DeepColor: Reinforcement Learning optimizes information efficiency and well-formedness in color name partitioning

Downloaded from: https://research.chalmers.se, 2019-01-21 04:05 UTC

Citation for the original published paper (version of record):
DeepColor: Reinforcement Learning optimizes information efficiency and well-formedness in color name partitioning

N.B. When citing this work, cite the original published paper.
DeepColor: Reinforcement Learning optimizes information efficiency and well-formedness in color name partitioning

Mikael Kågebäck, Devdatt Dubhashi (kageback, dubhashi@chalmers.se)
Dept. of Computer Science and Engineering
Chalmers University of Technology, Sweden

Asad Sayeed (asad.sayeed@gu.se)
Dept. of Philosophy, Linguistics, and Theory of Science
University of Gothenburg, Sweden

Abstract
As observed in the World Color Survey (WCS), some universal properties can be identified in color naming schemes over a large number of languages. For example, Regier, Kay, and Khetrapal (2007) and Regier, Kemp, and Kay (2015); Gibson et al. (2017) recently explained these universal patterns in terms of near optimal color partitions and information theoretic measures of efficiency of communication. Here, we introduce a computational learning framework with multi-agent systems trained by reinforcement learning to investigate these universal properties. We compare the results with Regier et al. (2007, 2015) and show that our model achieves excellent quantitative agreement. This work introduces a multi-agent reinforcement learning framework as a powerful and versatile tool to investigate such semantic universals in many domains and contribute significantly to central questions in cognitive science.

Keywords: color naming; world color survey; reinforcement learning

Introduction
Semantic partitioning of the color space in the lexicon
Color naming universals have a long history in linguistic research (Berlin & Kay, 1969). At an individual level, color perception is subjective; it differs for biological reasons across individuals (extreme examples being colorblindness and tetrachromacy). There are commonly-observed differences in individual color-naming choices. What is “turquoise” to one person may be a variant of “blue” to another. Nevertheless, within the same linguistic milieu, there is overall agreement as to color-naming; most English-speaking people recognise the typical supermarket tomato as “red”.

Berlin and Kay showed across a survey of 20 languages that there are strong consistencies in color naming and produced a set of universals: e.g., there are a maximum of eleven major color categories and, where fewer than eleven are realized for a given language, there is a standard pattern of emergence. This work came under methodological criticism (Lucy, 1997; Saunders, 1995), particularly the use of standardized color systems to abstract away from the interactional and cultural basis of color identification.

Given this methodological conflict, is it really the case that such universals are artifacts of a non-interactional method of investigation? Accounting for patterns of wide but constrained variation that have been observed empirically is a central challenge in understanding why languages have the particular forms they do. There is an increasingly influential proposal that language is shaped by the need for efficient communication (Piantadosi, Tily, & Gibson, 2011), which by its nature involves a trade-off between simplicity in an information-theoretic sense, which minimizes cognitive load, and informativeness which maximizes communicative effectiveness. Specifically, that good systems of categories have a near-optimal trade-off between these constraints.

Examples formalized in information-theoretic terms include suggestions that word frequency distributions, syllable durations, word lengths, syntactic structures, and case marking all facilitate efficient communication. This type of proposal has been particularly successful in explaining aspects of the structure of utterances as they unfold over time: linguistic structure optimizes information transfer between interacting language users. A different challenge is explaining the partitioning of semantic spaces—the common recognition of lexical-semantics differences across a speech community—likewise in terms of interactive behavior.

Color terms represent a limited semantic domain with easily manipulated parameters. By gradual changes of color value, an experimenter can manipulate red into orange, unlike other semantic domains, where the distinctions between potential referents (e.g., “car” vs. “truck”) are not easily captured in explicit terms. In addition, recent work (Regier et al., 2015; Gibson et al., 2017) argues that color categories in language should support efficient communication.

Our goal in this paper is to investigate the scientific question of the emergence of color terms via a computational framework for modeling the partitioning of semantic spaces. Our framework, DeepColor, is based on a deep reinforcement learning approach with independent justification as a cognitive modeling paradigm. The framework, inspired by Regier et al. (2015) and Gibson et al. (2017), takes the form of two agents, one of which is attempting to communicate a color to the other through a channel with a limited selection of signals intended to refer to the colors in a commonly-used color-identification experimentation system. Learning takes place through a reward system as the agents converge on approximate color-meanings for the signal vocabulary. We show that the simulations we run in this noisy-channel environment converge on characteristics similar to cross-language human-collected data using the same color system. This work val-
and control theory. RL, originally born out of mathematical
range of fields, including ethology, economics, psychology,
ing & van Otterlo, 2012). Such environments arise in a wide
range of environments in which actions lead
to rewards and punishments (Sutton & Barto, 1998; Wier-
them from one state or situation to the next, and can also lead
optimize their behavior in environments in which actions lead
Reinforcement learning (RL) studies the way that natural and
artificial systems can learn to predict the consequences of and
optimize their behavior in environments in which actions lead
from one state or situation to the next, and can also lead
to rewards and punishments (Sutton & Barto, 1998; Wier-
ing & van Otterlo, 2012). Such environments arise in a wide
range of fields, including ethology, economics, psychology,
and control theory. RL, originally born out of mathematical
psychology and operations research, provides qualitative and
quantitative computational-level models of these solutions.

There is an increasing realization that RL may offer more
than just a computational, theory for affective decision-
making but that RL algorithms appear to be directly instan-
tiated in neural mechanisms, such as the phasic activity of
dopamine neurons (Dayan & Niv, 2008; Niv, 2009; Niv &
Langdon, 2016). That RL appears to be so transparently
embedded implies that it can be seen as a general cognitive
mechanism and used in an immediate way to make hypothe-
ses about and interpretations of a wealth of behavioral and
neral data.

The availability of a growing suite of environments (from
simulated robots to Atari games), toolkits, and sites for com-
paring and reproducing results about RL algorithms applied
to a variety of tasks (Lazaridou, Pysakhovich, & Baroni,
2016; Havrylov & Titov, 2017; Evtimova, Drozdov, Kiela,
& Cho, 2017; Jorge, Kågebäck, & Gustavsson, 2016) makes
it possible to study cognitive science questions through a dif-
ferent lens. Cognitive science experiments are often carried
out in real life settings involving questionnaires and surveys
which are both costly and suffer from variability in responses.
If RL algorithms are indeed a good proxy for actual human
learning, then insights about questions of universals in lan-
guage learning could be obtained very cheaply and reliably
via controlled experiments in such in silico settings. We
 demonstrate this with a focus on questions about the univer-
sality of color categories and words in language.

This paper contributes a model of color partitioning based
in recent advances in deep reinforcement learning that di-
rectly simulates an interactive, communication-based model
of arriving at a color term consensus that partitions the color
space for optimal, lowest-cost communication. We thus ad-
vance work in measuring the role of communicative cost by
providing a procedure to generate color naming schemes that
reflect those results. Our system is a starting point in devel-
opling models to test further hypotheses about universals in
semantic partitioning while preserving a notion of interaction
in the generation of meaning.

Evaluating color–word schemes

An assignment of a color term to each chip corresponds to
a categorical partition of color space this arrangement repres-
ts. Given such a partition \( P \), Regier et al. (2007, 2015)
propose quantitative schemes to test the hypothesis that at-
tested color naming systems achieve a near-optimal trade–off
between informativeness and complexity—that is, that they
are nearly as informative as is theoretically possible for their
level of complexity (i.e., for their number of color terms).

Partitions of color space

The first criterion, proposed in Regier et al. (2007), is a mea-
sure of color space partition quality. The objective function
measures the extent to which such an assignment of cate-
gory labels to chips maximizes similarity within categories
and minimizes similarity across categories. In fact, this measure is well known in theoretical computer science and approximation algorithms as the (weighted) correlation clustering problem (Bansal, Blum, & Chawla, 2004; Giotis & Gupta, 2006; Ailon, Charikar, & Newman, 2008). Given a graph \( G = (V, E) \) with weights \( w^+, w^- : E \rightarrow R \); the maximizing agreements version of the problem seeks to partition the vertices into disjoint sets so as to maximize

\[
\sum_{\text{cat}(i) = \text{cat}(j)} w^+(i, j) + \sum_{\text{cat}(i) \neq \text{cat}(j)} w^-(i, j).
\]

Here \( \text{cat}(i) \) refers to the subset in the partition to which the vertex \( i \) belongs.

In our case, as in Regier et al. (2007), we take \( w^+(i, j) := \text{sim}(i, j) \) and \( w^-(i, j) := (1 - \text{sim}(i, j)) \), where \( \text{sim}(i, j) \), the similarity of two colors \( i \) and \( j \), is adopted from the psychological literature on categorization:

\[
\text{sim}(i, j) := \exp \left( -c \text{dist}(i, j)^2 \right), \quad (2)
\]

where \( \text{dist}(x, y) \) is the CIELAB distance between colors \( x \) and \( y \), and \( c \) is a scaling factor (set to 0.001 for all simulations reported here). It has a maximum value of 1 when chips \( x \) and \( y \) are the same (i.e., \( \text{dist}(x, y) = 0 \)) and a value that falls off approaching 0 as the distance between chips \( x \) and \( y \) becomes arbitrarily large. This similarity function thus captures the qualitative observation that beyond a certain distance colors appear "completely different" so that increasing the distance has no further effect on dissimilarity.

Given this choice, the problem is to maximize over all partitions \( \mathcal{P} \),

\[
\text{CC} (\mathcal{P}) := \sum_{\text{cat}(i) \neq \text{cat}(j)} 1 - 2 \text{sim}(i, j)
\]

The problem is NP-hard but can be computed exactly for small sizes with an integer LP. Regier et al. (2007) propose a heuristic to approximate the optimum.

**Information-theoretic analysis: communicating over a noisy channel**

The second criterion proposed in Regier et al. (2015) takes an information-theoretic perspective via a scheme in the form of a speaker and a listener communicating over a noisy channel. The speaker attempts to communicate a color from the WCS color grid. Each agent maintains a belief in the form of a probability distribution over colors. The speaker’s distribution \( s \) is concentrated on the single color \( i \) she is trying to convey by uttering a word \( w \). After listening to the uttered word, the listener tries to reconstruct the color, resulting in a probability distribution \( \ell \) over colors. The reconstruction error \( e(i) \) for color \( i \) is the KL divergence \( D(s || \ell) \) which equals the surprisal \( -\log \ell(i) \). The listener distribution \( \ell \) is concentrated on the category labeled with the word communicated and

\[
\ell(t) = s(t) / \sum_{i \in \text{Cat}(t)} s(i)
\]

Hence, \( E := \sum_i n(i) e(i) \).

**Communication efficiency**

In the work of (Gibson et al., 2017) the related measure communication efficiency is defined as

\[
\sum_c p(c) \sum_w p(w|c) \log_2 \frac{1}{p(c|w)}
\]

and computes the expected average surprisal of each color chip \( c \) while assuming a uniform prior \( p(c) \).

**Combined criterion**

An alternative suggestion is that the listener selects uniformly from the category labeled by the received word. In this case, the surprisal is \( \log_2 |\text{Cat}(w)| \). This suggests we optimize a combination of the two criteria, the quality of the color space and the surprisal, that is find a partition \( \mathcal{P} \) to minimize

\[
\frac{1}{n} \sum_w |\text{Cat}(w)| \log_2 |\text{Cat}(w)| + \text{CC}(\mathcal{P})
\]

\[
= \log n - H(\mathcal{P}) + \text{CC}(\mathcal{P}).
\]

Figure 1: Communication cost (left) and communication efficiency (right) for different levels of noise as a function of the number of allowed color terms. The points indicate the mean values from 50 independent runs and the shaded area constitutes the std. deviation \( \sigma / 4 \).
constitutes the std. deviation \( \sigma \) of the std. deviation \( \sigma / 4 \).

where \( \mathcal{H}(\mathcal{P}) \) is the entropy of the partition \( \mathcal{P} \) (with uniform measure on the elements).

A DeepColor framework for communication

We develop a version of the communication setup from Regier et al. (2015) via two automated agents implemented as feed forward neural networks trained via reinforcement learning. Though the structure of the model is more general, for the purposes of our game, the stimulus always consists of a color coordinate in CIELAB color space and the guess indicates which of the color chips in the palette the receiving agent believes the sender is trying to communicate. Both the sender and receiver is modeled using a multilayer perceptron with one hidden layer of \( k = 20 \) units.

The sender decides which message to send by sampling from its internal distribution over the set of color terms \( W \)

\[
w \sim p(W|t; \Omega_s) = \text{softmax}(g^T \tanh \theta^T t)
\]

given the stimulus \( t = \text{CIELAB}(c) + \beta \), where \( c \sim \mathcal{U} \) and the noise vector \( \beta \sim \mathcal{N}(0, \sigma^2) \). \( \Omega_s = \{ \theta_s \in \mathbb{R}^{k \times 3}, \phi_s \in \mathbb{R}^{|W| \times k} \} \) is the parameterization of the sender agent, and the softmax is defined as \( \text{softmax}(\mathbf{z}) = e^{\mathbf{z}} / \sum e^{\mathbf{z}} \). The bias terms have been omitted for brevity.

The receiver interprets the message and computes a distribution over all color tiles in \( U \) given the received message \( w \) as \( p(U|w; \Omega_r) = \text{softmax}(\phi^T \tanh \theta^T w) \), where \( \Omega_r = \{ \theta_r \in \mathbb{R}^{k \times |U|}, \phi_r \in \mathbb{R}^{|U| \times k} \} \) parameterize the receiver.

The sender part of the model is trained using the well known policy gradient method REINFORCE (Williams, 1992), an algorithm that aims to update the parameters of the model such that to maximize a given reward function. In our case this reward is chosen as the communication outcome \( r := \text{sim}(c, \text{argmax} p(U|w; \Omega_s)) \), where \( \text{sim} \) is the color similarity function defined in Equation 2. Plugging this reward into the REINFORCE cost function gives us

\[
J_s(\Omega_s) = -\frac{1}{N_b} \sum_n \log p(W = w_n|t_n; \Omega_s) * r_n,
\]

where \( N_b \) corresponds to the number of games that the cost is computed over. The receiver objective can be modeled directly, i.e., without using a reward, as the surprisal incurred by the receiver after seeing the intended color chip \( c \):

\[
J_r(\Omega_r) = -\frac{1}{N_b} \sum_n \log p(U = c_n|w_n; \Omega_r).
\]

The final objective function \( J(\Omega_s, \Omega_r) = J_s(\Omega_s) + \lambda J_r(\Omega_r) \) combines the sender and receiver parts using the mixing factor \( \lambda = 100 \), and the model is trained by minimizing this objective. At the beginning of training all parameters \( \{ \Omega_s, \Omega_r \} \) are initialized to random values and the optimization min\( \Omega_s, \Omega_r J(\Omega_s, \Omega_r) \) is performed using stochastic gradient descent with ADAM (Kingma & Ba, 2014). The batch size is set to \( N_b = 100 \) games, and the model is trained for a total of \( N = 10000 \) games.

After the model has been trained, the resulting color partitions \( \mathcal{P} \) is extracted from the sender agent as its most probable message corresponding to each color tile \( c \in U \).

Experiments

We conducted experiments to answer three primary questions: (1) Does DeepColor converge on high quality color partitions measured by both well-formedness of the resulting partitions and the information efficiency of the communication scheme; (2) how does varying noise during training influence the results in (1); and (3) does noise influence the number of words that the agents choose to employ?

To answer these questions we trained independent agents on all combinations of the number of allowed color terms \( k \)
Figure 4: Resulting color maps when allowing the agents to use 5 color terms and adding different amounts of noise \((\sigma^2 \in \{0, 25, 50, 100\})\) to the color chips. The color maps are presented in order of increasing noise from top to bottom.

\(\{3, 4, \ldots, 11\}\) and the variance of the noise added to the color chips \(\sigma^2 \in \{0, 25, 50, 100\}\).

In figures 1-3, we represent the performance of our models relative to the communication efficiency (see equation 4) communication cost (see equation 3), wellformedness (see equation 1), and combined criteria (see equation 5). The number of color words has minimal effect on efficiency, particularly at higher noise. As Gibson et al. (2017) suggest, this indicates the robustness of the WCS investigative paradigm: in our case, in an artificial agent setting. Our process is highly tolerant to noise, with the highest-noise scenario having similar overall cost reductions and well-formedness improvements relative to the number of words. More words represent a reduction in costs, with the curve reaching its asymptote at the higher number of words.

Figures 4 and 5 display sample color maps generated by the reinforcement learning process. A bounded region on each map represents a word that the system assigned to each color; they are not necessarily contiguous and sometimes wrap around the table. At 5 color terms, there is considerable stability among regions regardless of noise. At 11 color terms (this is a maximum, as the system is not obliged to assign all colors), we see more hierarchy and variation in the maps and more terms exploited as the noise increases. We also observe large similarities with the 5-color maps in terms of overall boundaries, with greater subdivisions.

Figure 6 presents a WCS color mode map of the Iduna language. Compared to the five-color maps of figure 4, we see large overlap in the color boundaries that appear. One major difference is that Iduna uses one color term for a small number of Munsell chips, in the manner that the noisiest two maps of figure 5 create singleton or small–group color terms.

Figure 7 shows that noise does have an effect on the number of words that the agents decided to employ—large amounts of noise decrease the active vocabulary of the agents, so precise terms may be less useful when referring to objects that vary in color. However, we also observe that no noise can have a similar detrimental effect which may be due to the noise acting as a stabilizing regularizer during training.

The presented example color maps have been sampled randomly from our experimental results that were compiled after training DeepColor using the hyper-parameters specified in the previous section.
Figure 7: Average number of words actually used by the agents after training under different amounts of noise. The points indicate the mean values from 20 independent runs and the shaded area constitutes the std. deviation $\sigma/4$.

Conclusions and future work

From where do apparent regularities in color naming patterns across language arise? A recent strain of research explains this via a perceptual well-formedness criterion as well as communicative cost. Implicit in communicative cost is the role of linguistic interaction. Our deep reinforcement model directly implements an interactional learning model, and the color term maps generated by this model replicate empirical observations in human–language color mapping.

Compared to other semantic domains, color is a simple test bed for interaction-based approaches to modeling the partitioning of meaning spaces. The initial success of this model not only allows for further experiments and comparisons in the quality and grounding of color partitions, such as through image data, but application to other semantic domains, such as the partitioning of WordNet synsets. It also demonstrates the viability of reinforcement learning in implementing information-theoretic approaches to representing semantic distinctions in linguistic interaction.

Acknowledgments The authors would like to acknowledge the project Towards a knowledge-based culturomics supported by a framework grant from the Swedish Research Council (2012–2016; dnr 2012-5738) as well as the SRC-funded Centre for Linguistic Theory and Studies in Probability (CLASP) at the University of Gothenburg.

References


