of a visual metaphor (ad, PSA, journalism) affect the frequency of particular errors?

RQ3: Property transfer (ads vs non-ads)
Forceville noted that people could more easily interpret the direction of property transfer when the visual metaphor was from an advertisement [7]. His reasoning was that advertisements must be making a statement about the product, and thus, the meaning is more likely "[product] is [property]", where the product is the target. Thus, we test if people can interpret direction correctly more often for advertisements than for PSA's and news articles (non-ads).

Results
Our data set consisted of 27 ads and 21 non-ads. For the ads, participants interpreted property direction correctly in 413 of 452 (91.4%) times. As Forceville thought, this is quite high. However, for non-ads, participants interpreted property direction correctly 304 of 344 (88.4%) times. A Chi-square independence test indicates no significant difference between the proportion of ads with correct direction and the proportion of non-ads with correct direction: $\chi^2(1) = 1.96$, $p = 0.16$. Therefore, to answer research question 3, our results do not support the idea that property direction is more easily identifiable for ads than for non-ads.

It is possible that the cues people use to infer direction in ads also exist for non-ads. For example, in ads, people use the cue that the product is probably the target (because it is an ad and ads send messages about their product). However, in non-ads, like PSAs, people may infer that a social issue (like the environment) is probably the target. Thus, Forceville may be right that contextual clues help us determine property direction. He just has not considered how widely this applied.

RQ4: Does familiarity affect interpretability?
Perhaps the interpretability of an image has less to do with its objects, direction, and property and more to do with how familiar its message is to its viewers. While evaluating interpretations, we found that metaphors with many correct interpretations seemed to have very familiar meanings, like "save the environment", while those with far fewer correct interpretations seemed to have obscure meanings like "Brazil takes off". Studies have shown that people can read words even when they are misspelled because they have strong priors on what the meaning will be [18]. A similar effect might be at play here, in which participants might be applying meanings they anticipate to images, paying less attention to the particular symbols and how they have been visually combined. We now address research question 4: does the familiarity of a visual metaphor's meaning predict how often people can interpret it?

To measure the familiarity of a meaning, we used the number of results returned when searching it on Google. This is a common approach in assessing how prevalent a topic is [16]. There were many options for choosing the phrase we searched on Google. We could have used the actual metaphor ("Popsicle is Iceberg") the text accompanying the image ("Icebergs are melting"), or the correct interpretations made by the participants ("global warming"). We used participants' correct interpretations as the search terms because we wanted to capture the fact that users get some messages correct precisely because they connect the image to a bigger cultural idea:
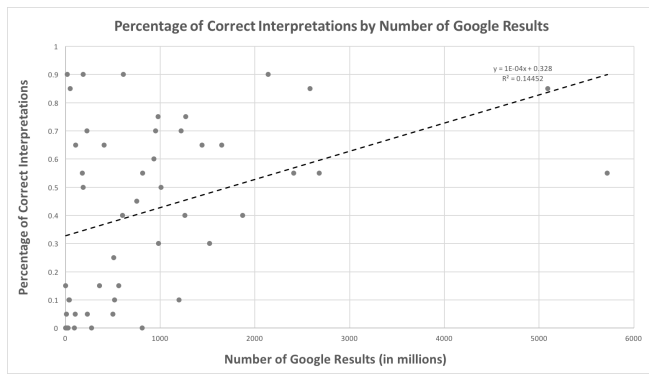
Percentage of Correct Interpretations by Number of Google Results

Figure 5. Scatter plot and regression line for Percentage of Correct Interpretations by Number of Google Results

like connecting melting icebergs (which has only 1.2 million results) to global warming (190 million results).

Results

Each image was now associated with a percentage of correct interpretations and a familiarity score, measured in millions of Google results. A linear regression was calculated to predict the percentage of correct interpretations based on the number of Google results. A significant regression equation was found ($F(1,46) = 7.771$, $p = 0.007$), with an R^2 of 0.144. The predicted percentage of correct answers is equal to $0.328 + .0000998$ (number of Google results in millions). An image's percentage of correct answers increased by .0000998 for each $1,000,000$ Google results. The line and scatter plot are shown in Figure 5. To answer research question 3, we find evidence that familiarity of the message correlates with the percentage of correct interpretations.

DISCUSSION

Improving visual metaphor understanding and creation

In this paper we have assessed 960 interpretations and have provided evidence for four common errors people make when interpreting visual metaphors. We discuss key points that could minimize each error and make visual metaphors more interpretable:

Object Identification

Incorrect object identification is the most basic error and was the second most common misinterpretation, making up 33.7% of errors. For example, in the "Nespresso wakes you up" of Figure 2, no participant could identify the Nespresso capsule. To better ensure an object is identifiable in a visual metaphor, designers should be careful not to blend that object too much with another. Removing too many identifying features like logo and color could render a product undetectable. For example, consider the "Absolut Vodka is dress" image in Figure 2. In this image, the only defining feature of the vodka bottle is its shape. It has been painted white, a color associated with elegant dresses, but in exchange, has lost its logo and the clear color of the vodka. Because of this, many of the viewers thought that the vodka bottle was actually a milk bottle and completely misinterpreted the meaning. In contrast, the Starbucks cups shown in Figure 4 have the Starbucks logo on them, which effectively identifies the brand and product.

Designers should ensure that the products in the blend each have prominent defining features so that they are interpretable.

Models for automatically understanding images should be able to identify objects without all of their visual features. For example, in order to automatically understand the "Absolut Vodka is dress" metaphor, the model should be able to identify the bottle by its shape and the dress by its color and mount. By identifying these two objects, the model can then develop an idea of the metaphor's target and purpose. With some priors about the symbolic meaning and sentiment of certain objects, the model could identify the dress as a positive symbol and thus determine that the image is a Vodka advertisement and not an alcoholism PSA. It is difficult however to identify stylized objects without all their features, but at the same time, it is essential. The objects provide significant insight towards the purpose and meaning of the advertisement.

Property Identification

Identifying the property transferred was the most common misinterpretation, making up 34.9% of errors. The cognitive process for interpreting the properties in visual metaphors is still an open problem. Participants could be identifying the direction of the property transfer first, then determining the property. Or participants could be interpreting the property first then determining the direction of transfer. However, as we show in study 4, familiarity of the meaning determines how often a visual metaphor is understood, and familiar symbols of abstract properties may be hinting at these common meanings.

In order for people to better interpret the property in visual metaphors, designers should use familiar objects that symbolize properties commonly used. For example, in many visual metaphors, guns and grenades are used to symbolize the property: *deadly*, and batteries are used to symbolize the property: *energy*. In the same vein, models for automatic ad understanding should be able to identify objects that are used to symbolize familiar properties.

The culture of the viewer matters

The background and culture of the viewer greatly affects their understanding and appreciation of the image [8] [21] [19] [7]. Interpreting visual metaphors requires that viewers can identify the objects in the image as well as their symbolic meanings. Therefore, a person's ability to decipher a visual metaphor's meaning is dependent on whether they have seen the objects and have learned their symbolic associations in the past. For example, consider the "Brazil takes off" image in Table 2. We are not surprised that participants in this study failed to interpret the Christ the Redeemer statue as Brazil. Some saw a statue with outstretched arms. Others recognized this was a statue of Jesus. But no one saw the statue as a symbol for Brazil. This metaphor was originally on a cover of the The Economist and thus was intended for their readers. Perhaps those who read The Economist would have understood the statue's symbolic meaning. One limitation of our study is that we picked random workers in the United States to interpret our data set. We did not target specific populations like Economist readers and advertisement creators. Thus we could not guarantee that the participants had the right cultural background to interpret these images.

It is important for designers to be aware of the culture of their audience when creating visual metaphors. The culture of their viewers directly affects what objects they can use as symbols. Culture and familiarity of meanings is also hugely important for automatic understanding of advertisements in both creating training data and for the algorithms classifying advertisements. The data sets created to train models for advertisement understanding were crowdsourced. Random Amazon Mechanical Turk workers were assessing the meanings of metaphors from a variety of countries and cultures. This data could be made better by having people interpret advertisements for products that they know and by having people interpret advertisements intended for their culture. This way people would hopefully know the symbols and cultural references being used and would be able to supply a more specific meaning for the image.

The algorithms for classifying the meaning of advertisements would also benefit from knowing the symbols pertaining to a particular culture. At the same time, these models could have a prior on common meanings depicted in certain cultures. A model could correctly interpret Christ the Redeemer as a symbol of Brazil for The Economist readers and as symbol as faith for other viewers. Knowing the advertisement's intended audience lends information on how to interpret it.

LIMITATIONS AND FUTURE WORK
We asked Mechanical Turk Workers to write their interpretation of visual metaphors. There are several possible flaws with this. First, all the images appear online and some probably appear in print. It is possible that they had seen the images before, with their surrounding text, thus they might know what the images meant and weren't inferring it. However, we think this is unlikely. More importantly, Mechanical Turk Workers may or may not be the target audiences of the visual metaphors. This could impede their ability to interpret the symbols or the messages. This is possibly the case for the The Economist cover about "Brazil Takes Off", or for the Nespresso ad. However, a majority of the ads are for fairly broad US audiences like Tabasco, McDonald's, Starbucks, and anti-smoking.

For research question 4, we used the number of Google results for a phrase as a measure of familiarity. This is an imperfect measure, and there are certainly messages that seem very familiar like "Global warming" that got comparatively few Google results. Although this increases the variance, it does not pose a great threat to validity. A bigger threat would be if there were something correlated with Google results that were actually driving the increase in interpretability. For example, if an ad were popular and widely blogged about, it would have many Google results and high familiarity as well as interpretability simply because it was newsworthy.

These studies have not fully explained why some visual metaphors are more interpretable than others. There is much more future work to be done to test other hypotheses. For example:

- The background color of the image may give a clue as to whether the message is positive or negative. Ads are typically positive and PSAs are typically negative, so this is

a visual contextual clue separate from the objects that may contribute to interpreting their meaning.

- The appearance of motion may indicate which object is the source, and thus what the direction of property transfer is. Objects that are in motion like a rocket taking off, ice cream melting, or a dandelion florets blowing away indicate that it symbolizes a verb. Verbs are always the source in visual metaphors, which imply the other object, like the earth receives the property of melting.

- Visual metaphors have an object that is seen literally, like the french fries in panel 2 of Figure 3 and an object that is implied (cigarettes). In 43 of the 48 images the literal object is the target. In the remaining 5 of 48 images the literal object is the source. It is possible that using the literal object as the target helps people interpret direction. However, this is a challenging hypothesis to test because finding examples of each case is hard.

It is likely that no single visual cue will fully indicate the meaning, however, by using multiple cues together it may increase the probability that more individuals will see one of the cues and thus interpret it correctly.

CONCLUSION
Visual metaphors are a creative technique used in print media to convey a message through images. This message is not said directly, but implied through symbols and how those symbols are juxtaposed in the image. The messages we see affect our thoughts and lives, and it is an open research challenge to get machines to automatically understand the implied messages in images. We find people can infer visual metaphors without their surrounding text and do so 41.3% of the time. A major source of errors is actually quite basic - that viewers don't recognize the objects in the blend. This is potentially easy to check for and improve. The other major source of error is when viewers correctly identify the objects, and the direction of property transfer, but infer the wrong property. This is a hard problem. We find that when the message is already familiar people are more likely to interpret it correctly. They are probably using their world knowledge as a factor in interpreting it. Thus, automated approaches should also take world knowledge into account and not just rely solely on the information in the image. We discuss future work to explore what other smaller cues people or computers could use to enhance interpretation of visual metaphors including extra context from the background color, signs of motion to indicate one object is a verb, and using the literal object as the target.

REFERENCES
[1] Chun-Tuan Chang and Ching-Ting Yen. 2013. Missing ingredients in metaphor advertising: The right formula of metaphor type, product type, and need for cognition. *Journal of advertising* 42, 1 (2013), 80–94.

[2] Lydia B. Chilton, Savvas Petridis, and Maneesh Agrawala. 2019. VisiBlends: A flexible workflow for visual blends. In *Proceedings of the 2019 ACM Conference on Human Factors in Computing Systems*

(CHI '19). ACM, New York, NY, USA. DOI:
http://dx.doi.org/10.1145/3290605.3300402

[3] Joao M Cunha, Joao Gonçalves, Pedro Martins, Penousal Machado, and Amílcar Cardoso. 2017. A pig, an angel and a cactus walk into a blender: A descriptive approach to visual blending. *arXiv preprint arXiv:1706.09076* (2017).

[4] Joao M Cunha, Pedro Martins, and Penousal Machado. 2018. How shell and horn make a unicorn: Experimenting with visual blending in emoji. In *Proceedings of the Ninth International Conference on Computational Creativity*.

[5] Charles Forceville. 1994. Pictorial Metaphor in Advertisements. *Metaphor and Symbolic Activity* 9, 1 (1994), 1–29. DOI:
http://dx.doi.org/10.1207/s15327868ms0901_1

[6] Charles Forceville. 2002. The identification of target and source in pictorial metaphors. *Journal of Pragmatics* 34, 1 (2002), 1–14.

[7] Charles Forceville and others. 2017. Visual and multimodal metaphor in advertising: Cultural perspectives. *Styles of Communication* 9, 2 (2017), 26–41.

[8] Jos Hornikx and Rob le Pair. 2017. The influence of high-/low-context culture on perceived Ad complexity and liking. *Journal of Global Marketing* 30, 4 (2017), 228–237.

[9] Zaeem Hussain, Mingda Zhang, Xiaozhong Zhang, Keren Ye, Christopher Thomas, Zuha Agha, Nathan Ong, and Adriana Kovashka. 2017. Automatic understanding of image and video advertisements. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 1100–1110.

[10] Se-Hoon Jeong. 2008. Visual metaphor in advertising: Is the persuasive effect attributable to visual argumentation or metaphorical rhetoric? *Journal of marketing communications* 14, 1 (2008), 59–73.

[11] Allison J Lazard, Benita A Bamgbade, Jennah M Sontag, and Carolyn Brown. 2016. Using visual metaphors in health messages: a strategy to increase effectiveness for mental illness communication. *Journal of health communication* 21, 12 (2016), 1260–1268.

[12] RG Le Pair and MJP Van Mulken. 2008. Perceived complexity and appreciation of visual metaphors by consumers with different cultural backgrounds. (2008).

[13] Edward F McQuarrie and David Glen Mick. 1999. Visual rhetoric in advertising: Text-interpretive, experimental, and reader-response analyses. *Journal of consumer research* 26, 1 (1999), 37–54.

[14] Edward F McQuarrie and Barbara J Phillips. 2005. Indirect persuasion in advertising: How consumers process metaphors presented in pictures and words. *Journal of advertising* 34, 2 (2005), 7–20.

[15] Marijn HC Meijers, Paola Remmelswaal, and Anke Wonneberger. 2018. Using Visual Impact Metaphors to Stimulate Environmentally Friendly Behavior: The Roles of Response Efficacy and Evaluative Persuasion Knowledge. *Environmental Communication* (2018), 1–14.

[16] Jean-Baptiste Michel, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K Gray, Joseph P Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, and others. 2011. Quantitative analysis of culture using millions of digitized books. *science* 331, 6014 (2011), 176–182.

[17] Barbara J Phillips and Edward F McQuarrie. 2004. Beyond visual metaphor: A new typology of visual rhetoric in advertising. *Marketing theory* 4, 1-2 (2004), 113–136.

[18] Keith Rayner, Sarah J White, and SP Liversedge. 2006. Raeding wrods with jubmled lettres: There is a cost. (2006).

[19] Elisabeth El Refaie. 2003. Understanding visual metaphor: The example of newspaper cartoons. *Visual communication* 2, 1 (2003), 75–95.

[20] Renske van Enschot and Hans Hoeken. 2015. The occurrence and effects of verbal and visual anchoring of tropes on the perceived comprehensibility and liking of TV commercials. *Journal of Advertising* 44, 1 (2015), 25–36.

[21] Margot Van Mulken, Rob Le Pair, and Charles Forceville. 2010. The impact of perceived complexity, deviation and comprehension on the appreciation of visual metaphor in advertising across three European countries. *Journal of Pragmatics* 42, 12 (2010), 3418–3430.

[22] Margot Van Mulken, Andreu van Hooft, and Ulrike Nederstigt. 2014. Finding the tipping point: Visual metaphor and conceptual complexity in advertising. *Journal of Advertising* 43, 4 (2014), 333–343.

[23] Stephanie K Van Stee. 2018. Meta-analysis of the persuasive effects of metaphorical vs. literal messages. *Communication Studies* 69, 5 (2018), 545–566.

[24] Thomas Whiteside. 1971. *Selling death: cigarette advertising and public health*. Liveright.

[25] Keren Ye and Adriana Kovashka. 2018. Advise: Symbolism and external knowledge for decoding advertisements. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 837–855.