Title

Natural selection subsumes and unites multiple theories of perceptual compression

Author

Victor Quintanar-Zilinskas
Address correspondence to: vquintanar@hotmail.com

Upload Date

December 27, 2018

Abstract

An ideal model of perceptual processing would be congruent with experimentally-observed perceptual performance and be theoretically generalizable. In the extant models with experimental predictive power, the most-invoked signal processing paradigms are rate-distortion theory, efficient coding, and Bayesian inference. A theoretical reconciliation of these alternatives has yet to emerge. Here I endeavor to constrain the model space by clarifying which processing paradigms align with natural selection objectives. I find that signal processing and fitness objectives often deviate; for example, in a finite game where risk-avoidant bet-hedging beats maximization of average performance. When these objectives do align, the specific signal processing performance metric that aligns with fitness over the most general set of circumstances is the minimization of utility-weighted signal distortion. Together, these results reveal that selection favors perceptual system adaptations to (at least) three types of noise; environmental, encoding, and decoding; and establish utility-weighted distortion minimization as the optimal adaptation to the second.

Main Text

An ideal model of perceptual processing would be congruent with experimentally-observed perceptual performance and be theoretically generalizable. In models with experimental predictive power, the most-invoked signal processing paradigms are rate-distortion theory (1-3), efficient coding (4-6), and Bayesian inference (7-9). Though some models (2, 3, 6-8) partially reconcile these paradigms by framing perceptual encoding (i.e. environmental information compression) and decoding (i.e. inference) as separate processes, a theoretical characterization of their respective domains of applicability has yet to emerge. Adding further confusion is that all can be related to maximum-value decision-making and have therefore been asserted to be favored by natural selection (2, 3, 6, 7, 10). Here I endeavor to constrain the model space by framing fitness as a function of the encoding scheme and optimizing fitness directly.

To characterize the relationship between perceptual compression and reproductive fitness, I devise a framework for evaluating the fitness consequences of a particular compression configuration: in its most general form, it involves an organism being presented with n items, perceiving them through its perceptual interface, choosing one, and adding the item's utility value to its total lifetime utility score. Not all interfaces are equally fit: their fitness hinges on their effective mapping of high- and low-utility items to different percepts, thereby enabling low-utility choices to be avoided. Also, the interface represents a compression: multiple items are mapped to the same percept, possible real-world items outnumber the possible percepts, and the mapping does not fully preserve the details of an item's real-world attributes (Fig. 1A). Notably, this framework's discretization of choice options renders the compression amenable to formal mathematical parameterization, while its discretization of choice instances and utilities renders the compression's fitness consequences measureable. Despite these abstractions, the framework retains perception's fundamental relationship to fitness: the information it transmits to the organism informs actions (11).

In the present study, I apply the above framework across several decision-making environments, each characterized by their own combination of selection pressures and distribution of item utilities. In each environment, certain interface parameters, cast as alleles that are subject to selection, are shown to be more adaptive than others, and perhaps even optimal in that particular environment. The various congruencies revealed between fitness-favoring parameterizations and well-known compression and decision strategies collectively provide an expanded understanding of perceptual decision-making in the natural world.

In environment one, fitness is linearly proportional to average decision utility, item utilities are a function of a subset of their real-world attribute values, and the values of each utility-relevant attribute are contained within the interval $[0,b_{a,1}]$, where a is the index of a specific attribute. In a version of this environment simplified even further, utility depends linearly on just one attribute (in which case, I write $b_{a,1}$ as b_1), and item probability is uniform over the attribute space. In such a one-dimensional utility landscape, it is adaptive to associate percept boundaries with specific values of the utility-determining attribute, and assign all items between boundaries b_i and b_{i+1} to the same percept. Thus, an interface with β output bits divides the item space into $m = 2^\beta$ percepts P_i (i = 1:m), whose index indicates both the percept's ordinal desirability (relative to other percepts) and its upper boundary b_i. While b_1 is a constant, the other b_i (i=2:m) can be regarded as the alleles subject to selection.

In this environment, there is congruence between the utility maximization objective and the minimization of signal distortion, with the loss function being the compression's utility cost (details in Supplementary Equations, Part 1). These objectives are demonstrably equivalent in the general case of a β-bit interface and n item choices (Supplementary Equations, Part 2). They are also equivalent, when $\beta = 1$ and n = 2, on utility landscapes exhibiting simple deviations from the assumptions of 1) a linear attribute-utility relationship 2) uniformly probable item attribute values (Fig. 1b1 and 1b2; Supplementary Equations, Parts 3 and 4; respectively). Moreover, these results straightforwardly extend to multidimensional utility landscapes. The extension is trivial for an organism with an idealized interface (12) that can map items to percepts with infinite flexibility. If utilities are unitary (e.g. a weighted sum of attribute values), then the interface will in effect rearrange the items onto a unidimensional utility continuum (Fig. 1c, left). Alternatively, if the organism experiences a variety of situations, each corresponding to a different utility-determinative attribute, then each dimension individually can be optimally divided (Fig. 1c, right). Granted, real-world interfaces are often composites whose individual components (e.g., vision and olfaction) are constrained to processing a subset of item attributes (e.g., reflected wavelengths and emitted chemical compounds). Under such circumstances, the utility and distortion loss objectives will dictate equivalent segmentations of individual dimensions and, at least for a two-bit interface making binary decisions between items from a two-dimensional utility space (Fig. 1d; Supplementary Equations, Part 5), an equivalent assignment of interface bits to the available dimensions.

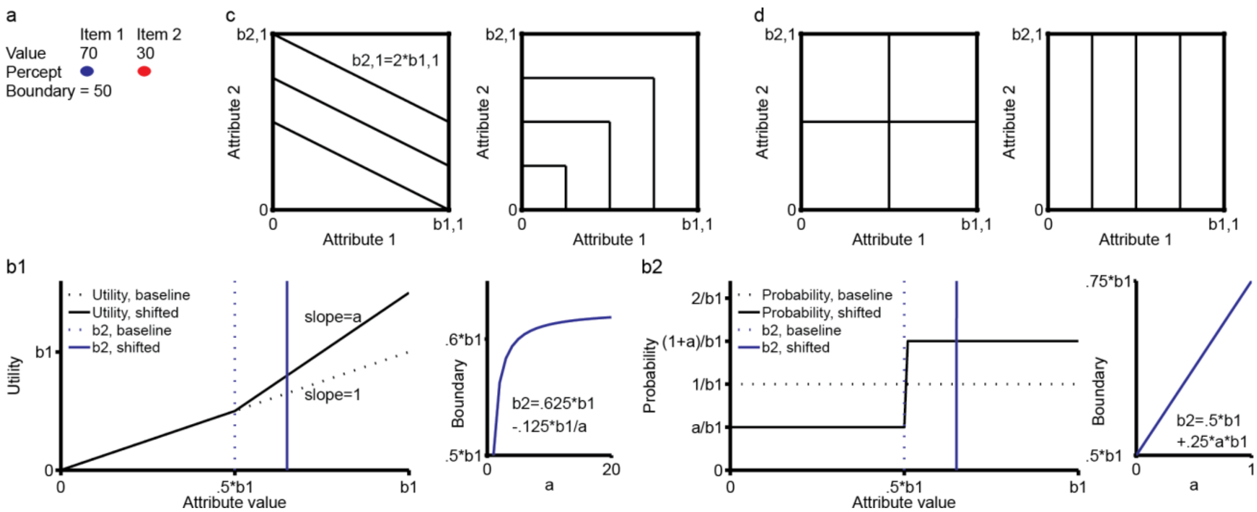

Figure 1. Illustration of utility-maximizing assignment of real-world items to percepts. a, Individual real-world items (1 and 2) have quantifiable attributes. Perceptual interfaces inevitably compress this information; here, an interface with a 1-bit output perceives items as blue or red, based on one particular attribute value being above or below a boundary. The organism favors blue options when choosing an item. **b**, Each pair of subpanels illustrates how (for a 1-bit interface making binary decisions) a particular deviation from a simple baseline utility landscape shifts the optimal perceptual boundary. **c**, The boxes in this panel (and the next) show the optimal division, by a 2-bit interface that makes binary decisions, of an item landscape in which utilities are determined by two uncorrelated attributes. Right and left boxes respectively segment items whose utilities exhibit situation-specific single-attribute dependency or are determined by their attribute sum. These interfaces are free from item-percept assignment constraints. **d**, These landscape divisions are optimal subject to the constraint that attributes are perceived separately. The right-box division is “optimal” (e.g., in environment 2) when utility hinges on attribute 1 far more than on attribute 2. The left-box division is favored evolutionarily when utility is similarly dependent on both attributes, or even in versions of environment 2 (where $b_{2,1} >> b_{1,1}$) in which this “risk averse” division benefits from avoidance of rare, but large, utility-harvesting failures.

Notably, when I set $\beta = 1$ and $n > 2$, the utility and distortion loss objectives diverge from the frequently-studied objective of item utility estimation error minimization (5), which is achieved neither across all perceived items nor across all chosen items; stated alternatively, neither perceptual accuracy nor action outcome prediction are optimized (Supplementary Equations, Part 6). However, optimized action outcome prediction has been previously shown (7) to provide a lower bound on utility obtained.

Environment two is mostly similar to environment one; the main differences, seemingly subtle, are that fitness is determined by utility accumulated over a finite lifespan (duration measured in decisions) rather than by average decision utility, and that after every generation, the fixed-size population undergoes a membership refresh during which the lowest-utility members leave no offspring. Additional key features of this environment: individual decisions depend on just one attribute, the utility landscape is two-dimensional, and although $b_{1,1} << b_{2,1}$, decisions contingent on attribute 2 are so rare that mean utility maximization dictates the allocation of both perceptual bits to attribute 1. Organisms interact with this environment via a 2-bit interface configured to either maximize mean utility ("optimal", Fig. 1d, right) or avoid large utility-harvesting failures during those few decisions contingent on attribute 2 ("risk averse", Fig. 1d, left).

Table 1. Simulations of competition between "risk averse" and "optimal" interface alleles

Condition	Lifetime Decisions	Probability of high-stakes (HS) scenario	Risk averse allele, initial prevalence	Instances of "risk averse" allele fixation
A	1000	.02	80%	8
B	1000	.015	80%	0
C	100	.015	80%	7
D	100	.015	50%	1
E	100	.02	50%	9
F	100	.02	20%	1
Comparison	**p**	**Parameter**		
A vs. B	.0007	p(HS)		
B vs. C	.0031	Lifetime length		
C vs. D	.0198	Initial prevalence		
D vs. E	.0011	p(HS)		
E vs. F	.0011	Initial prevalence		

The parameterization conditions A-F were each simulated 10 times. Then, for various pairs of conditions differentiated by just one parameter, it was hypothesized (and tested with Fisher's exact test, two-sided) that their "risk averse" fixation probabilities differed.

Simulations of competition between these interface strategy alleles (details in Methods) mostly favor "optimal"; notably, all observed fixations of "risk averse" occur when reproduction is restricted to the best-performing 20%, and this outcome is also favored by short organism lifetimes (Table 1). Thus, the victories of "risk averse" can best be understood as successful bet-hedges (13, 14), defined as the sacrifice of maximized average performance in order to avoid unacceptable risks (such as falling into the population's non-reproductive 80%), or as demonstrations of the advantageousness of risk-taking in competitions that reward relative standing (15, 16). These results show that when reproductive fitness deviates from linear dependence on average decision utility, strong correspondence between distortion loss and fitness objectives is still possible, but cannot be assumed.

In environments three through six, item utilities are not fixed in evolutionary memory; instead, they must be experienced and learned. It is assumed that percept utilities are learned, and that percept preference order is established, rapidly. It is further assumed, except when otherwise noted, that average post-learning decision utility determines fitness. Environments three and four consist of 100 items with randomly sampled utilities; the expected value derived from a particular division of these items amongst percepts is computed as described in Methods and Fig. 2.

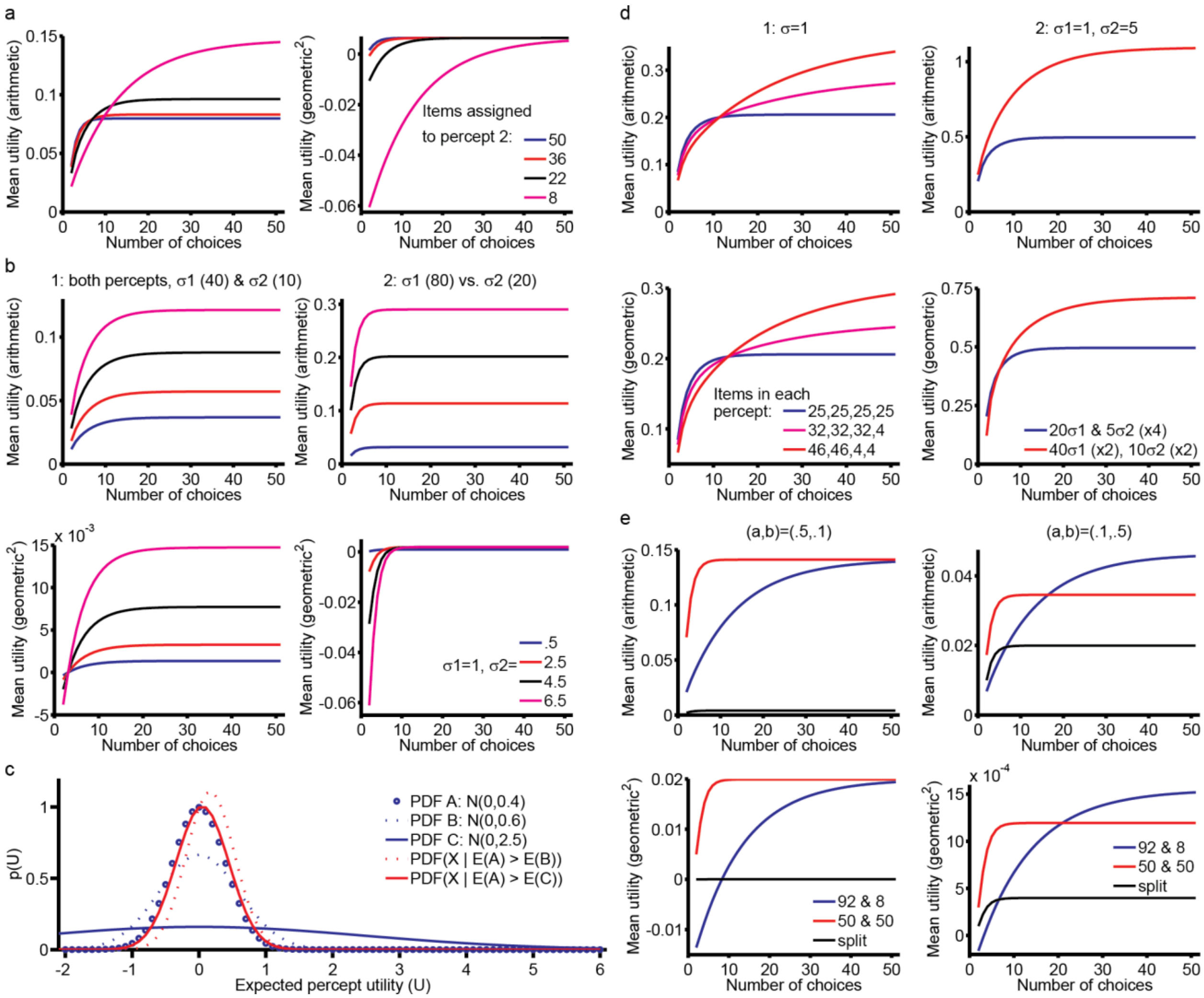

Figure 2. Optimizing the assignment of unknown-utility items to percepts whose utilities are learnable. a, Displays mean (arithmetic and geometric) decision utilities obtained using a 1-bit interface to distinguish amongst 100 items with utilities drawn independently from N(0,4); mean utility varies according to the number of choices presented and the division of items between percepts. **b**, Similar to (a), but with 80 and 20 utilities respectively drawn from $N(0,(\sigma 1)^2)$ and $N(0,(\sigma 2)^2)$; item-percept assignments are indicated in each plot-pair's title. **c**, When percept variances are similar (A vs. B), the low-variance percept's high-utility random values are more likely than its low-utility random values to become learned preferences (dotted red line). However, this likelihood distinction mostly disappears (solid red line) when percept variances differ significantly (A vs. C), because the high-variance percept's value almost wholly determines which percept is preferred. **d**, Displays mean utilities obtained by a 2-bit interface when using item-percept assignment strategies akin to either risk homogenization (blue), a "barbell" portfolio (red), or a hybrid (magenta). **e**, Here, item utilities belong to one of two clusters, $x_1+N(0,b^2)$ and $x_2+N(0,b^2)$; x_1 and x_2 are themselves drawn from $N(0,a^2)$. Line labels indicate either the sizes of percept-assigned clusters, or that each cluster's members are "split" between percepts.

In environment three, in which item utilities are uncorrelated, I find that the highest-performance strategies for maximizing the extraction of value from randomly-valued items use high-variance percepts to obtain large gains and reliably-available low-variance percepts to obtain steady gains. One-bit interfaces are faced with a trade-off between these strategies (Fig. 2a and 2b); thus, which configuration is favored depends on whether an environment selects for arithmetic or geometric mean utility. Fig. 2c illustrates the genesis of this trade-off; it clarifies why the pursuit of large gain potential from one percept comes at the cost of decreased gain potential from the other. This trade-off between percepts, in turn, suggests that unbalanced percepts deliver low geometric mean utility because their pursuit of utility is undiversified (17). Two-bit interfaces, meanwhile, are able to use low- and high-variability percepts in conjunction to form a percept volatility barbell (18) that can simultaneously exceed the arithmetic and geometric mean utilities obtained with homogenized volatilties (Fig. 2d).

In environment four, utilities are correlated: specifically, they belong to one of two clusters. I find (Fig. 2e) that even when within-cluster utility variation appreciably exceeds between-cluster variation, assigning the clusters to distinct percepts delivers greater mean utlity, arithmetic and geometric, than does splitting each cluster's items evenly between the percepts. The latter strategy is an advantageous bet-hedge only if items from one of the clusters are very rarely encountered, in which case environment four approximates environment three.

Environment five is like the simplest version of the first, save for one difference: while utility still varies linearly with a single attribute, the attribute value corresponding to the utility peak is no longer fixed at b_1, but instead assumes a random value within the interval $[0,b_1]$. (For a perhaps more biologically relevant formulation, let the organism encounter a multiplicity of decision situations, each with a unique utility-maximizing attribute value.) This change brings, if $\beta = 1$ or $n = 2$, the estimation error objective into alignment with the utility and distortion loss objectives (Supplementary Equations, Part 7).

Finally, environment six is an elaboration of environment four in which there are many item clusters (characterized by strong item utility correlations) of equal size, arranged in a 1D chain along which cluster utilities are (weakly) correlated with those of their immediate neighbors. If the number of clusters is a multiple of the interface's available percepts, then a cluster-percept assignment that evenly segments the chain serves as environment six's analog of the item-percept assignments of environment five. This analogy clarifies what information the interfaces of environments four and five transmit in common that enables their simultaneous optimization of utility and estimation fidelity: it is not item properties, but item similarities.

In this study, evolutionary optimization emerges as a potential unifying theoretical framework that establishes a hierarchy of perceptual processing adaptations and allows them to be understood as adaptations to noise. The present results position adaptations to environmental noise as primal, taking precedence over all others. This interpretation is corroborated by the existence of perceptual circuitry that is sensitive to (19) and well-adapted to the processing of (20) risk-indicative stimuli. The present results also put forth as a candidate secondary objective the mitigation of noise created by perceptual compression, i.e. the optimization of encoding, and clarify that the signal processing metric that describes its achievement is cost-weighted distortion. This objective and its associated solution serves to formalize long-standing notions that perceptual systems preferentially transmit information that reveals consequential utility gradients (21, 22). It also predicts categorical perception; for example, in birds whose strongest color difference sensitivity along the orange-red spectrum corresponds to a beak color indicator of mate desirability (23). Finally, the results from environments four through six demonstrate the benefits of perceptual similarity between items with correlated situational utilities, which may be provably adaptive in the face of noise associated with generalization of previous learning to novel situations, and previous studies present Bayesian inference as an effective adaptation to decoding noise (absent in this study). Thus, future elaborations of the present study, were they to incorporate other forms of noise, would very likely demonstrate the adaptiveness of additional information processing constructs and unveil further nuances of natural selection's superset of information processing objectives.

Methods

Interface allele competition. Three populations are simulated in parallel. The size of each is fixed at 100 individuals. Individuals differ according to whether their perceptual interface is "risk averse" (RA) or "optimal" (O). At simulation initialization, the prevalence of O is either 20%, 50%, or 80%; all three populations are initialized equivalently. Then, within each population, a lifetime of decisions is simulated for each member (the nth member of each population faces the same life decisions), the members are ranked by their lifetime's accrued utility, and the membership of the next generation is generated according to rules specific to each population. Lifetime simulation and population refresh procedures repeat until each population converges to fixation for one of the two alleles.

The first and second sets of refresh rules are simple: the bottom 80% and 50% (respectively) of the utility accumulators are eliminated, and the set remaining is copied until the population size is restored to 100. The third set eliminates the bottom 10% and replaces each member thereof with an individual whose probability of exhibiting the O allele is proportional to the O allele's prevalence in the top 90%.

The decisions made within a lifetime involve choosing between two items sampled (with uniform probability) from the utility landscape in Fig. 1d, with $b_{1,1}$ = 100 and $b_{2,1}$ = 1100. Most decision utilities are contingent upon attribute 1; those contingent upon attribute 2 are deemed "high stakes" (HS), and arise with probability p(HS). In these simulations, p(HS) = .02 or .015; the result in Supplementary Equations, Part 5 indicates that the O interface (Fig. 1d, right) achieves higher average decision utility than the RA interface (Fig. 1d, left) when P(HS) < .022.

Expected utility computation. Because percepts are composed of items whose utilities are drawn from normal distributions, the distributions of mean percept values are also normal. Also, once percept utilities are learned, the organism will preferentially select items mapped to higher-utility percepts; a percept's expected value, therefore, is related to its position in the percept preference order. If it is further specified that the interface has N percepts, and that a particular sampling of item utilities resulted in percept 1 being the highest-utility percept, then the expected value of percept 1, $E(P_1)$, can be expressed as

$$\int_{-\infty}^{\infty} x * N_{PDF}(0, \sigma_1^2) * \prod_{i=2}^{N} N_{CDF}(0, \sigma_i^2) * x dx$$

Alternatively, if $E(P_1) < E(P_2)$ and $E(P_1) > E(P_i)$ ∀ i > 2, then

$$E(P_1) = \int_{-\infty}^{\infty} x * \left(1 - N_{CDF}(0, \sigma_2^2)\right) * N_{PDF}(0, \sigma_2^2) * \prod_{i=3}^{N} N_{CDF}(0, \sigma_i^2) * x dx$$

The expressions for all expected percept values across all preference orderings are constructed similarly. In the present study, these expressions are evaluated by numerical integration. The results in Fig. 2 are mostly obtained by integrating over [-5,5] with dx = .001, those in Fig. 2c are obtained by integrating over [-6,6] with dx = .01.

A notable feature of the above expressions is that each percept has an associated σ-value. These are computed individually for every percept across all item-percept assignments, using standard methods. For example, for a percept containing 25 items with utilities drawn from N(0,1), σ = 1/5.

Upon computing the ordering-specific expected percept values, it becomes possible to combine them to compute the mean utility obtained by a particular assignment of 100 items to the available percepts. Specifically, the respective arithmetic and geometric mean utilities are $\sum_j p(O_j) * E(O_j)$ and $\prod_j \left(E(O_j)\right)^{p(O_j)}$, where the O_j represent the assignment's potential percept preference orderings. For each O_j, using $s_{j,k}$ to represent the size (in items) of the kth percept in ordering j,

$$E(O_j) = \sum_{i=1}^{N} E(P_{j,i}) * \left(\left(\sum_{k \le i} s_{j,k}\right)^n - \left(\sum_{k > i} s_{j,k}\right)^n\right) / 100^2$$

For the 1-bit interface, $p(O_j) = .5$ ∀ j, and for the 2-bit interface, $p(O_j) = .25$ ∀ j when all percept variances are equal. For the orderings deriving from the 2-bit interface and unequal-variance percepts, the $p(O_j)$ are numerically approximated as follows. Define g as the probability that a percept has a higher utility than all of the smaller-variance percepts. For the interface with three similarly-constructed percepts and a 4th, higher-variance percept (Fig. 2d1, magenta),

$$g = \int_{-\infty}^{\infty} x * N_{PDF}(0, \sigma_4^2) * \prod_{i=1}^{3} N_{CDF}(0, \sigma_i^2)\, dx$$

For the interfaces with two pairs of equal-variance percepts (Fig. 2c, red), if the high-variance percepts are designated as P_3 and P_4, then

$$g - g^2/2 = \int_{-\infty}^{\infty} x * N_{PDF}(0, \sigma_4^2) * \prod_{i=1}^{3} N_{CDF}(0, \sigma_i^2)\, dx$$

The g-$g^2/2$ expression arises because the integral represents the probability (g) that $E(P_4) > E(P_1)$ and $E(P_2)$, minus the probability (g^2) that $E(P_3) > E(P_1)$ and $E(P_2)$ is also true, plus the probability ($g^2/2$) that under such circumstances $E(P_4) > E(P_3)$. Table 2 (below) shows how $p(O_j)$ values derive from g, and catalogs the numerical approximation outcomes used in the computation of mean utilities for the 2-bit interface with heterogeneous percepts.

Code and Data are available at github.com/victorqz1/PerceptualCompression1.

Table 2: Percept ordering probabilities, 2-bit interface

Percept σ values	Ordering (O_j)	Probability integral	Ordering Probability ($p(O_j)$)	
			Numerical	Symbolic
P_1: 0.1768 P_2: 0.5	P_1, P_1, P_1, P_2		0.3864	g
	P_1, P_1, P_2, P_1		0.1136	f/2
	P_1, P_2, P_1, P_1		0.1136	f/2
	P_2, P_1, P_1, P_1	g = 0.3864	0.3864	g
P_1: 0.1474 P_2: 0.5	P_1, P_1, P_2, P_2		0.1834	g^2
	P_1, P_2, P_1, P_2		0.1228	2*f*g
	P_1, P_2, P_2, P_1		0.0206	f^2
	P_2, P_2, P_1, P_1	$g-g^2/2$ = 0.3366	0.1834	g^2
	P_2, P_1, P_2, P_1	0.3366	0.1228	2*f*g
	P_2, P_1, P_1, P_2	0.3366	0.3669	$2*g^2$
P_1: 0.1581 P_2: 1.5811	P_1, P_1, P_2, P_2		0.2256	g^2
	P_1, P_2, P_1, P_2		0.0475	2*f*g
	P_1, P_2, P_2, P_1		0.0025	f^2
	P_2, P_2, P_1, P_1	$g-g^2/2$ = 0.3622	0.2256	g^2
	P_2, P_1, P_2, P_1	0.3622	0.0475	2*f*g
	P_2, P_1, P_1, P_2	0.3622	0.4512	$2*g^2$

f = 1-2*g

References

1. Sims CR. Rate-distortion theory and human perception. Cognition. 2016;152:181-98.
2. Sims CR. Efficient coding explains the universal law of generalization in human perception. Science. 2018;360:652-6.
3. Marzen SE, DeDeo S. The evolution of lossy compression. J R Soc Interface. 2017;14:20170166.
4. Linsker R. Self-organization in a perceptual network. IEEE Computer. 1988;21:105-17.
5. Wei X-X, Stocker AA. Lawful relation between perceptual bias and discriminability. Proc Nat Acad Sci. 2017;114:10244-9.
6. Polanía R, Woodford M, Ruff CC. Efficient coding of subjective value. Nature Neuroscience. 2018;22:134-42.
7. Friston K. The free-energy principle: a unified brain theory? Nat Rev Neurosci. 2010;11:127-38.

8. Knill DC, Pouget A. Bayesian brain: the role of uncertainty in neural coding and computation. Trends Neurosci. 2004;27:712-9.
9. Moreno-Bote R, Knill DC, Pouget A. Bayesian sampling in visual perception. Proc Natl Acad Sci. 2011;108:12491-6.
10. Geisler WS, Diehl RL. A Bayesian approach to the evolution of perceptual and cognitive systems. Cog Sci. 2003;27:379-402.
11. Singh M, Hoffman D. Natural selection and shape perception. In: Dickinson SJ, Pizlo Z, editors. Shape Perception in Human and Computer Vision. London: Springer Verlag; 2013. p. 171-86.
12. Mark JT, Marion BB, Hoffman DD. Natural selection and veridical perceptions. J Theor Biol. 2010;256:504-15.
13. King OD, Masel J. The evolution of bet-hedging adaptations to rare scenarios. Theoretical Population Biology. 2007;72:560-75.
14. Frank SA. Natural selection. I. Variable environments and uncertain returns on investment. J Evol Biol. 2011;24:2299–309
15. Rtischev D. Risk-bearing in a winner-take-all contest. Gakushuin Economic Papers. 2008;45:65-79.
16. Dekel E, Scotchmer S. On the evolution of attitudes towards risk in winner-take-all games. J Econ Theory. 1999;87:125-43.
17. Markowitz HM. Portfolio Selection: Efficient Diversification of Investments. 2nd ed. Malden, MA: Blackwell Publishers Ltd; 1991.
18. Geman D, Geman H, Taleb NN. Tail Risk Constraints and Maximum Entropy. Entropy. 2015;17:3724-37.
19. DeLoache JS, LoBue V. The narrow fellow in the grass: human infants associate snakes and fear. Developmental Science. 2009;12:201-7.
20. Hein AM, Gil MA, Twomey CR, Couzin ID, Levin SA. Conserved behavioral circuits govern high-speed decision-making in wild fish shoals. Proc Nat Acad Sci. 2018;115:12224-8.
21. Wehner RJ. 'Matched filters' — neural models of the external world. Comp Physiol A. 1987;161:511-31.
22. Shepard RN. Toward a universal law of generalization for psychological science. Science. 1987;237:1317-23.
23. Caves EM, Green PA, Zipple MN, Peters S, Johnsen S, Nowicki S. Categorical perception of colour signals in a songbird. Nature. 2018;560:365-7.

Supplementary Equations

Note: in these analyses, the range of utility values is assumed to be spanned by a density of items sufficient to justify an integral approximation for the calculation of: utility estimation error, item-percept mutual information, expected utility, and distortion loss.

Part 1: derivation of the objective function for distortion loss minimization

Distortion during a single decision: given a set of n items available during a particular decision round, one can label their utilities x_1:x_n, indexed in order of descending utility. A perceptual system that transmits environmental information perfectly would obtain utility x_1, as would a perceptual system that assigns the x_1-utility item to a category of its own. However, if there are k items that appear to be of the same, most-preferred percept, then the expected utility of the decision round is $\sum_{i=1}^{k} x_i \,/k$. The distortion loss is the expected magnitude of the utility loss that results from the perceptual system's compression of the world's information; it's magnitude is $x_1 - \sum_{i=1}^{k} x_i \,/k$.

Total distortion (TD), then, is simply the sum of the distortion losses over all possible utility value sets X = {x_1,…,x_n} when the interface uses the boundaries B = {b_2,…,b_m}. As an illustration of how boundaries affect the distortion, consider the set Q of all item sets {x_1,…,x_n} that satisfy $b_i < x_1$:$x_k < b_{i+1} < x_{k+2}$:x_n. The contribution of Q to TD is

$$Q = \int_{b_{i+1}}^{b_i} \int_{b_{i+1}}^{x_1} \dots \int_{b_{i+1}}^{x_{k-1}} \left(x_1 - \frac{\sum_{i=1}^{k} x_i}{k} \right) \int_0^{b_{i+1}} \int_0^{x_{k+1}} \dots \int_0^{x_{n-1}} p(X) dx_n dx_{n-1} \dots dx_1$$

Note that the limits of integration reflect the ordinal arrangement of the x_i, and their position relative to the b_i. Note also that the distortion depends on p(X). Generalizing this expression across all boundaries i and best-percept item counts k, and noting that $Q_{i1} = 0$, yields $TD = \sum_{i=1}^{m} \sum_{k=2}^{n} Q_{ik}$.

Part 2: proof that for the general game with m percepts, n choices, a linearly increasing and unidimensional utility continuum, and a uniform distribution of item probabilities over that continuum, the perceptual category boundaries that optimize total distortion (TD) and utility (U) are equivalent

Approach: the optima of U and TD are equivalent if ∀ i, $dU/db_i = C*dTD/db_i$, C constant.

2.1: expression for dU/db_i

Recall from main text that: b_i and b_{i+1} are the upper and lower boundaries of P_i; note also that $b_{m+1} = 0$, $E(P_i) = (b_i+b_{i+1})/2$, and

$$p(P_i) = P(\text{no items from percept } x < i)*P(\text{at least one item from percept } i)$$

$$= (b_i/b_1)^n*(1-(b_{i+1}/b_i)^n) = ((b_i)^n-(b_{i+1})^n)/(b_1)^n$$

Thus, utility can be written as

$$U = \sum_{i=1}^{m} E(P_i)P(P_i) = \sum_{i=1}^{m} (b_i + b_{i+1})(b_i^n - b_{i+1}^n)/2b_1^n$$

$$= (\ldots + (b_{i-1} + b_i)(b_{i-1}^n - b_i^n) + (b_i + b_{i+1})(b_i^n - b_{i+1}^n) + \cdots)/2b_1^n$$

Setting $dU/db_i = 0$ and solving yields

$$b_i = \left(\frac{b_{i-1}^n - b_{i+1}^n}{n(b_{i-1} - b_{i+1})}\right)^{1/(n-1)}$$

2.2: simplifying Q_{ik}

$$Q_{ik} = \int_{b_{i+1}}^{b_i} \int_{b_{i+1}}^{x_1} \ldots \int_{b_{i+1}}^{x_{k-1}} \left(x_1 - \frac{\sum_{i=1}^{k} x_i}{k}\right) \int_0^{b_{i+1}} \int_0^{x_{k+1}} \ldots \int_0^{x_{n-1}} p(X) dx_n dx_{n-1} \ldots dx_1$$

First, set aside the p(X) term; in this environment, it is constant ∀ X
Second,

$$\int_0^{b_{i+1}} \int_0^{x_{k+1}} \ldots \int_0^{x_{n-1}} 1 dx_n \ldots dx_{k+1} = b_{i+1}^{n-k}/(n-k)!$$

Third, define $C_1 = (b_{i+1})^{n-k}/k(n-k)!$
Fourth,

$$x_1 - \sum_{i=1}^{k} x_i/k = \left((k-1)x_1 - \sum_{i=2}^{k} x_i\right)/k$$

These four steps allow Q_{ik} to be re-written as

$$C_1 \int_{b_{i+1}}^{b_i} (k-1)x_1 \int_{b_{i+1}}^{x_1} \ldots \int_{b_{i+1}}^{x_{k-1}} 1 dx_k \ldots dx_1 - C_1 \int_{b_{i+1}}^{b_i} \int_{b_{i+1}}^{x_1} \ldots \int_{b_{i+1}}^{x_{k-1}} \sum_{i=2}^{k} x_i \, dx_k \ldots dx_1$$

Fifth,

$$\int_{b_{i+1}}^{x_1} \ldots \int_{b_{i+1}}^{x_{k-1}} 1 dx_k \ldots dx_2 = (x_1 - b_{i+1})^{k-1}/(k-1)!$$

Sixth,

$$\int_{b_{i+1}}^{x_1} \dots \int_{b_{i+1}}^{x_{k-1}} \sum_{i=2}^{k} x_i \, dx_k \dots dx_2 = (b_{i+1} + x_1) * (x_1 - b_{i+1})^{k-1} / (2 * (k-2)!)$$

Therefore,

$$Q_{ik} = C_1/(k-2)! \int_{b_{i+1}}^{b_i} (x_1 - b_{i+1})^k / 2 \, dx_1 = \frac{b_{i+1}^{n-k}(k-1)(b_i - b_{i+1})^{k+1}}{2(k+1)!\,(n-k)!}$$

2.3: deriving and simplifying TD_i

Defining TD_i as $\sum_{k=2}^{n} Q_{ik}$ leads to the expression given in TD_{i1}, expansion of the binomial exponential reveals DL_{i2}, and expansion of the double summation followed by the grouping of terms by exponent of b_i reveals DL_{i3}.

$$TD_{i1} = \sum_{k=2}^{n} \frac{b_{i+1}^{n-k}(k-1)(b_i - b_{i+1})^{k+1}}{2(k+1)!\,(n-k)!}$$

$$TD_{i2} = \frac{1}{2} * \sum_{k=2}^{n} \frac{(k-1)\sum_{j=0}^{k+1}\binom{k+1}{j}(-1)^j b_i^{k+1-j} b_{i+1}^{n-k+j}}{(k+1)!\,(n-k)!}$$

$$TD_{i3} = \frac{1}{2} * \sum_{L=0}^{n+1} C_L b_i^L b_{i+1}^{n+1-L}$$

In TD_{i3}, the coefficients C_L can be derived from noting that in TD_{i2}, for all powers of b_i, there is a one-to-one correspondence between the outer sum's k values and the inner sum's j values. Thus,

$$C_L = \sum_{k=2}^{n} \frac{(-1)^{k+1-L}\binom{k+1}{k+1-L}(k-1)}{(k+1)!\,(n-k)!} \rightarrow \frac{1}{L!} \sum_{k=\max(2,L-1)}^{n} \frac{(-1)^{k+1-L}(k-1)}{(k+1-L)!\,(n-k)!}$$

The rightmost expression is the consequence of observing that k+1-L is potentially negative for certain combinations of (k,L), and that in DL_{i2} there is no b_i^L term in the inner sum while L < k+2. This formulation yields

$$C_0 = \sum_{k=2}^{n} \frac{(-1)^{k+1}(k-1)}{(k+1)!\,(n-k)!} = \frac{\sum_{k=0}^{n}(-1)^k(k-2)T_{k,n+1} - \sum_{k=0}^{2}(-1)^k(k-2)T_{x,n+1}}{(n+1)!} = \frac{1-n}{(n+1)!}$$

$$C_1 = \sum_{k=2}^{n} \frac{(-1)^k (k-1)}{(k)!\,(n-k)!} = \frac{1}{n!} * \left(\sum_{k=0}^{n} (-1)^k (k-1) T_{k,n} - \sum_{k=0}^{1} (-1)^k (k-1) T_{k,n} \right) = 1/n!$$

$$C_2 = \frac{1}{2} \sum_{k=2}^{n} \frac{(-1)^{k-1}}{(k-2)!\,(n-k)!} = \sum_{k=0}^{i} (-1)^k T_{k,i} = 0$$

$$C_{2<L<n-1} = \frac{1}{6} \sum_{k=L-1}^{n} \frac{(-1)^{k-2}(k-1)}{(k-L+1)!\,(n-k)!} = \sum_{k=0}^{i} (-1)^k (L-2+k) T_{k,i} = 0$$

$$C_n = -1/n!$$

$$C_{n+1} = (n-1)/(n+1)!$$

Combining all of the terms derived yields

$$TD_i = \frac{b_i^{n+1}(n-1)}{2(n+1)!} - \frac{b_i^n b_{i+1}}{2n!} + \frac{b_i b_{i+1}^n}{2n!} - \frac{b_{i+1}^{n+1}(n-1)}{2(n+1)!}$$

In the above expressions, $T_{k,i} = \binom{i}{k}$ and represents the kth term of the ith row of Pascal's triangle. In Part 2.3.1 it is shown that certain sums over $T_{k,i}$ terms equal 0.

2.3.1: 0 = $\sum_{x=0}^{i} (-1)^x T_{x,i}$ = $\sum_{x=0}^{i} (-1)^x (X+x) T_{k,i}$ (X is an arbitrary integer)

A well-known feature of Pascal's triangle is that $T_{k,i} = T_{k-1,i-1} + T_{k,i-1}$ ($T_{-1,i} = T_{i+1,i} = 0$). Therefore, the $T_{k,i}$ terms in the sum $\sum_{k=0}^{i} (-1)^k T_{k,i}$ can be replaced with $T_{k,i-1}$ terms; additionally decomposing the sum itself into its positive and negative components yields

$$\sum_{k=-1}^{i} T_{k,i-1} - \sum_{k=0}^{i-1} T_{k,i-1} = T_{-1,i-1} + T_{i,i-1} + \sum_{k=0}^{i-1} T_{k,i-1} - \sum_{k=0}^{i-1} T_{k,i-1} = 0$$

It is also true that $T_{k,i} = T_{k-2,i-2} + 2*T_{k-1,i-2} + T_{k,i-2}$. Proposing that $\sum_{k=0}^{i} (-1)^k (X+k) T_{k,i}$ can be expressed as $\sum_{k=-2}^{i} D_k T_{k,i-2}$ the coefficients D_k can be derived by noting first that individual $T_{k,i-2}$ terms only contribute to $T_{n,i}$ for n∈[k:k+2]; within [k:k+2], observe that

$$\begin{aligned}
&(X+k)*T_{k,i}-(X+k+1)*T_{k+1,i}+(X+k+2)*T_{k+2,i} \\
&= (X+k)*(T_{k-2,i-2}+2*T_{k-1,i-2}+T_{k,i-2}) \\
&- (X+k+1)*(T_{k-1,i-2} + 2*T_{k,i-2}+T_{k+1,i-2}) \\
&+ (X+k+2)*(T_{k,i-2}+2*T_{k+1,i-2}+T_{k+2,i-2}) \\
&= \text{Constant}+T_{k,i-2}*(X+k-2*(X+k+1)+X+k+2) = \text{Constant}+T_{k,i-2}*0
\end{aligned}$$

Since the individual $D_k = 0$, $\sum_{k=-2}^{i} D_k T_{k,i-2} = 0 = \sum_{k=0}^{i} (-1)^k (X+k) T_{k,i}$

2.4: Expression for dTD/db_i

$$TD = \ldots + TD_i + TD_{i+1} + \ldots$$

$$= \cdots + \frac{b_{i-1}^{n+1}(n-1)}{2(n+1)!} - \frac{b_{i-1}^{n} b_i}{2n!} + \frac{b_{i-1} b_i^{n}}{2n!} - \frac{b_i^{n+1}(n-1)}{2(n+1)!} + \frac{b_i^{n+1}(n-1)}{2(n+1)!} - \frac{b_i^{n} b_{i+1}}{2n!} + \frac{b_i b_{i+1}^{n}}{2n!} - \frac{b_{i+1}^{n+1}(n-1)}{2(n+1)!} + \cdots$$

Setting $dTD/db_i = 0$ and solving yields

$$b_i = \left(\frac{b_{i-1}^n - b_{i+1}^n}{n(b_{i-1} - b_{i+1})}\right)^{1/(n-1)}$$

Final note: rewriting TD and U (from Part 2.1) as shown below reveals that TD and U vary in opposite directions with respect to $b_{i-1}(b_i)^n - b_i(b_{i-1})^n$. So, the b_i that maximize utility will minimize distortion.

$$TD = C_1 b_1^{n+1} + C_2 \sum_{i=2}^{m} (b_{i-1} b_i^n - b_{i-1}^n b_i)$$

$$U = C_3 b_1^{n+1} + C_4 \sum_{i=2}^{m} (b_{i-1}^n b_i - b_{i-1} b_i^n)$$

Part 3: proof that, for two-item decisions on the attribute-utility landscape in Figure 1b1, the perceptual boundary that optimizes utility also optimizes distortion

In this decision environment, setting $a > 1$, utility as a function of attribute value x is

$$u(x) = \begin{cases} x, x < b_1/2 \\ b_1/2 + a_1(x - b_1/2), x > b_1/2 \end{cases}$$

From Part 2.1, for a 1-bit interface, $U = p(P_1)E(P_1)+p(P_2)E(P_2)$. Also, assuming (reasonably) that $b_2 > b_1/2$,

$$p(P_1) = ((b_1)^2-(b_2)^2)/(b_1)^2,\ p(P_2) = (b_2/b_1)^2$$

$$E(P_1) = (b_1/2+a_1{*}(b_2-b_1/2)+(1+a_1){*}b_1/2)/2 = (b_1+a_1{*}b_2)/2$$

$$E(P_2) = \left(\int_0^{b_1/2} x dx + \int_{b_1/2}^{b_2} b_1/2 + a_1(x - b_1/2) dx\right)/b_2$$

Setting $dU/db_2 = 0$ and solving yields $b_2 = b_1{*}(5/8-a^{-1}/8)$.

Considering now the optimization of TD, the present decision environment affords three simplifications to the integrand, $p(X)(x_1 - \sum_{i=1}^{k} x_i / k)$. First, p(X) is a constant. Second, $x_1 - \sum_{i=1}^{k} x_i / k$ simplifies to $(x_1-x_2)/2$. Third, when x_1 and x_2 correspond to different percepts, the integrand is zero. Therefore,

$$TD = \int_0^{b_2} \int_0^{x_1} \frac{x_1 - x_2}{2} dx_2 dx_1 + a_1 \int_{b_2}^{b_1} \int_{b_2}^{x_1} \frac{x_1 - x_2}{2} dx_2 dx_1$$

Here, it is illuminating to split the first double integral into the following components:

$$\int_0^{b_1/2} \int_0^{x_1} \frac{x_1 - x_2}{2} dx_2 dx_1 + \int_{b_1/2}^{b_2} \int_0^{b_1/2} \frac{b_1/2 + a_1(x_1 - b_1/2) - x_2}{2} dx_2 dx_1 + a_1 \int_{b_1/2}^{b_2} \int_{b_1/2}^{x_1} \frac{x_1 - x_2}{2} dx_2 dx_1$$

Setting $dTD/db_2 = 0$ and solving yields $b_2 = b_1*(5/8-1/(8*a_1))$.

Part 4: proof that, for two-item decisions on the attribute-utility landscape in Figure 1b2, the perceptual boundary that optimizes utility also optimizes distortion

In this decision environment, with $a_2 \in [0,1/b_1]$, the item PDF as a function of attribute value x, and the resulting utility function, are given below:

$$p(x) = \begin{cases} 1/b_1 - a_2, x < b_1/2 \\ 1/b_1 + a_2, x > b_1/2 \end{cases}$$

$$U = \left(1 - \left(\int_0^{b_2} p(x)\right)^2\right) * \left(\frac{\int_{b_2}^{b_1} p(x)u(x)}{\int_{b_2}^{b_1} p(x)}\right) + \left(\int_0^{b_2} p(x)\right)^2 * \left(\frac{\int_0^{b_2} p(x)u(x)}{\int_0^{b_2} p(x)}\right)$$

U reduces to a more familiar form when $a_2 = 1$. For $a_2 \in [0,1/b_1]$, $\int_{b_2}^{b_1} p(x)u(x) / \int_{b_2}^{b_1} p(x)$ simplifies to $(b_1+b_2)/2$ because p(x) is constant in $[b_2,b_1]$. Also, for specificity's sake:

$$\int_0^{b_2} p(x) = \int_0^{b_1/2} 1/b_1 - a_2 dx + \int_{b_1/2}^{b_2} 1/b_1 + a_2 dx$$

$$\int_0^{b_2} p(x)u(x) = \int_0^{b_1/2} x(1/b_1 - a_2)dx + \int_{b_1/2}^{b_2} x(1/b_1 + a_2)dx$$

Setting $dU/db_2 = 0$ and solving yields $b_2 = b_1*(1/2+a_2*b_1/4)$, which is a linear function of a_2 with range $[b_1/2, 3*b_1/4]$.

Meanwhile, letting $p_1(x_1,x_2) = (1/b_1-a_2)^2$, $p_2(x_1,x_2) = (1/b_1-a_2)*(1/b_1+a_2)$, and $p_3(x_1,x_2) = (1/b_1+a_2)^2$,

$$TD = \int_0^{\frac{b_1}{2}} \int_0^{x_1} p_1(x_1,x_2)\frac{x_1-x_2}{2}dx_2dx_1 + \int_{b_1/2}^{b_2}\int_0^{b_1/2} p_2(x_1,x_2)\frac{x_1-x_2}{2}dx_2dx_1 + \int_{b_1/2}^{b_2}\int_{b_1/2}^{x_1} p_3(x_1,x_2)\frac{x_1-x_2}{2}dx_2dx_1 + \int_{b_2}^{b_1}\int_{b_2}^{x_1} p_3(x_1,x_2)\frac{x_1-x_2}{2}dx_2dx_1$$

Setting $dTD/db_2 = 0$ and solving yields $b_2 = b_1*(1/2+a_2*b_1/4)$.

Part 5: demonstration that the bit-attribute assignments that maximize utility are the same as those that minimize distortion (2 bits, 2 independent and independently-perceived attributes, 2 item choices, uniform item probabilities over attribute values)

5.1: definition of problem and notation

The ranges of attributes 1 and 2 are $[0,b_{1,1}]$ and $[0,b_{2,1}]$, respectively. If item utilities are the sum of attribute values, utility and distortion are abbreviated U_s and D_s, respectively. Alternatively, if utility is situationally-determined, depending on attributes 1 and 2 with probabilities p and 1-p (respectively), I use the abbreviations U_p and D_p. These abbreviations are in turn superscripted according to the attributes in which perceptual bits are invested.

Thus, if both bits are invested in differentiation along attribute 1,

$$U_s^{11} = \frac{b_{2,1}}{2} + \sum_{i=1}^{4}(b_{1,i}+b_{1,i+1})(b_{1,i}^n - b_{1,i+1}^n)/2b_{1,1}^n$$

$$U_p^{11} = (1-p)*\frac{b_{2,1}}{2} + p*\sum_{i=1}^{4}(b_{1,i}+b_{1,i+1})(b_{1,i}^n - b_{1,i+1}^n)/2b_{1,1}^n$$

$$D_s^{11} = \sum_{i=1}^{4}\sum_{k=2}^{n} Q_{ik}^{(1)} + \int_0^{b_{2,1}}\int_0^{x_1}\dots\int_0^{x_{n-1}}\left(x_1 - \frac{\sum_{i=1}^{n} x_i}{n}\right)$$

$$D_p^{11} = p*\sum_{i=1}^{4}\sum_{k=2}^{n} Q_{ik}^{(1)} + (1-p)*\int_0^{b_{2,1}}\int_0^{x_1}\dots\int_0^{x_{n-1}}\left(x_1 - \frac{\sum_{i=1}^{n} x_i}{n}\right)dx_n\dots dx_1$$

$$Q_{ik}^{(1)} = \int_{b_{1,i+1}}^{b_{1,i}}\int_{b_{1,i+1}}^{x_1}\dots\int_{b_{1,i+1}}^{x_{k-1}}\left(x_1 - \frac{\sum_{i=1}^{k} x_i}{k}\right)\int_0^{b_{1,i+1}}\int_0^{x_{k+1}}\dots\int_0^{x_{n-1}} dx_n dx_{n-1}\dots dx_1$$

If one bit is invested in differentiation along each attribute,

$$U_s^{12} = \frac{(b_{1,1}+b_{1,2})(b_{1,1}^n - b_{1,2}^n) + b_{1,2}^{n+1}}{2b_{1,1}^n} + \frac{(b_{2,1}+b_{2,2})(b_{2,1}^n - b_{2,2}^n) + b_{2,2}^{n+1}}{2b_{2,1}^n}$$

$$U_p^{12} = p * \frac{(b_{1,1}+b_{1,2})(b_{1,1}^n - b_{1,2}^n) + b_{1,2}^{n+1}}{2b_{1,1}^n} + (1-p) * \frac{(b_{2,1}+b_{2,2})(b_{2,1}^n - b_{2,2}^n) + b_{2,2}^{n+1}}{2b_{2,1}^n}$$

$$D_s^{12} = Q^{(2)} + Q^{(3)}$$

$$D_p^{12} = p * Q^{(2)} + (1-p) * Q^{(3)}$$

$$Q^{(2)} = \int_0^{b_{1,2}} \int_0^{x_1} \dots \int_0^{x_{n-1}} \left(x_1 - \frac{\sum_{i=1}^n x_i}{n}\right) dx_n \dots dx_1 + \sum_{k=2}^n \int_{b_{1,2}}^{b_{1,1}} \int_{b_{1,2}}^{x_1} \dots \int_{b_{1,2}}^{x_k} \left(x_1 - \frac{\sum_{i=1}^k x_i}{k}\right) \int_0^{b_{1,2}} \int_0^{x_{k+1}} \dots \int_0^{x_{n-1}} dx_n \dots dx_1$$

To obtain $Q^{(3)}$ from $Q^{(2)}$, respectively replace $b_{1,1}$ and $b_{1,2}$ with $b_{2,1}$ and $b_{2,2}$.

5.2: along attribute a, for a particular n and boundaries b_i, $U_a = C_1 * b_{a,1}$ and $D_a = C_2 * b_{a,1}$; so, if $b_{a,1}$ is scaled by f, so are U_a and D_a.

First, from Parts 2.1 and 2.4, $b_i = \left(\frac{b_{i-1}^n - b_{i+1}^n}{n(b_{i-1} - b_{i+1})}\right)^{1/(n-1)}$. If b_1 is scaled by f, and all other b_i are also scaled by f, then this formula still holds true:

$$\left(\frac{f^n b_{i-1}^n - f^n b_{i+1}^n}{n(f * b_{i-1} - f * b_{i+1})}\right)^{1/(n-1)} = \left(\frac{f^n}{f} * \frac{b_{i-1}^n - b_{i+1}^n}{n(b_{i-1} - b_{i+1})}\right)^{1/(n-1)} = f * b_i$$

Second, from Part 2.1, $U = \sum_{i=1}^m (b_i + b_{i+1})(b_i^n - b_{i+1}^n)/2b_1^n$; scaling all b_1 by f gives

$$\sum_{i=1}^m (f * b_i + f * b_{i+1})(f^n b_i^n - f^n b_{i+1}^n)/(2f^n b_1^n) = f \sum_{i=1}^m (b_i + b_{i+1})(b_i^n - b_{i+1}^n)/2b_1^n = f * U$$

Third, from Parts 1 and 2.4,

$$TD = \frac{p(X)}{2n!} * \left(b_1^{n+1}(n-1) + \sum_{i=2}^m b_{i-1} b_i^n - b_{i-1}^n b_i\right)$$

Assuming uniform item probability over each attribute's span, $p(X) = 1/(b_{a,1})^n$. Thus, substituting $f * b_{a,1}$ for $b_{a,1}$,

$$\frac{1}{f^n b_1^n} * \frac{1}{2n!} * \left(f^{n+1} b_1^{n+1}(n-1) + \sum_{i=2}^{m} f^{n+1} b_{i-1} b_i^n - b_{i-1}^n b_i \right) = f * TD$$

5.3: for all U^{11} and D^{11} vs. U^{12} and D^{12}, the optimal bit-attribute allocations are analogous

Knowing that utilities and distortions are linear functions of $b_{a,1}$ allows several simplifications to be made to the $U^{()}$ and $D^{()}$ expressed in Part 5.1. The simplifications make use of the following expressions (which correspond to n = 2; when β > 1, closed-form expressions are not obtainable for n > 2; see Part 5.3.1):

$U_{\beta=k} = \sum_{i=1}^{k}(b_i + b_{i+1})(b_i^2 - b_{i+1}^2)/2b_1^2$ = b_1*(1/2,5/8,21/32) for k=(0,1,2)

$D_{\beta=k} = \frac{2^k}{b_1^2}\int_0^{b_1/2^k}\int_0^{x_1}\frac{x_1-x_2}{2}dx_2dx_1$ = b_1*(1/12,1/48,1/192) for k=(0,1,2)

In the above solutions, using the result from Part 2.4, $b_2 = b_1/2$ when β = 1, and $(b_2,b_3,b_4) = (.75b_1,.5b_1,.25b_1)$ when β = 2. For further simplicity, $b_{1,1} = 1$ and $R = b_{2,1}/b_{1,1}$; now, the $U^{()}$ and $D^{()}$ become:

$$U_s^{11} = \frac{R}{2} + \frac{21}{32},\ U_s^{12} = \frac{5}{8} + \frac{5}{8}R,\ U_s^{22} = \frac{1}{2} + \frac{21}{32}R$$

Thus, $U_s^{11} > U_s^{12}$ ($U_s^{22} > U_s^{12}$) when R < .25 (R > 4)

$$U_p^{11} = (1-p) * \frac{R}{2} + p * \frac{21}{32},\ U_p^{12} = p * \frac{5}{8} + (1-p) * \frac{5}{8}R,\ U_p^{22} = \frac{(1-p)}{2} + pR * \frac{21}{32}$$

Thus, $U_p^{11} > U_p^{12}$ ($U_p^{22} > U_p^{12}$) when R*(1-p)/p < .25 (R*(1-p)/p > 4)

$$D_s^{11} = \frac{1}{192} + \frac{R}{12},\ D_s^{12} = \frac{1}{48} + \frac{R}{48},\ D_s^{22} = \frac{R}{192} + \frac{1}{12}$$

Thus, $D_s^{11} > D_s^{12}$ ($D_s^{22} > D_s^{12}$) when R < .25 (R > 4)

$$D_p^{11} = \frac{p}{192} + (1-p) * \frac{R}{12},\ D_p^{12} = \frac{p}{48} + (1-p) * \frac{R}{48},\ D_p^{22} = \frac{pR}{192} + \frac{1-p}{12}$$

Thus, $D_p^{11} > D_p^{12}$ ($D_p^{22} > D_p^{12}$) when R*(1-p)/p < .25 (R*(1-p)/p > 4)

The bit allocation optimality transitions are equivalent for U_s vs. D_s, and for U_p vs. D_p.

5.3.1: derivation of b_i, from formula in Part 2.4

For β = 0 and β = 1, the derivations are trivial. For β = 2, $b_4 = b_3 * n^{-1/(n-1)}$, which leads to $b_2 = \left(\frac{b_1^n - b_3^n}{n(b_1-b_3)}\right)^{1/(n-1)}$, which leads to an expression for b_3 that is intractable in general but is exactly solvable n = 2.

$$b_3 = \left(\frac{\left(\frac{b_1^n - b_3^n}{n(b_1 - b_3)}\right)^{n/(n-1)} - b_3^n * n^{-n/(n-1)}}{n\left(\left(\frac{b_1^n - b_3^n}{n(b_1 - b_3)}\right)^{1/(n-1)} - b_3 * n^{-1/(n-1)}\right)} \right)^{1/(n-1)}$$

$$b_3 = \frac{1}{4} * \frac{\left(\frac{b_1^2 - b_3^2}{(b_1 - b_3)}\right)^2 - b_3^2}{\left(\frac{b_1^2 - b_3^2}{(b_1 - b_3)} - b_3\right)} = \frac{1}{4} * \frac{(b_1 + b_3)^2 - b_3^2}{(b_1 + b_3 - b_3)} = \frac{b_1 + 2b_3}{4} \rightarrow b_3 = b_1/2$$

Part 6: proof that the percept boundary placement that maximizes evolutionary utility differs from those which either maximize percept-item mutual information or minimize utility estimation error

From Part 2.1, the 1-bit interface's utility is optimized when $b_2=b_1/n^{1/(n-1)}$.

Meanwhile, the following expression serves as a generalized mean error function:

$$p(L) \int_0^{b_2} p_{x|L} f(x, b_2/2) dx + p(H) \int_{b_2}^{b_1} p_{x|H} f(x, (b_1 + b_2)/2) dx.$$

In the absolute and squared error cases, f(a,b) = |a-b| and $(a-b)^2$, respectively. Also, if one is considering E_P (E_C), the utility estimate error of items perceived (chosen), the respective values of p(L) and p(H) are b_2/b_1 and $1-b_2/b_1$ ($(b_2/b_1)^n$ and $((b_1)^n-(b_2)^n)/(b_1)^n$). Finally, $p_{x|L} = 1/b_2$ and $p_{x|H} = 1/(b_1-b_2)$.

It is clear by inspection that maximizing E_P with respect to b_2 is not equivalent to maximizing utility because only the latter depends on n. Below it is confirmed that the maxima of utility and E_C are not equivalent either.

$$E_C\text{, squared} = \left(\frac{b_2}{b_1}\right)^n \int_0^{b_2} \left(x - \frac{b_2}{2}\right)^2 / b_2 dx + \left(\frac{b_1^n - b_2^n}{b_1^n}\right) \int_{b_2}^{b_1} \left(x - \frac{b_1 + b_2}{2}\right)^2 / (b_1 - b_2) dx$$

Setting $dE_C/db_2 = 0$ and simplifying yields

$$b_1*(1+n)*(b_2)^{n+1}+(b_2)^2*(b_1)^n-b_2*(b_1)^{n+1}-n*(b_2)^n*(b_1)^2/2 = 0$$

This does not afford an analytic solution for all n; it does not reduce to $b_2 = b_1/n^{1/(n-1)}$. (Illustratively, for the case n = 2, it reduces to $b_2 = b_1/3^{1/2}$.)

$$E_C\text{, absolute} = \left(\frac{b_2}{b_1}\right)^n * \frac{1}{b} * \left(\int_0^{\frac{b_2}{2}} \left(\frac{b_2}{2} - x\right) dx + \int_{\frac{b_2}{2}}^{b_2} \left(x - \frac{b_2}{2}\right) dx\right)$$
$$+ \left(\frac{b_1^n - b_2^n}{b_1^n}\right) * \frac{1}{b_1 - b_2} * \left(\int_{b_2}^{\frac{b_1 + b_2}{2}} \left(\frac{b_1 + b_2}{2} - x\right) dx + \int_{(b_1 + b_2)/2}^{b_1} (x - \frac{b_1 + b_2}{2}) dx\right)$$

Setting $dE_C/db_2 = 0$ and simplifying yields $2*(n+1)*(b_2)^{n+1}-n*b_1*(b_2)^n-b_2*(b_1)^n = 0$.

This does not afford an analytic solution for all n; it does not reduce to $b_2 = b_1/n^{1/(n-1)}$. (Illustratively, for the case n = 2, it reduces to $b_2 = (1+7^{1/2})*b_1/6$.)

Part 7: proof that the percept boundaries that minimize estimation error across all percepts correspond to an optimum of the distortion loss and utility objectives when, within an attribute interval $[0,b_1]$, 1) attribute probabilities are uniform 2) utility varies linearly with attribute value 3) the max-utility attribute value is random 4) choices are binary

Let point p ($p \in [0,b_1]$) emerge as the randomly-determined peak utility attribute value; denote the corresponding distortion and utility functions as D_p and U_p, respectively.

The objective function corresponding to the interface's overall utility: $OU = \int_0^{b_1} U_p \, dx$. To optimize the boundary placement of a 1-bit interface, set $dOU/db_2 = 0$ and solve.

$$\frac{dOU}{db_2} = \int_0^{b_1/2} \frac{dU_p}{db_2} dx + \int_{b_1/2}^{b_1} \frac{dU_p}{db_2} dx + \frac{dU_{.5*b_1}}{db_2}$$

Estimation error is minimized when $b_2=b_1/2$; fortuitously, at this value of b_2, the integral terms are additive inverses and $dU_{.5*b1}/db_2 = 0$. Moreover, this result is independent of n (the number of item choices presented).

To construct an optimal 2-bit interface, it is instructive to assume that $b_1/2$ is the optimal placement for b_3 and to define OU_1 and OU_2 such that $U_p \in OU_1$ (OU_2) when $p \in [0,b_1/2]$ ($p \in [b_1/2,b_1]$). Within the interval $[b_3,b_1]$,

$$\frac{dOU_2}{db_2} = \int_{b_3}^{(b_1+b_3)/2} \frac{dU_p}{db_2} dx + \int_{(b_1+b_3)/2}^{b_1} \frac{dU_p}{db_2} dx + \frac{dU_{(b_1+b_3)/2}}{db_2}$$

By analogy with the 1-bit interface, the optimal placement of b_2 with respect to OU_2 is midway between b_1 and b_3. This analogy extends to β-bit interfaces as follows: the optimal placement of b_i with respect to OU_i, $U_p \in OU_i$ when $p \in [b_{i-1},b_{i+1}]$, is $(b_{i-1}+b_{i+1})/2$.

Meanwhile, within the interval $[b_3,b_1]$, all $U_p \in OU_1$ reduce to c-x (c constant); the division of these linear utility gradients is therefore equivalent and as described in Part 2. Thus, the placement of b_2 at $(b_3+b_1)/2$ is optimal uniquely when n = 2. This result extends to β-bit interfaces as follows: with n = 2, the optimal placement of b_i with respect to OU_x, $U_p \in OU_x$ when $p \neg\in [b_{i-1},b_{i+1}]$, is $(b_{i-1}+b_{i+1})/2$.

Finally, the above expressions continue to hold true if D_p is substituted for U_p, and so do the corresponding conclusions.