How Much Value Should an Uncertainty Palette Suppress if an Uncertainty Palette Should Suppress Value? Statistical and Perceptual Perspectives

Matthew Kay* School of Information University of Michigan

ABSTRACT

Correll et al. introduced value-suppressing uncertainty palettes (VSUPs): bivariate color palettes that encode the value of a datapoint and its uncertainty, where the datapoint's value is suppressed at higher levels of uncertainty by combining cells in the palette using a tree structure. These tree VSUPs do not suppress uncertainty monotonically (i.e., higher uncertainty for the same value does not guarantee equal or greater suppression), potentially leading to suboptimal decision-making. This paper introduces two alternative VSUP algorithms that (1) use monotonic uncertainty suppression and (2) are grounded in statistical and perceptual perspectives on uncertainty: shrinkage VSUPs and perceptual VSUPs. These techniques allow incorporation of domain and task-relevant knowledge to answer questions like, how much value should we suppress due to uncertainty—or should we suppress value at all?

Index Terms: Human-centered computing—Visualization—Visualization techniques

1 Introduction

Correll et al. [2] introduced the notion of *value-suppressing uncertainty palettes* (VSUPs): bivariate color palettes that encode the value of a datapoint and its uncertainty, where the datapoint's value is *suppressed* at higher levels of uncertainty (Fig 1: *Tree* vsup). Such palettes are motivated by the observation that people often ignore or discount uncertainty when interpreting data: perhaps, then, we can adjust the values shown to people in a color palette to mitigate decision-making errors that result from them ignoring uncertainty. A common example is election polling maps: people tend to be overconfident when presented with probabilities of a candidate winning an election [9], and so—the reasoning goes—perhaps we can reduce this overconfidence by making a candidate's lead *look* less strong when uncertainty is high.

I find this idea very appealing. However, the particular algorithm used in the original definition of VSUPs—what I will dub **tree VSUPs**—suppresses value in a way that is not fully grounded in either a statistical or perceptual perspective on uncertainty. The tree VSUP algorithm splits the uncertainty in the values in a dataset into several levels. As uncertainty increases, value bins are merged together, and the color shown for a value is the midpoint of the merged bin. At the highest level of uncertainty, everything is given the same color, at the center of the scale (Fig 1: *Tree VSUP*).

Perhaps there is a connection between VSUPs and the idea of shrinkage in hierarchical models: for groups with higher uncertainty in a hierarchical model, estimates will be pulled more towards the mean, making them more conservative (or, say, information from more certain groups is borrowed to improve estimates in groups that

are less certain). At first I thought that that is what tree VSUPs would do, but then I realized they do so inconsistently: sometimes they pull values towards the mean, and sometimes they pull values away from the mean. This led me to formulate one desirable property that VSUPs should have that tree VSUPs do not: **monotonic suppression**; that is, the amount of value suppression should be a monotonic function of the uncertainty. This way, any value that has more uncertainty associated with it will always be as or more suppressed than a different datapoint with the same value but less uncertainty. Without this property, I suspect tree VSUPs may lead people to make worse decisions in some cases (Fig 2).

A second issue with tree VSUPs is that what data values get the maximum amount of value suppression is unclear: should it just be the datapoints with the largest uncertainty in the dataset? But then if a new datapoint were added to a dataset with more uncertainty than has been so far observed, it would become the most-suppressed value, and the previously most-suppressed value might be suppressed much less. It is important to have clear, domain- and task-relevant guidance on how to determine how much value to suppress—or if to suppress value at all—due to uncertainty.

I propose two alternative approaches to constructing valuesuppressing uncertainty palettes: **shrinkage VSUPs** and **perceptual VSUPs**. Both approaches use monotonic suppression: given the same observed value, greater uncertainty always yields equal or greater value suppression. In addition, the amount of suppression is based either on a statistical foundation (shrinkage VSUPs, which can be derived either from a regularization perspective or a Bayesian perspective, wherein a prior is used to shrink values towards some meaningful point) or on a perceptual foundation (perceptual VSUPs, which use a linear-in-probit model of human probability perception [10]² to adjust the displayed value so that tail probabilities are more likely to be perceived as they actually are). I show how—at least with a Normal approximation—value suppression in both cases is a simple function of an observation and its variance, making both forms of VSUP straightforward to implement.

More generally, the framework under which I derive both alternative VSUP algorithms admits several interesting extensions, such as applying a fully Bayesian approach to shrinkage VSUP construction using expert priors, or explicitly incorporating knowledge of probability perception in a particular decision-making context to construct a perceptual VSUP aimed at a specific application.

2 SETTING THE STAGE: UNCERTAINTY

To develop a different form of VSUP grounded either in statistical or perceptual considerations about uncertainty, it helps to first establish exactly what we're talking about when we talk about uncertainty. I tend to think about uncertainty from a Bayesian perspective; or put another way, I characterize uncertainty using probability distributions. This gives us a unified way of thinking about uncertainty and the ability to easily ask certain questions, like the probability of

^{*}e-mail: mjskay@umich.edu

¹Uncertainty may mean different things depending on dataset or task: e.g., the variance in a measurement, or perhaps the probability of a value being above or below some reference point. I will define this precisely later.

²Zhang and Maloney [10] actually describe a linear-in-log-odds model, but as I discuss later, linear-in-log-odds and linear-in-probit are virtually identical, but linear-in-probit makes the math work out better in this case.

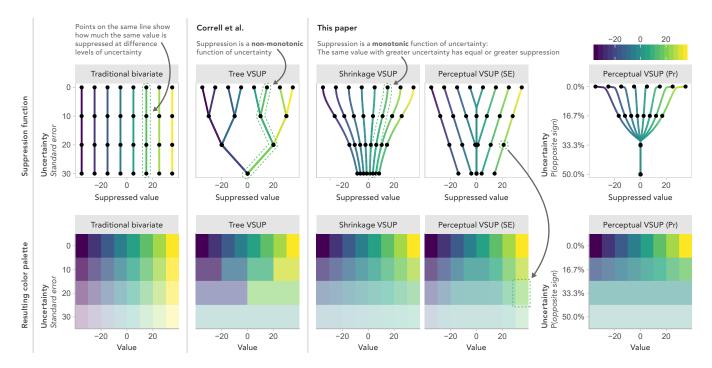


Figure 1: Comparison of traditional bivariate uncertainty color palettes with different forms of value-suppressing uncertainty palettes (VSUPS): VSUPS as originally proposed in Correll et al. [2] (what I call **tree vSUPS**), and **shrinkage vSUPS** and **perceptual vSUPS**, proposed in this paper. These examples all show suppression towards 0, but any reference value could be chosen. The amount of suppression in the palettes proposed in this paper could be more or less than shown here, depending on domain knowledge incorporated into a shrinkage prior (for shrinkage VSUPS) or bias correction based on a linear-in-probit model of probability perception for a given domain (for perceptual VSUPS).

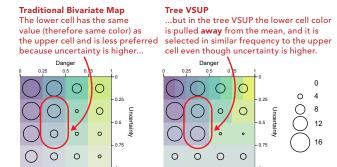


Figure 2: Non-monotonic suppression may lead to less optimal decisions: e.g., preference for an option with similar value to but higher uncertainty than an alternative (from Correll [2], annotation added).

some particular outcome or some particular state in the world.³

From that perspective, say we have observed some value: perhaps the difference in polling support between two candidates in an election (call this the *lead*, where a positive lead means the candidates is polling higher than their opponent, and a negative lead means the opponent is polling higher). Such a dataset was used as an example in Correll et al. [2]. Adopting a Bayesian perspective, we might characterize the *i*th observed lead using a random variable X_i :

$$X_i \sim \text{Normal}(\mu_i, \sigma_i^2)$$

This could be thought of as a measurement error model: we do

not observe a single point μ_i , we observe it with error σ_i . For the sake of both shrinkage and perceptual VSUPs, I am going to assume everything is Normally distributed. This will make it possible to straightforwardly generate examples of both type of VSUP; however, in principle one can apply these same techniques to other distributions using analytical or numerical methods (e.g., Monte Carlo).

Given our two assumptions—uncertainty is probability, and everything is Normally distributed—we must now consider how to map the uncertainty of each datapoint onto a number that we can use in a bivariate color palette. The *value* of the *i*th datapoint is μ_i ; what is its *uncertainty*? Let's consider two different approaches:

- 1. **Precision of an estimate:** How precise or not is our estimate of some quantity? We might quantify this precision (or really, lack of precision) using the standard deviation of our uncertainty in X_i : that is, σ_i . Traditionally we might call this the *standard error*.
- 2. **Probability of an outcome:** What is the probability of some particular outcome of interest? Taking the election polling example, what is the probability that the lead is positive: i.e., that the candidate wins the popular vote? This is a one-sided probability, such as $\mathbb{P}(X_i < 0)$ (candidate loses) or $\mathbb{P}(X_i > 0)$ (candidate wins).

Each approach will give us a different way to systematically vary "how much uncertainty there is" separately from the actual value (μ_i) of some datapoint. The first leads us to shrinkage VSUPs, and the second to perceptual VSUPs (though as we'll see, due to some nice simplifications, both forms of uncertainty are easy to use with perceptual VSUPs).

3 SHRINKAGE VSUPS

The first sense of uncertainty—as precision of an estimate—leads us to a regularization or shrinkage VSUP, wherein we adjust a data-

³For the purposes of this paper, I will assume we are only interested in uncertainty that can be quantified statistically (leaving aside, for example, qualitative *ontological uncertainty* [7] in a model or data generating process).

point's value statistically according to its standard error.

We will proceed by setting a goal of deriving a modified version of X_i that we will call X_i^* . This will also be a random variable (i.e., a probability distribution), but will represent a statistically adjusted version of X_i that has had its value *suppressed* according to how imprecise it is (where more imprecise means more suppressed).

One approach that yields a monotonic suppression function is to translate X_i to X_i^* using a Bayesian measurement error model. We will assume that all of the X_i 's come from a Normal distribution, say Normal (θ, τ^2) . We can think about this as a prior for what we think X_i is before we observe it. Using an Empirical Bayesian approach, we will estimate the prior mean (θ) and standard deviation (τ) from the data: we'll take θ as the mean of all μ_i and τ as the standard deviation of all μ_i . This allows us to derive the posterior distributions (X_i^*) analytically [3]:

$$X_i^* \sim \text{Normal}(\mu_i^*, \overset{*}{\sigma_i^2})$$

$$\mu_i^* = \frac{\frac{\theta}{\tau^2} + \frac{\mu_i}{\sigma_i^2}}{\frac{1}{\tau^2} + \frac{1}{\sigma_i^2}}$$

$$\frac{1}{\overset{*}{\sigma_i^2}} = \frac{1}{\tau^2} + \frac{1}{\sigma_i^2}$$

The posterior mean (μ_i^*) is the observed value (μ_i) *shrunken* towards the prior mean (θ) ; our suppression function for a **shrinkage VSUP** is to replace each observed value (μ_i) with the mean of its posterior (μ_i^*) in the color palette. Another way to view this is as a weighted average of the prior mean (θ) and the observed value (μ_i) , weighted by the prior precision $(1/\tau^2)$ and the precision of the observation $(1/\sigma_i^2)$, respectively. This causes values to be smoothly suppressed, as in the *Shrinkage VSUP* panel in Fig 1. The amount of suppression is a product of both how confident we are in the prior and how confident we are in any given observation.

We can even tweak this suppression function to be more appropriate by incorporating application-specific domain knowledge into the prior. In the polling example, we might fix the prior mean (θ) for the *lead* to be 0, to ensure for values to be suppressed equally for candidates in either party. With θ fixed to 0, we can simplify the suppression function even further:

$$\mu_i^* = \frac{\mu_i}{1 + \frac{\sigma_i^2}{\sigma_i^2}} \qquad \text{When } \theta = 0$$

This makes for a pleasantly straightforward suppression function.

4 PERCEPTUAL VSUPS

The second sense of uncertainty—as probability of an outcome—leads us to a linear-in-probit perceptual VSUP, wherin we adjust a datapoint's value *perceptually* according to the probability of it being above or below a reference value of interest.

Assume our viewer is interested in either $\mathbb{P}(X_i \leq x)$ or $\mathbb{P}(X_i > x)$ for some value of x. Our bias-correction strategy will be to present the viewer with some other distribution whose tail probabilities are *perceived* as if they came from X_i , given what we know about biases in people's perception of probability. We will start with a linear-in-log-odds model of probability perception (Zhang and Maloney [10]), which is a finding that biases in probability perception tend to follow a function like this:

$$\mathsf{llo}(p) = \mathsf{logit}^{-1} \left(\alpha + \beta \cdot \mathsf{logit}(p) \right)$$

⁴We could use a fully Bayesian approach, estimating parameters and hyperparameters simultaneously using Markov chain Monte Carlo (MCMC), but for the purposes of this example the Empirical Bayesian has the nice property of giving us a closed-form way of writing the value suppression function, which makes it easy to create the color palette.

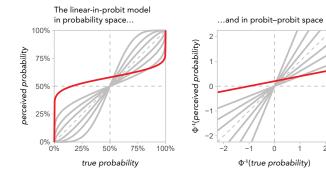


Figure 3: Several linear-in-probit curves with latent intercept (α) = 0 and various slopes (β), and **one curve with nonzero** α that is based on eyeballing polling perception results from Westwood et al. [9]. This model behaves nearly identically to the linear-in-log-odds model of people's perception of probabilities [10] but is more convenient for deriving an uncertainty suppression function. The slope and intercept typically must be empirically derived for a given domain or task.

That is, given a displayed probability p, people tend to perceive it as llo(p): a linear transformation of that probability in log odds space. This relationship shows up across many fields: it takes a similar form as prospect theory in behavioral economics (Tversky and Fox [8]), and generalizes the cyclical power model [10], a variant of Stevens' power law as applied to the perception of proportions in visualizations (Hollands and Dyre [5]).

For our purposes we will actually use a slightly different model: a linear-in-probit model. The logit function and the probit function are both S-shaped functions that are empirically very hard to distinguish from each other (Amemiya [1]). Following in a long statistical tradition, I will adopt whichever happens to be more mathematically convenient for the problem at hand.⁵ As we will see, because I have assumed all errors are Normal, and the probit function is the inverse Normal cumulative distribution function (CDF), the probit function will allow us to simplify things nicely.

We will define the linear-in-probit function, lpr(p), analogously to the linear-in-log-odds model. Here, Φ is the standard Normal CDF, and Φ^{-1} is its inverse, also known as the probit function:

$$\operatorname{lpr}(p) = \Phi\left(\alpha + \beta \cdot \Phi^{-1}(p)\right)$$
$$\Longrightarrow \operatorname{lpr}^{-1}(p) = \Phi\left(\frac{\Phi^{-1}(p) - \alpha}{\beta}\right)$$

This gives us a latent intercept (α) and slope (β) that control how a person perceives a given probability. These parameters are domain-specific and must be derived empirically [10]. The function lpr(p) translates from displayed probability onto perceived probability, and $lpr^{-1}(p)$ translates from perceived probability back onto the probability we should display. We will also need the following definition of the CDF for a scaled-and-shifted (i.e. non-standard) Normal distribution:

$$\mathbb{P}(X \le x) = \mathcal{F}_{Normal}\left(x \middle| \mu, \sigma^2\right) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

We can use these functions to transform the CDF for a given X_i into a different CDF (and therefore a different probability distribution) that we should display so that the viewer perceives left-tail probabilities correctly. I will dub this new distribution $X_i^{<}$:

⁵If one wishes to translate, $\alpha_{\rm llo} \approx 1.6\alpha_{\rm lpr}$ [1] and β remains the same.

probability we display probability that is perceived
$$\mathbb{P}(X_i^< \leq x) = \operatorname{lpr}^{-1}\left(\mathbb{P}(X_i \leq x)\right)$$

$$= \operatorname{lpr}^{-1}\left(\Phi\left(\frac{x - \mu_i}{\sigma_i}\right)\right)$$

$$= \Phi\left(\frac{x - \mu_i - \alpha\sigma_i}{\beta\sigma_i}\right)$$

$$= F_{\text{Normal}}\left(x\Big|\mu_i + \alpha\sigma_i, (\beta\sigma_i)^2\right)$$

$$\Longrightarrow X_i^< \sim \operatorname{Normal}\left(\mu_i^<, \tilde{\sigma}_i^2\right)$$
where $\mu_i^< = \mu_i + \alpha\sigma_i$
and $\sigma_i^< = \beta\sigma_i$

We can similarly derive a distribution to help viewers correctly perceive right-tailed probabilities, $X_i^>$ (details omitted for space):

probability we display probability that is perceived
$$\mathbb{P}(X_i^>>x) = \operatorname{lpr}^{-1}(\mathbb{P}(X>x))$$

$$\Longrightarrow X_i^> \sim \operatorname{Normal}\left(\mu_i^>, \overset{>}{\sim}_i^2\right)$$
where $\mu_i^> = \mu_i - \alpha\sigma_i$
and $\sigma_i^> = \beta\sigma_i$

I found this result pleasingly tidy: human biases in probability perception imply that a simple bias correction approach when faced with Normal uncertainty is to shift the mean by a multiple of the standard deviation. One nice upshot of this is that perceptual suppression can be used to easily create color palettes that vary in terms of either form of uncertainty we have considered: **precision of an estimate** (e.g., standard error) or **probability of an outcome**. A minor complication arises in that the correction has the opposite sign (either adding or subtracting $\alpha \sigma_i$ to μ_i) depending on if the viewer is interested in left- or right-tailed probabilities. I think domain knowledge can help here.

Take again the election polling example from Correll et al. [2]. Literature suggests that people have a slight positive bias (positive α) in the perception of chances of candidates winning, consistent with folk knowledge that people interpret probabilistic election forecasts overconfidently. When considering polling differences in two-party elections, we also have a natural reference point, r = 0 (even support). Given such a reference point, we could use $\mu_i^{<}$ to suppress μ_i when above r and $\mu_i^{>}$ to suppress μ_i when below r. This corresponds to a perceptual bias correction for the probability that the opposite candidate wins the election. We should also constrain the suppressed values not to cross the reference point to avoid reversals of sign (this also allows for merging of values, akin to the merging in tree VSUPs). Therefore, we'll define $\mu_i^{>}$, a perceptually-suppressed value with respect to some reference point r, as follows:

$$\mu_i^{\lozenge} = \begin{cases} \max(\mu_i^{<}, r) & \text{if } \mu_i \leq r \\ \min(\mu_i^{>}, r) & \text{if } \mu_i > r \end{cases}$$

⁶To fully transform the distribution one could simultaneously scale the standard deviation by β as well; for a color palette this makes no difference, since all σ_i 's in the palette would be scaled by the same constant *then* mapped into a color space the result would be the same as not scaling the σ_i 's at all.

⁷Based on me eyeballing a linear-in-probit function from figures in Westwood et al. [9].

⁸See, for example, overconfidence in probabilistic predictions in the 2016 US presidential election [4], and particularly negative reactions to FiveThirtyEight's predictions—despite FiveThirtyEight being historically well-calibrated [6] and giving Trump a reasonable ~ 30% chance at winning.

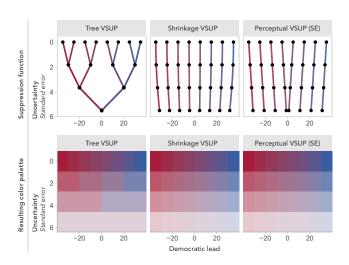


Figure 4: Different VSUPs applied to the polling example from Correll et al. [2], though with uncertainty as precision (SE) instead of lead/error (based on advice [9] not to emphasize election win probabilities). Tree VSUPs arguably over-suppress uncertainty in this data.

This suppresses values towards the reference point (r) as a linear function of the error (σ_i) , as can be seen in the *Perceptual VSUP* (SE) panel in Fig 1. An alternative approach might be to display the VSUP along a probability axis for the uncertainty rather than an error axis, as in the *Perceptual VSUP* (Pr) panel, though this yields a sharp (and perhaps undesirable) change in the amount of suppression.

5 So, How Much Value Should We Suppress?

Thinking statistically and perceptually about uncertainty can help guide designers in determining how much (or even if) value should be suppressed in an uncertainty palette. The answer to this question inevitably comes back to tasks and domain knowledge: do users care about precision or tail probabilities? Is there domain knowledge that could form a reasonable prior? Is there a sensible reference point around which to derive a perceptual VSUP? Do we know enough about biases in probability perception for the domain at hand to confidently derive α ?

Absent clear answers to these questions, a conservative approach might be to use a traditional bivariate color map, or perhaps a shrinkage VSUP with an empirical prior (or a fully Bayesian shrinkage VSUP based on a hierarchical model).

While how much value to suppress may not always be obvious, shrinkage VSUPs and perceptual VSUPs can help us probe that question more deeply. For example, when applied to the election map data from [2], the amount of value suppression from both techniques is markedly smaller than the suppression applied by a tree VSUP (Fig 4). I would argue that this suggests that there is little statistical or perceptual justification for strong suppression in this dataset, unless one had a very strong prior that polling margins should cluster tightly around 0—in which case a shrinkage VSUP could be employed.

6 CONCLUSION

Statistical and perceptual perspectives offer an approach to bias correction in uncertainty visualization that allows for monotonic suppression and which can help designers incorporate domain and task knowledge into uncertainty suppression techniques. The perceptual uncertainty correction result could also be incorporated into bias correction in uncertainty visualization outside of color palettes.

REFERENCES

- [1] T. Amemiya. Qualitative response models: A survey. *Journal of Economic Literature*, 19(4):1483–1536, 1981.
- [2] M. Correll, D. Moritz, and J. Heer. doi: 10.1145/3173574.3174216
- [3] A. Gelman, H. S. Stern, J. B. Carlin, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian data analysis*. Chapman and Hall/CRC, 2013.
- [4] J. H. Gross. How to better communicate election forecasts in one simple chart, nov 2016.
- [5] J. G. Hollands and B. P. Dyre. Bias in Proportion Judgments: The Cyclical Power Model. *Psychological Review*, 107(3):500–524, 2000. doi: 10.1037//0033-295X.107.3.500
- [6] N. Silver. When We Say 70 Percent, It Really Means 70 Percent, apr 2019.
- [7] D. Spiegelhalter. Risk and Uncertainty Communication. *Annual Review of Statistics and Its Application*, 4(1):31–60, mar 2017. doi: 10.1146/annurev-statistics-010814-020148
- [8] A. Tversky and C. R. Fox. Weighing risk and uncertainty. *Psychological review*, 102(2):269, 1995.
- [9] S. Westwood, S. Messing, and Y. Lelkes. Projecting confidence: How the probabilistic horse race confuses and demobilizes the public. *Available at SSRN 3117054*, 2019.
- [10] H. Zhang and L. T. Maloney. Ubiquitous log odds: A common representation of probability and frequency distortion in perception, action, and cognition. *Frontiers in Neuroscience*, 6(JAN):1–14, 2012. doi: 10. 3389/fnins.2012.00001