# **Coloring and Drawing Spot:**
# **The Impact of Perception on the Accurate Depiction of a Robot Dog**

Dave Feltner and Doug Jonidis

North Carolina State University

PSY 500: Visual Perception

Dr. Yingchen He

# Introduction

Literature on visual perception and attention has consistently demonstrated the relationships between these two processes, suggesting that attention lies at the crossroads between the taking in of perceptual information and the incorporation of that information into the conscious experience (Carrasco, 2011). The relationship between attention and visual perception can be explained through multiple resource theory, where human performance is supported through a pool of finite cognitive resources (Wickens, 2008). In terms of selective attention, the focus of attention results in a greater availability of cognitive resources to the target location, although at the expense of resources distributed to the unattended location, as a result of the finite pool of total resources (Carrasco, 2011).

In terms of visual perception, the tradeoff relationship between attended and unattended regions regarding cognitive resources has been demonstrated by studies showing retinotopically specific neural signal enhancement at the focal point of attention and a converse reduction in signal strength in the same location when attention is placed elsewhere, referred to as the biased competition theory of selective attention. (Beck & Kastner, 2007). The assumptions of biased competition theory are as follows: representation of stimuli in the visual system is competitive, both top-down and bottom-up processes bias this competition, and the competition is integrated across cognitive processes. In essence, attention allows for the optimization of the visual system by enhancing the representation and perception of relevant stimuli, while diminishing that which is outside the focus of attention (Carrasco, 2011; Smith, Singh, & Greenlee, 2000).

This study demonstrates how verbal cueing affects visual perception through attention. Through a two part study, participants will hear verbal cues that generate selective attention towards a specific body part of a robot. After being cued towards a specific area, participants will attempt to discern contrast patterns in the robot's image and test their ability to accurately draw the image. Previous research indicates that the selective attention, caused by the cue, will make that cued area brighter, more noticeable, and perceptually larger in memory.

## Contrast

Luminance contrast, the difference in luminance between an object and its background, has been identified as an appropriate variable of interest for the relationship between attention and visual perception under contrast discrimination paradigms (Carrasco, 2004). Evidence consistent across both experimental and neurophysiological trials indicates an increase in both perceived contrast and neuronal contrast sensitivity (contrast gain) for attended stimuli, relative to that which is unattended (McAdams & Maunsell, 1999)

In terms of cueing visual attention, verbal cues have been shown to increase perceptual sensitivity (d') in object discrimination tasks, quite literally leading to an affect where hearing a word made otherwise invisible stimulus visible, an effect that was not found to be present when visual cueing was used for the same stimulus (Lupyan & Spivey, 2010).

## Visually Accurate Size Drawings

Drawing objects that you observe is a multi-stepped process from attending to the object, encoding it, and recalling it to accurately re-create the object (Purves & Howe, 2005). A visually accurate drawing is "one that can be recognized as a particular object…rendered with little addition of visual detail that cannot be seen in the object represented or with little deletion of visual detail (Cohen & Bennett, 1997, p. 609). To create an accurate drawing, people need to utilize a set of visuo-cognitive skills to deploy their attention, mental imagery, and visual memory (Lou, 2018); however, this task is difficult and fraught with errors when people attempt to create that visually accurate drawing.

The errors of people's perceptual judgments do not just come from a lack of drawing skills; the misperception hypothesis indicates that drawing inaccuracy is related to an inaccurate perceptual encoding or recall of a stimulus. Despite participants being instructed beforehand to selectively attend to specific stimuli within a cluttered visual field, their errors were spread across all objects that they drew (Ostrofsky et al., 2015). This indicates that initial instructions are insufficient in encoding specific visual stimuli. Additionally, literature has shown that as more selective attention was given to an object, that object took up an inordinate amount of space of the drawing which presented as erroneously enlarged aspects of their drawings (Pepperell and Haertel, 2014; Mitchell et al., 2005). The drawing errors could also be attributed to selective attention generating more salient memories. As people attempt to accurately draw what they recently draw, those areas with the most salient memories are drawn disproportionately larger than other areas (Gazzaley & Nobre, 2012). Therefore, as selective attention is maintained on specific portions of an image, it is larger in memory and therefore re-created larger in later renderings.

## Present Study

This study will be completed through a pilot study followed by a two-part experiment. The pilot study serves to validate that the verbal cues generate selective attention towards the intended AOIs along with the boundaries of the AOIs themselves. The study will then be a two-part within-subjects experiment with Part 1 focused on Contrast and Part 2 focused on Size Perception. Part 1 of the experiment employs a 4x2x2 within-subjects factorial design where the independent variables are Gabor location (Head, Body, Legs, None), attentional cue (specific, neutral), and Gabor orientation ("/", "\"). Part 2 of the experiment employs a 3x3 within-subjects design where the independent variables are cued body parts (Head, Body, Legs) and image size (Small, Medium, Large). Based on previous research we hypothesize the following:

H1: Under the specific attentional cue condition participants will perceive the physical contrast of the test and standard patches to be equal when the physical contrast of the test patch is at a lower physical contrast than the standard patch.

H2: Participants are able to detect smaller contrasts in cued Areas of Interest.

H3: Participants have more errors when drawing the portion of the body that was cued.

This experiment addresses a significant gap in the literature. While both size and contrast have been used throughout visual perception of basic research, the specific effects of verbal cues have not been investigated on them. Through this experimental paradigm, we explore the efficacy of verbal cues beyond Lupyan and Spivey's (2010) object discrimination. By addressing this gap and understanding the effects of verbal cues across the domain of visual perception, we can better understand future implications of verbal cues in design.

# Methods

## Pilot Study

Fifteen undergraduate students will complete the pilot study to validate that all verbal cues are clear along with ensuring that the cues generate selective attention on the intended AOIs, as well as the incorporation of neutral cues (not meant to consistently direct attention) for use in Part 1. Participants will be screened to meet the inclusion criteria, participants must have 20/20 corrected vision, have full color vision, and have no formal training in art or drawing.

### *Materials*

The verbal cues will indicate which part of Spot that the visual attention will be cueing.

Below are example cues for each body part, with each verbal cue containing two separate references to that same portion of Spot. Participants are cued with statements like those seen below in Table 1. As they are cued, participants' visual attention will be measured through eye tracking on Spot's AOIs. The initial AOIs as depicted in Figure 1 have different surface areas and will be refined based on the eye tracking data from the pilot study.

**Table 1**

*Verbal Cues to Generate Selective Attention*

| Body Part | Cue |
|---|---|
| Head | Spot looked around at the intersection, trying to see if there were any people hiding behind the trees. |
| Head | Spot heard people nearby in the woods. The noise only seemed to be about 30 feet away. |
| Head | Spot smelled campfire smoke coming from its right; the pine needles gave off a unique odor. |
| Body | Spot carried additional equipment to sustain itself for the mission; the additional gear did not weigh Spot down. |
| Body | Spot's back tensed as it prepared to take on additional weight in it's carrying bag; Spot is able to carry over 100 pounds on its back. |
| Body | Spot's radio antennae on it's back extended higher so it could get the message through the dense forest. |
| Legs | Spot walked through the mud at the same pace that it walked through open fields due to special footwear. |
| Legs | Spot stepped in the shallow water as it attempted to cross an intermittent creek in the woods. |
| Legs | Spot walked through the wooded area for the patrol, where Spot had to step over fallen tree branches on the route. |
| Neutral | Spot is manufactured by the Boston Dynamics Company. |
| Neutral | Spot is able to withstand dust, rain, and tempreatures as high as 125 degrees. |
| Neutral | Spot can perform a variety of applications and tasks. |

**Figure 1**

*Initial AOIs on Spot*

*Note.* The different colors represent the Head, Body, and Leg AOIs.

***Procedure***

After the pilot participants meet the inclusion criteria, they will be calibrated on an eye-tracking computer with a  chin rest to avoid any excess movement.  The purpose of this eye tracking is to ensure that each verbal cue generates selective attention - as measured through gaze time - on each of the cued AOIs.  Once calibrated, each participant will be read a verbal cue for 10 seconds while the eye tracker records their gaze and then have 5 seconds of silence.  The cues will be presented randomly until each cue has been read 5 times each.

Once complete with all 60 trials, a Repeated Measures ANOVA will reveal if the cues generate a visual attention on the intended AOIs.  Through this ANOVA we expect that the body-part specific cues will generate significantly more gaze time on their respective body-parts. Also, we expect to see no specific AOI receiving more gaze time in the neutral cues.

**Part 1 Perceived Contrast**

***Participants***

Based on an a priori power analysis of a Repeated Measures ANOVA for the proposed experiment in Part 2, 86 undergraduate students at a large university will be recruited to find an expected small effect size (d = .20). For Part 1, of the 86 participants recruited for Part 2, all will be asked to participate in a within-subjects forced-choice task. For this experiment a minimum sample size of 13 per trial condition has been found to be sufficient under similar experimental design (Carrasco, 2004). To meet the inclusion criteria, participants must have 20/20 corrected vision, have full color vision, and have no formal training in art or drawing.

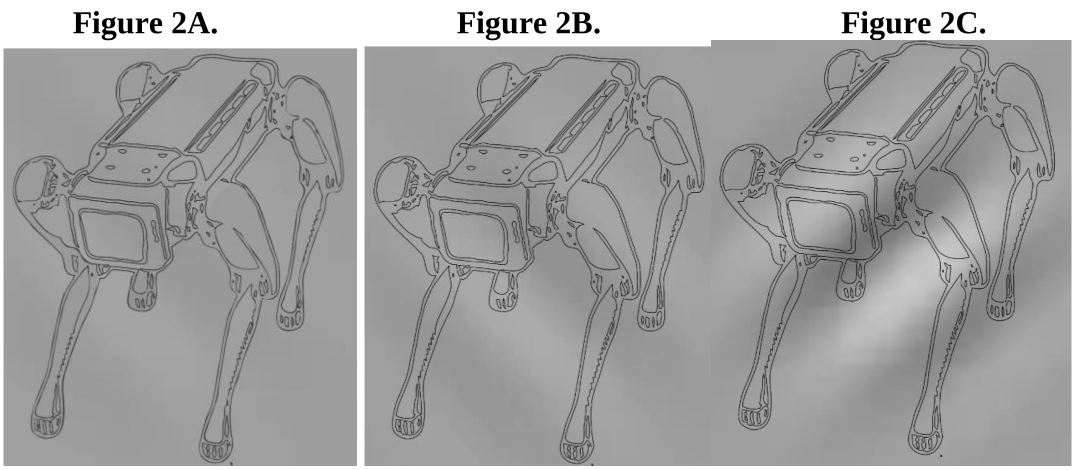
The stimuli used will be Gabor patches, sinusoidal gratings of 2 or 4 cpd with a Gaussian envelope with an orientation of 45° to the left or right. One patch (the standard patch) will be kept at a contrast near-threshold level of 6%  (defined using the Michelson Contrast of the brightest and darkest areas of each patch). The second patch (test patch)  will range from 2.5% - 16% across trials as this range of values has been found to be effective in similar experimental designs (Carrasco, 2004). The two patches will be laid over an outline of Spot the dog, depicted in Figure 2 below, with opposing orientations (i.e. one tilting left, another tilting right).

***Procedure***

To keep participants naive to this experimental purpose, rather than reporting contrast directly, participants' will instead be asked to report the orientation of the patch with the higher perceived luminance contrast. A 4x2x2, within-subject factorial design will be utilized (test patch location: head, legs, body, none, attentional cue: cued, neutral, orientation: “/”, “\”). For the attentional cues, the neutral condition will receive an irrelevant cue (unintended to direct attention to any one feature), and the specific cue condition will receive a verbal cue (seen above in Table 1) relevant to the given feature of the robotic dog in which the test patch has been overlaid. Verbal cues will be in the form of an audio message played prior to orientation selection. The cues used will be uninformative of either orientation or contrast.

**Figure 2**

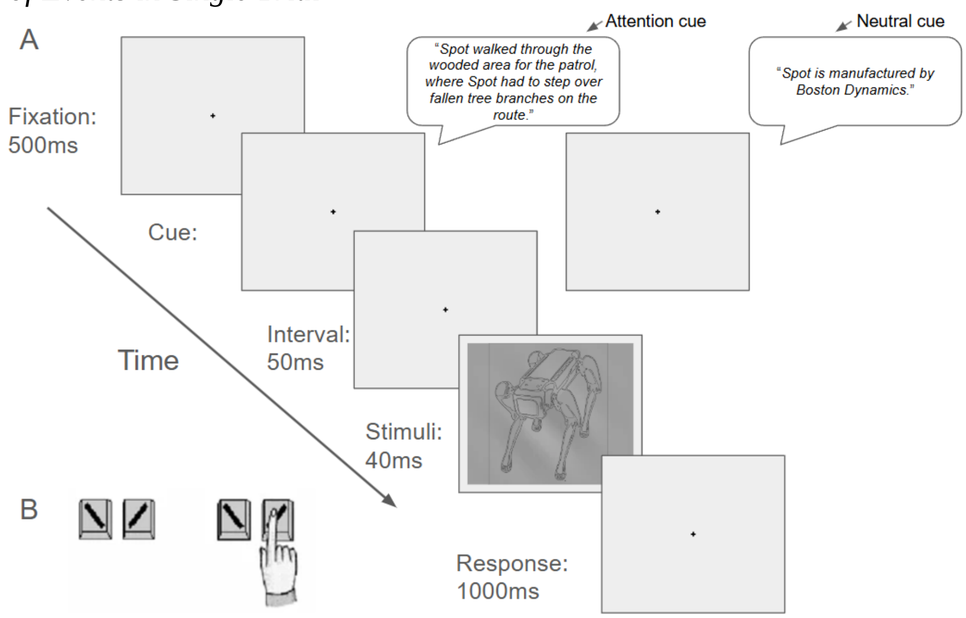
*Examples of Gabor Patches Overlaid on an Outline of Spot*




*Note*: Figure 2A: An example of two Gabor patches with differing orientations overlaid on Spot. The first with orientations starting from the top left going down to the bottom right is at a contrast of 6%% and is overlaid over the entire image except for the legs. The legs have been overlaid with a patch with the opposing orientation at a contrast of 3.5%. Figure 2B: Similar to 2A, but here the contrast of the patch traveling from top left to bottom right is shown at 8% for enhanced visual clarity. Figure 2C: shows a single patch overlaid across the entire image with a contrast value of 16%. The previous two examples are shown for demonstration purposes only.

**Figure 3**

*Sequence of Events in Single Trial*




*Note:* A: Each trial begins with a fixation, followed by a verbal cue. B: Participants will be given a response window of 1000ms.

As the research focus is on perceived contrast rather than orientation, our dependent variable of interest will be the selected area (i.e. the location perceived as having the highest contrast) relative to the actual physical contrast of the unselected areas (DV = $contrast_{unselected}$ - $constrast_{selected}$). The cues will be randomized, but will always correspond to the respective placement of the test patch for that trial. For the neutral condition, the test patch will simply be placed at random.

Participants will first perform a practice block of 75 trials before completing ten blocks of 200 trials each, which is estimated to take around 60 minutes.Participants will be instructed specifically that the higher contrast area will only be found on either the whole head, body, or legs such that they know that the contrast orientation between the four legs will be held constant. A chin rest will be utilized to avoid distraction and a gaze-contingent variable will be used to control for differences in gaze time across participants and trials. This will be recorded as total duration time on the three areas of interest (head, legs, face). Time spent outside of these areas will be considered as a fourth AOI.

As a result of the combination of verbal cueing effect on visual attention and the contrast gain effect of visual attention, it is hypothesized that under neutral cues , as physical contrast between the three locations of interest approaches equality, participant performance on the task will approach chance probability (reflecting a lack of perceivable difference in contrast when contrast for the three areas is held equal). Conversely, for under the specific cue  condition, as contrast between the locations of interest approaches equality, participant selection of the highest contrast area will be biased to the location cued in the verbal message, over and above that of chance (reflecting the subjective contrast gain resulting from an increase in visual attention to the area of interest).

***Data Analysis***

To model the change in contrast sensitivity, a nested hypothesis test will be conducted utilizing the local linear method. Hypotheses will be tested in a nested fashion, first assessing a single psychometric model across conditions for goodness of fit before comparing separate models for each condition. This will ensure that a significant difference in perceived contrast exists across conditions.

The local linear method is a method of modeling probability distributions of trial-over-trial human performance data without the need to make certain assumptions, such as the true guessing rate, which are assumed under similar techniques, such as the Weibull or Gaussian distributions (Zychaluk & Foster, 2009). The model has been assessed in its effectiveness in modeling psychometric functions compared to more traditional, parametric methods, and has been shown to perform often better but never worse than these methods (Zychaluk & Foster, 2009). The choice in selecting the local linear method is to effectively manage the non-traditional use of stimuli in our experimental procedure. As the method is non-parametric in nature, it is not necessary for the experimenters to make assumptions for which there is a lack of empirical support.

To model the relationship between verbal semantic information and contrast sensitivity, the function will be approximated locally, in a nearest-neighbor fashion such that for a given trial $t_1$, the value of the function $h(x)$ (the relationship between trial conditions and the response) is approximated within the neighborhood of time $t_0$ with the use of a Taylor expansion:

$$h(u) \approx h(t_0)+(u-t_0)h'(t_0)$$

where:

- $h(t_0)$ is the value of the function at $t_0$,
- $h'(t_0)$ is the first derivative of $h(t_0)$ at $x_0$,
- u is a point in the neighborhood of $x_0$.

Observations are then weighted using a kernel function to provide parameter estimates of $h(t_0)$ and the derivative $h'(t_0)$, $\underline{h}(t_0)$ and $\underline{h}'(t_0)$, respectively. These estimates are then used as intercept and slope parameter values in the modeling of the psychometric function. The estimate of the psychometric function P at time $t_1$, is then:

$$\underline{P}(t_0) = g^{-1}[\ \underline{h}(t_0)\ ]$$

where:

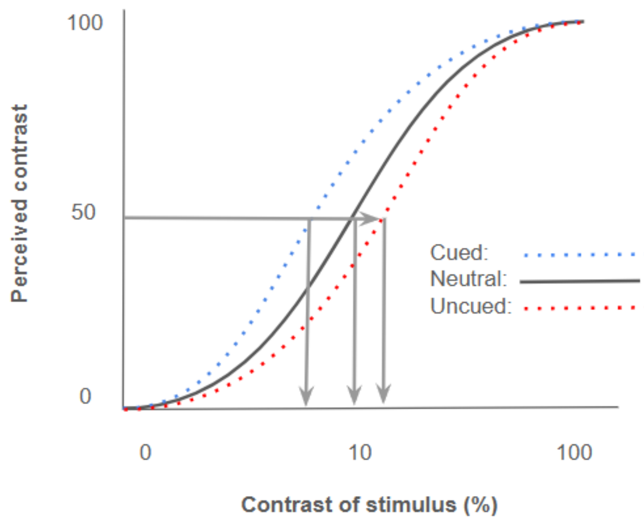
- $g$ is the link function (local log likelihood using the weights obtained from the kernel function) between observation and prediction for a given point $t$.

This is repeated for all values of $t$ such that $\underline{P}(t)$ sufficiently describes P (the true psychometric function between the experimental condition and contrast sensitivity). This function will then be used in a nested hypothesis test to examine for significant difference in performance by condition. (Local linear estimation with separate fits for each condition and compared against a single fit for both conditions). Thus, functions are obtained to describe the probability of a participant choosing the test patch in relation to the standard patch, as a direct function of the relationship between perceived and actual contrast. The value of the test contrast when the function reaches 50% represents the point of subjective equality (PSE), when guessing is predicted to be at chance).

Percentage of responses where the participant reports the contrast of the test patch as higher than the standard, will be plotted against the test patch's physical contrast for both conditions. Lines will be included to indicate the PSE for each gabor orientation used in the experiment.

**Figure 4**

*Hypothetical graph of the psychometric functions of the contrast experiment*




*Note:* Percentage of responses in which the participants select the test patch orientation over the standard patch, plotted as a function of the test patch's physical contrast. The grey arrows intersecting the curve represent the contrast values necessary for the test and standard patch to attain PSE.

To control for size discrepancy between the areas of the robotic dog model, a predictor variable is created using a difference score between the area of location in which the test patch is placed and the area selected by the participant, in terms of total pixel space. This will be used to control for possible effects of size on contrast sensitivity. Additionally, to control for gaze duration, pairwise difference scores between the selected location AOI and the remaining AOI's and will be assessed for significant differences.

## Part 3: Drawing Spot

### *Overview and Design*

As previously described, experiment 2 employs a 3x3 within subjects design. The independent variables are cued body part (Head, Body, Legs) as seen previously in Figure 1 and image size (Small, Medium, Large) seen below in Figure 2. The dependent measure will be the drawing size error based on the relative surface area of the cued area. For example, in the presented image the head AOI is 13% of the total body. If the participant draws a head that represents 25% of the total body surface area, then the drawing size error score is 12%.

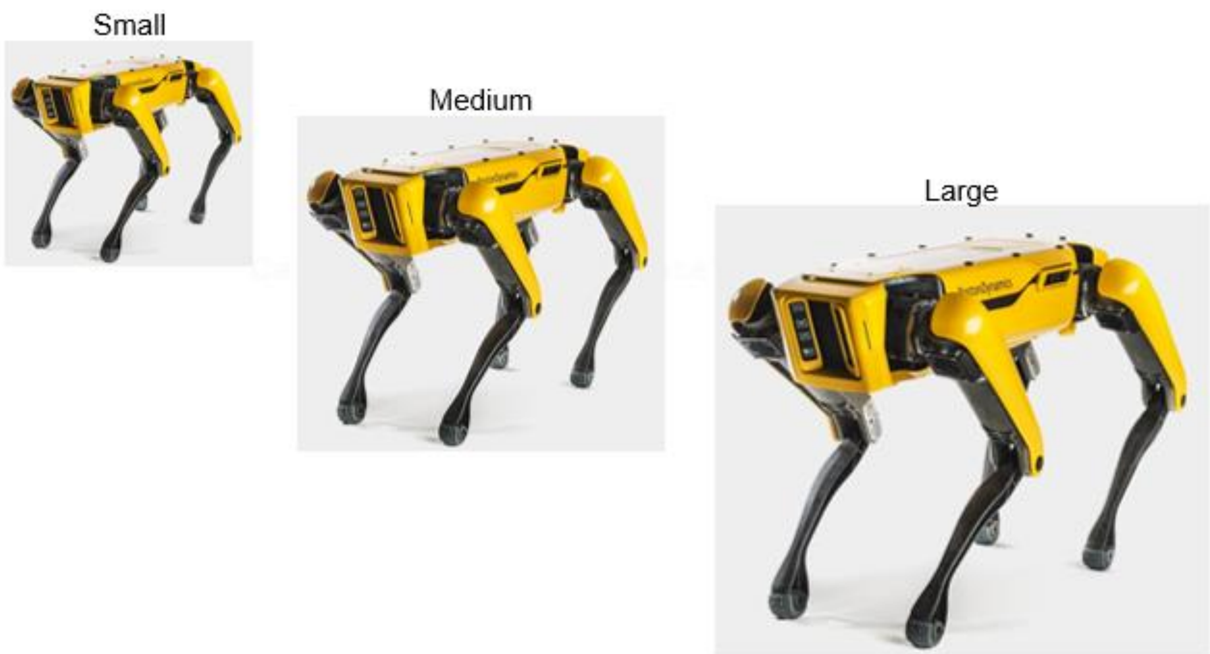
A drawing task was constructed with a 30 second exposure of Spot's image followed by 3 minutes of drawing. The verbal cue was repeated 3 times so that participants heard the cue during the entire 30 second image exposure. This exposure gave participants enough time to study the image and hear multiple references to the cued body part. Thirty seconds after the exposure, participants are given 3 minutes to draw the image as accurately as they can. The image size and verbal cue orders are counterbalanced in order to account for learning and practice effects.

***Stimuli***

The verbal cues listed in Table 1 will be refined based on the results of the pilot study and only the most salient cue for each body part will be used for the drawing task. These cues are chosen to precisely focus participants' selective attention on Spot's body parts. Spot's three images are seen below in Figure 2. The Small image is 1.5"x 1.5". The Medium image is 2.25" x 2.25", which is 50% larger than the Small image. The Large image is 3" x 3" which is 100% larger than the Small image. The images will be presented on a computer screen for the duration of their exposure.

**Figure 5**

*Small, Medium, and Large Spot Images*

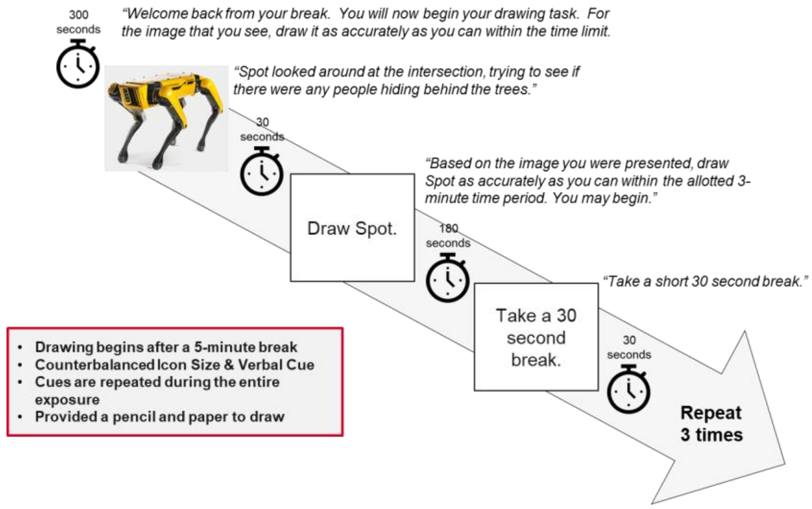



***Procedure***

After a 5 minute break, the same 86 personnel from Part 2 will start their drawing task which will take approximately 25 minutes. Participants will sit down in front of a computer and be given three sheets of paper along with a pencil. As seen below in Figure 6, the participants are briefed that they will be shown an image for 30 seconds and to draw the image as accurately as they can. During that 30 second exposure, participants will also hear a verbal cue repeated throughout the entire exposure. At the conclusion of the 30 seconds they have 3 minutes to draw Spot as accurately as they can within the allotted time period. After their 3 minutes has elapsed, the drawing will be turned in and the participant will take a 30 second break. The participant will complete a total of three drawings in this experiment; the presented image size and verbal cues are counterbalanced across participants to account for any learning or order effects.

**Figure 6**

*Example Procedure for the Drawing Spot Task*



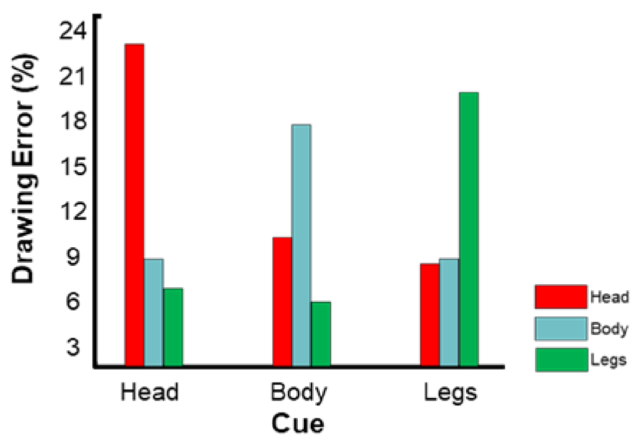

***Data Analysis***

A Repeated Measures ANOVA will be used to understand the differences that participants show in their drawings based on the cue that they receive. Descriptive statistics will be used to gather the mean and standard deviation of the error score as well determine the normality of the data.

Based on attention literature, we expect to see a main effect of cue across all three body parts (Head, Body, Legs) as seen below in Figure 7. For example, when participants hear the verbal cue indicating Spot's head we expect their drawing to have a proportionally larger drawing of the head. We do not expect any significant effect to come from the size of the image that the participant sees. Additionally, AOIs will be tested for any moderation effects they might have on the participants' visual attention and perception.

**Figure 7**

*Anticipated Drawing Size Results*



# Discussion

## Application

This basic research attempts to better understand how specific verbal cues impact people's visual perception. As researchers continue to develop the technology of robots in the military, they will refine their design and maintenance of the robots. The military requires that Soldiers can provide field maintenance to any new piece of equipment. Given the variety of conditions that Soldiers perform field maintenance, a simple design feature could help direct Soldiers' attention at the most pertinent time. For example, if a semi-autonomous robot detects a small unknown error, then it can give a verbal cue to its operator directing their attention. This captures the Soldiers' attention and directs it where their discernment is needed most. This is the most direct translation of this line of research, but the idea of a verbal cue can apply to other fields where directed attention provides an advantage in a cluttered visual field. For example, UAV operators, radar operators, or Air Traffic Controllers. Designs that build upon basic visual research are incredibly powerful and best utilize the strengths of the robot's human operator.

## Limitations

The most significant limitation of this study is the usage of the Areas of Interest and tethering that visual gaze to the participants' increased perception of that body part. The areas are close together and there could be an issue of participants looking at something that they find visually stimulating on Spot. These areas will be validated with an initial pilot experiment and the researchers are prepared to find any moderating effects of the attention on perception.

Another significant limitation of this study are the verbal cues themselves. While the pilot study will ensure that the cue triggers the participants' gaze to the correct AOI, it is not known how long each cue's effect lasts. Repeating the cue re-orients the participant's gaze, but it is also not known if that effective time diminishes after being repeated. Also, the cues are sentences rather than just location words which causes the participant to process the cue's meaning and associate that meaning with a body part rather than being directed to a specific location with one word.

One potential limitation of the analysis of experimental data from Part 1 is the use of the nonparametric local linear method. As previous work has demonstrated, a correct parametric model will always perform better than a nonparametric model, as the parametric model assumes more about the data (Zychaluk & Foster 2009). However, an ad hoc models can be compared using simpler parametric model fits for cross-validation in order to account for this limitation.

## Future Work

In future studies, we anticipate studying the effective time of long cues (seen in Table 1) against that of short cues (i.e. front legs, head, torso). Cues direct perceptual attention, but it is not known what specific information needs to be in the cue to best orient people. This information would drive the information held within the cue as well as inform how often the cue needs to be reinforced to maintain a perceptual advantage. This future study about the timing and effectiveness of verbal cues could directly inform designers that need to address a communication gap between a robot and a human - what information is communicated and how often does the robot need to orient the operator's visual perception.

**References**


Beck, D. M., & Kastner, S. (2007). Stimulus similarity modulates competitive interactions in human visual cortex. *Journal of Vision*, 7(2), 1–12. https://doi-org.prox.lib.ncsu.edu/10.1167/7.2.19

Carrasco M. (2011). Visual attention: the past 25 years. *Vision research*, 51(13), 1484–1525. https://doi-org.prox.lib.ncsu.edu/10.1016/j.visres.2011.04.012

Carrasco, M., Ling, S. & Read, S. (2004). Attention alters appearance. Nat Neurosci 7, 308–313 . https://doi-org.prox.lib.ncsu.edu/10.1038/nn1194

Cohen D. J., Bennett S. (1997). Why can't most people draw what they see? *Journal of Experimental Psychology: Human Perception and Performance*, 23, 609–621. https://doi.org/10.1037/1931-3896.2.1.8

Gazzaley, A., & Nobre, A. C. (2012). Top-down modulation: bridging selective attention and working memory. *Trends in cognitive sciences*, *16*(2), 129-135.

Lou L. (2018). Artists' innocent eye as extended proximal mode of vision. *Art and Perception*, 6, 1–40. https://doi.org/10.1163/22134913-00002100

Lupyan, G., & Spivey, M. J. (2010). Making the invisible visible: verbal but not visual cues enhance visual detection. *PloS one*, 5(7), e11452. https://doi-org.prox.lib.ncsu.edu/10.1371/journal.pone.0011452

McAdams, C. Maunsell, J. (1999). Effects of Attention on Orientation-Tuning Functions of Single Neurons in Macaque Cortical Area V4. *Journal of Neuroscience*, (1) 431-441, 10.1523/JNEUROSCI.19-01-00431.1999

Mitchell, P., Ropar, D., Ackroyd, K., & Rajendran, G. (2005). How perception impacts on drawings. *Journal of Experimental Psychology: Human Perception and Performance*, 31(5), 996.

Ostrofsky, J., Kozbelt, A., & Cohen, D. J. (2015). Observational drawing biases are predicted by biases in perception: Empirical support of the misperception hypothesis of drawing accuracy with respect to two angle illusions. *Quarterly Journal of Experimental Psychology*, 68(5), 1007-1025.

Pepperell R., Haertel M. (2014). Do artists use linear perspective to depict visual space? *Perception*, 43, 395–416. https://doi.org/10.1068/p7692

Smith, A. T., Singh, K. D., & Greenlee, M. W. (2000). Attentional suppression of activity in the human visual cortex. *NeuroReport: For Rapid Communication of Neuroscience Research*, 11(2), 271–277. https://doi-org.prox.lib.ncsu.edu/10.1097/00001756-200002070-00010

Wickens, C. D. (2008). Multiple Resources and Mental Workload. *Human Factors*, 50(3), 449-455. https://doi-org.prox.lib.ncsu.edu/10.1518/001872008X288394

Zychaluk, K., & Foster, D. H. (2009). Model-free estimation of the psychometric function. *Attention, perception & psychophysics*, 71(6), 1414–1425. https://doi.org/10.3758/APP.71.6.1414