Beauty requires thought

Aenne A. Brielmann[1], **Denis G. Pelli**[1,2,*]

[1]Department of Psychology, New York University, New York, NY, 10003, USA

[2]Center for Neuroscience, New York University, New York, NY, 10003, USA

Summary

The experience of beauty is a pleasure, but common sense and philosophy suggest that feeling beauty differs from sensuous pleasures such as eating or sex. Immanuel Kant [1,2] claimed that experiencing beauty requires thought but that sensuous pleasure can be enjoyed without thought and cannot be beautiful. These venerable hypotheses persist in models of aesthetic processing [3–7], but have never been tested. Here participants continuously rated the pleasure felt from a nominally beautiful or non-beautiful stimulus, then judged whether they had experienced beauty. The stimuli, which engage various senses, included seeing images, tasting candy, and touching a teddy bear. The observer reported the feelings that the stimulus provoked. The time course of pleasure, across stimuli, is well-fit by a model with one free parameter, pleasure amplitude. Pleasure amplitude increases linearly with the feeling of beauty. To test Kant's claim of a need for thought, we reduce cognitive capacity by adding a "2-back" task. This added task greatly reduces the beauty and pleasure experienced from stimuli that otherwise produce strong pleasure, and spares that of less-pleasant stimuli. We also find that strong pleasure is always beautiful, whether produced reliably by beautiful stimuli, or just occasionally by sensuous stimuli. In sum, we confirm Kant's claim that only the pleasure associated with feeling beauty requires thought and disconfirm his claim that sensuous pleasures cannot be beautiful.

Results

Philosophers and scientists agree that the experience of beauty is a kind of pleasure [1, 2 (pp. 96), 3–4, 8–11]. Kant, still the preeminent authority on beauty [12], claimed that beauty requires thought, and that sensuous pleasures do not [1]. He later proposed that beauty stems from a harmonious interplay between the faculties of sensation and understanding [2 (pp. 102), 7]. Experimental findings [13,14] and current models of aesthetic appreciation [3–5] favor this interactionist view.

We here ask whether the experience of beauty is distinguished from other pleasures by needing thought [5–7]. Philosophers agree that there is an experience common to people

who are "feeling beauty", but disagree on how to define it [2,8–9]. Neither wishing to encumber our participants with philosophical baggage, nor to spoil the test by revealing our hypothesis, we left "beauty" undefined and simply asked the participant at the end of each trial: "During this trial, did you get the feeling of beauty from the object?" We used various stimuli: seeing a plain or beautiful image; sucking a candy; or a touching a teddy bear (see Figure 1A).

Beautiful images were of two kinds. First, before coming to the lab, each participant was asked to choose six images from the internet that are "movingly beautiful" to him or her. Second, we selected images from the international affective picture set (IAPS) with very high valence ratings (i.e., rated "completely happy" [15]) as unfamiliar stimuli likely to elicit stronger beauty responses in most participants. We used valence as a proxy for beauty because the IAPS database does not include beauty ratings. As a shorthand, we will refer to the self-selected beautiful and the high-valence-IAPS images as nominally *beautiful images*, and to the rest of the stimuli as nominally *non-beautiful*, including *plain images* (mid-valence IAPS and IKEA), and *sensuous pleasures* (candy and teddy bear). We obtained several images of each kind, had several flavors of candy, and several teddy bears with different wool textures. Self-selected images were seen only by the participant who selected them. No participant saw any image more than once. Each participant did two trials with each stimulus kind, one with and one without an added task.

In each trial without added task (explained below), participants usually (57/62 = 92%) gave the highest rating on our 4-point beauty scale ("definitely" beautiful, scored as 3) to their self-selected images and in more than half of trials without task (35/62 = 56%) saw high-valence IAPS pictures that way too. Roughly a third of participants "definitely" experienced beauty from non-visual stimuli in trials without added task, i.e., from sucking a candy (23/62 = 37%) or touching a teddy bear (6/20 = 30%). They hardly ever (3%) claimed to definitely experience beauty from neutral images and seldom (19%) for IAPS images rated moderately high in valence, i.e., neither happy nor unhappy (see also Table 1 and Figure S4).

To test the claim that beauty requires thought, we manipulated cognitive capacity by adding a secondary task to reduce the capacity of participants' executive functions, including attention and working memory. This added task was unrelated to the stimuli presented, thus serving to diminish cognitive resources devoted to the stimulus. On each trial, participants rated the pleasure and beauty they experienced from a stimulus. There were two trials for each stimulus: one with the added task and one without. The order of all trials was randomized for each participant. Apart from stimuli, our experiments (1A, 1B, and 2) differed only in the added task. In each trial, we asked the participant to continuously rate the pleasure he or she feels from the stimulus during and after presentation, for a total of 90 s. Participants indicated the amount of pleasure they felt by adjusting the spread between their index and middle finger on an iPad. At the end of each trial, they were asked: "During this trial, did you get the feeling of beauty from the object?": definitely not (0), perhaps not (1), perhaps yes (2), definitely yes (3). (Observers were unaware of the numerical encoding.) The finger spread and beauty judgement were recorded by our web app emotiontracker.com [16].

As previous work has shown [16], continuous pleasure ratings are well fit by a simple model, refined here (Eqs. 1–3 and Figure 1B). The model supposes a stable initial response level $r_{initial}$. After stimulus onset, pleasure asymptotically approaches the steady-state r_{steady}. After stimulus offset, pleasure asymptotically approaches the final response level r_{final}. The initial asymptotic approach, beginning at stimulus onset, is a decaying exponential with time constant τ_{short}. The second asymptotic approach, beginning at stimulus offset, is a weighted sum of two decaying exponentials with time constants τ_{short} and τ_{long}:

$$\widehat{R} = \alpha_{on}(t)r_{initial} + \left(1 - \alpha_{on}(t)\right)\left(\alpha_{off}(t)r_{steady} + \left(1 - \alpha_{off}(t)\right)r_{final}\right) \quad (1)$$

$$\alpha_{on}(t) = \exp\left(\frac{-t - t_{on}}{\tau_{short}}\right) \quad (2)$$

$$\alpha_{off}(t) = w_{short}\exp\left(\frac{-t - t_{off}}{\tau_{short}}\right) + \left(1 - w_{short}\right)\exp\left(\frac{-t - t_{off}}{\tau_{long}}\right) \quad (3)$$

where $x = \max(0, x)$ is the "floor" function.

The model has 6 free parameters: $r_{initial}$, r_{steady}, r_{final}, τ_{short}, τ_{long}, w_{short}. The steady-state response r_{steady} is the asymptotic pleasure of the stimulus response. For our initial fit, for each condition (stimulus kind, with or without added task) we averaged across experiments and participants to obtain a curve describing pleasure response over time. We fit all these curves at once, allowing a different r_{steady} for each curve, and a single value for each of the remaining parameters across all curves. The best fit has $r_{initial} = 1.13$; $r_{final} = 1.08$; $\tau_{short} = 3.05$ s; $\tau_{long} = 104$ s; $w_{short} = 0.215$. For our final model, we froze all parameters but r_{steady}. The best fit of this one-parameter model was solved analytically for each trial (Eq. 4 in Experimental Procedures; see Table S1 for fits per participant). The model's excellent fits to the curves capture the effects of condition entirely in the r_{steady} parameter (see Figure 1B and Figure S2A for RMSE and Figure S1 for residuals). Hence, the dynamics of the pleasure response, including the time needed for pleasure to rise to steady state and the duration of its decay, were conserved across conditions. In the following, we refer to r_{steady} as pleasure, and designate it as P for trials without task or P_{task} for trials with task. Pleasure was correlated with the beauty judgment, $r(328) = 0.60$, 95% CI [0.55, 0.65].

Experiment 1 – The 2-back task demands executive function.

In Experiments 1A and 1B, participants performed an auditory 2-back task. This task demands a broad set of executive functions, including working memory and attentional control. It is thus ideally suited to interfere with Kant's postulated "harmonious interplay" of sensation and cognition [2, p.102], since updating and manipulating working memory content seems likely to demand much the same resources as binding dynamic sensory input

together with higher-level cognitions. In Experiment 1A ($N = 20$), participants were asked to press a key when the current letter was the same as the one two letters ago. In Experiment 1B ($N = 20$), they were instead asked to give a verbal response. (This freed a hand, allowing us to add the teddy-bear stimulus.) We did not find meaningful differences between the results of Experiments 1A and 1B, so we merged the results together as *Experiment 1*.

The 2-back task reduced the average beauty felt from beautiful stimuli, but hardly affected that from non-beautiful stimuli. In particular, performing a 2-back task reduced the end-of-trial beauty judgement of beautiful images by 0.5 and 0.6 on our 4-point scale: $d = -0.90$, 95% CI [-1.37, -0.44] for self-selected beautiful, and $d = -0.92$, [-1.39, -0.44], for high-valence IAPS, where d is Cohen's d, the difference in means divided by the pooled standard deviation (https://en.wikiversity.org/wiki/Cohen%27s_d). Average beauty judgments of non-beautiful stimuli were unaffected by the added task (Figure 1D, see Table 1 for raw distributions). In a similar way, the added task also greatly decreased pleasure from self- and pre-selected beautiful images (by 2.1, $d = -0.79$, [-1.26, -0.32], and 1.8, $d = -0.66$, [-1.12, -0.20], see Figure 1C). We know that the added task did not just attenuate the participants' manual response, because the pleasure ratings of non-beautiful stimuli are unaffected by the added task. One control experiment confirmed that accuracy of tracking distance between moving dots on the screen was not changed by adding the 2-back task (see Supplemental Table 3). Another control experiment, in which stimuli were present before the task onset or after the task offset, provided evidence that the pleasure rating is indeed continuous and tracks some internal state (see Supplemental Figure 4). ROC analysis (Figure 1E) confirmed the conclusion, finding a significant effect of task only for the beautiful stimuli.

Having assessed the effect of added task per stimulus kind, we now consider its relation to pleasure and beauty without the added task. Only strong, not weak, pleasures were affected (Figure 2A,B). We find a linear relationship between pleasure and beauty (Figure 2C) that is conserved across stimuli and added tasks. A simple model (Figure 2E) captures the linear relation between mean beauty and pleasure, the non-linear effect of task, and the effect of coarse quantization in the categorical beauty response. It fits all our pleasure and beauty ratings well (RMSE = 0.15 and 0.6), with and without task, with only 10 degrees of freedom: 6 for stimulus strength (one per type), 2 for task effect (P_{beau} and g), and 2 for the linear relation between pleasure and beauty (a, b). The fitted value of the threshold pleasure parameter $P_{\text{beau}} = 4.30$, above which pleasure is attenuated by task (Figure 2A,B), lies between the estimated thresholds $P_1 = 3.7$ and $P_2 = 6.0$ for perhaps and definitely feeling beauty (Figure 1F).

The linear relation between beauty and pleasure supports the claim that beauty is interchangeable with "aesthetic pleasure" [10]. The difference between beautiful and non-beautiful pleasures is that the beautiful pleasures are greater. It may help to draw an analogy to color perception here. In general, when a color's luminance is reduced, it becomes darker (e.g., red becomes dark red). But yellow becomes brown. This series of physical stimuli, varying only in luminance, spans two different color names: yellow and brown. Yellow exists only at high luminance. Similarly, our data indicate that "definitely felt" beauty exists only at high pleasure.

Participants' reports of "definitely" experiencing beauty from non-visual stimuli in trials without added task (37% for candy, 30% for teddy bear) made us wonder whether the occasional beauty of (usually) non-beautiful stimuli might be like that of beautiful stimuli, just less frequent. We gauged the task-susceptibility by the relative frequency of "definitely" feeling beauty with and without added task. The ratio of instances of definitely feeling beauty with over without added task was 27/61=0.44, 95% CI [0.33, 0.57], (N=80) for beautiful stimuli and 19/28=0.68, [0.49, 0.82], (N=140) for non-beautiful stimuli, which are not significantly different. Thus, the experience of beauty is equally susceptible to the 2-back task, regardless of the nominal beauty of the stimulus.

Experiment 2 — Digit-span task demands just short-term memory.

To further explore which cognitive capacities are needed to experience beauty, participants in Experiment 2 ($N = 22$) were given a string of digits to remember at the beginning of the trial and were asked to report it only at the end of the trial. Unlike Experiment 1, this task merely required retention. A three-way interaction of experiment × task × stimulus kind indicated that the pattern of effects differed for the two experiments, $F(4, 240) = 3.28$, $p = 0.012$. Remembering the digit string slightly decreased beauty across stimuli, $F(1, 84) = 4.38$, $p = 0.049$, $d = 0.22$, 95% CI [–0.05, 0.49], but had no stimulus-specific effect (see Figure 3B and Table 1 for raw distributions), $F(1, 84) = 0.40$, $p = 0.810$. It did not affect participants' overall pleasure, $F(1, 84) = 1.15$, $p = 0.296$, nor the pleasure from a particular stimulus (see Figure 3A), $F(4, 84) = 0.71$, $p = 0.588$.

Discussion

Beauty does & sensuous pleasures do not require thought—Kant confirmed.

We find that broadly restricting executive function decreases the pleasure taken from beautiful, and not from non-beautiful, stimuli (Experiment 1, Figures 1–2). Only nominally beautiful stimuli produce the high average pleasure (exceeding P_{beau}) that is susceptible to reduction by a secondary task (Figures 1C,D and 2A,B). This confirms Kant's claim that the experience of beauty requires thought and that non-beautiful pleasures do not.

Beauty and pleasure were spared by the digit-span task, which was much less demanding, impairing only memory capacity, with little attentional and cognitive load (Experiment 2, Figure 3). The much higher cognitive and attentional load of the 2-back task may account for its greater impairment of beauty and pleasure.

A Reviewer (Stephen Palmer) suggested an alternate account. Suppose, when both stimulus and task are attended, that observers cannot distinguish the pleasures of stimulus and task, and can report only their average. Then, the effect of a secondary task on pleasure is unrelated to its degree of cognitive engagement and depends only on the task's pleasure. To test this, in an additional experiment, participants (N=7) rated the pleasure of doing just the task for 90 s. We measured pleasure ($P = r_{steady}$) for the three tasks: 2-back manual (2.72), 2-back verbal (2.42), and digit span (1.91). Since the digit-span task is the least pleasant, the compulsory-averaging model predicts more pleasure reduction by the digit-span than by the

2-back task, whereas, in fact, the digit-span task hardly affects reported pleasure. This rejects the compulsory-averaging account.

Future experiments might further reveal the cognitive requirements of feeling beauty. Any account of the reduction of beauty by the 2-back that is based on limiting a cognitive resource—whether by loading executive function, cognitive effort, or attention—builds on our main finding that beauty requires thought.

One could imagine that beauty-related thinking increases over the stimulus duration, predicting greater task effects later in the stimulus presentation. The excellent fit of our model (with fixed time constants) to data with and without task rejects this hypothesis. Or one might suppose that novel stimuli require more beauty-related thought than familiar ones, predicting greater task effect on novel stimuli. Our model fits equally well for familiar (self-selected) and unfamiliar (high-valence IAPS) stimuli, indicating conservation of beauty-related thinking independent of familiarity.

The beauty of sucking candy—Kant disconfirmed.

Our evidence rejects Kant's claim that sensuous pleasures cannot be beautiful. The same definitely-beautiful rating can arise from our entire range of stimuli. Beautiful stimuli are just more likely to produce it (90% for self-selected images vs. 3% for IKEA furniture). Surprised that sucking candy could be "beautiful," we queried some participants who had responded "definitely yes" for beauty on candy trials. Most of them remarked that sucking candy had personal meaning for them, like a fond childhood memory. One participant replied, "Of course, anything can be beautiful." The occasional reports of beauty from sensuous stimuli cannot be attributed to a subset of participants who give beauty ratings promiscuously. First, reports of beauty ($B = 3$) for IKEA furniture were very rare (2/62= 3% across all experiments). Second, across observers, the distribution of beauty ratings for each stimulus kind without added task was unimodal (see Table 1 and Supplemental Figure 4), failing to give evidence for a subpopulation with a higher mode. Finding that beautiful things yield higher pleasure than sensual stimuli challenges widespread claims that the highest pleasures are sensual, especially "sex, drugs, and chocolate" [17]. Our findings encourage extending the scope of current models of beauty beyond images and music to include sensuous pleasures [18,19].

Conclusion.

Beauty matters. Homer's *Iliad* recounts a war over a woman's beauty [20]. The cosmetics industry sold $460 billion worldwide in 2014 [21]. Mathematicians consider beauty essential to their craft: "Beauty is the first test: There is no permanent place in the world for ugly mathematics." [22] Having tested Kant's claims about beauty [1,2], we draw several conclusions about beauty. First, the feeling of beauty increases linearly with pleasure, and strong pleasure is always beautiful, whether produced reliably by beautiful stimuli, or just occasionally by sensuous stimuli. Second, impairing thought attenuates the excess of pleasure over a threshold P_{beau}. Third, only beautiful stimuli produce average pleasure strong enough (exceeding P_{beau}) to be susceptible to reduction by a secondary task. In sum, we confirm Kant's claim that only the pleasure associated with feeling beauty requires

thought and disconfirm his claim that sensuous pleasures can never be beautiful. We thus demonstrate that psychological experiments can test philosophical theories, [6] and mathematical models can describe aesthetic experiences.

STAR Methods

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Denis Pelli (denis.pelli@nyu.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

62 naive participants (20 in Experiment 1A, 20 in 1B, 22 in 2) were recruited from the immediate environment of New York University. We did not record the gender of our participants because we had no hypotheses relating to this information and collected only necessary personal information in line with the ethics board guidelines. All but one participant (aged 17) were adults and all (plus one of the minor's parents) gave written informed consent according to the declaration of Helsinki. Participants received either $10 or course credit as compensation.

METHOD DETAILS

Stimuli.—Participants were exposed to stimuli of six kinds to explore beauty and pleasure. Images were presented on a 21.5" iMac display, scaled to the full width of the 1920✕1080 pixel screen, and were viewed from a distance of approximately 1 m. The luminance of the walls of the testing room was about 40 cd/m^2. When blank, the screen luminance was 340 cd/m^2. Images were dimmer, e.g. mid-valence IAPS image #7160 (a dark abstract pattern) had a luminance of 108±68 cd/m^2 (mean±sd), sampled with a spot photometer. These screen and wall luminances are in the typical range for computers in offices, so luminance *per se* would not be expected to affect arousal. A range of images was provided to increase the likelihood that, on different trials, participants would experience beauty, mild pleasure, and neutral feelings. First, each participant was asked to provide four to six images that were "movingly beautiful" to him or her (for examples see Figure 1a and Figure S4). We refer to them as *self-selected images*. Second, we selected six images from the International Affective Picture System (IAPS [15]) that had extremely positive valence (7–8 on a scale from 1–8) and elevated arousal (5–6 on a scale from 1–8) ratings (*high-valence IAPS*; picture numbers: 1710, 5600, 5621, 5833, 7330, 7508). Results show that the self-selected are the most beautiful. By also including the high-valence IAPS—which they had never seen before—we produced beauty with and without familiarity. As the IAPS does not provide beauty ratings, and as there was to our knowledge no other image data base with normative beauty ratings available at that time, we selected images based on the most closely related available rating, i.e., valence (also anchored with the terms pleasant and unpleasant). Third, we chose images from the IAPS database that had positive valence (5–7) and medium arousal (3–4) ratings (*mid-valence IAPS*; picture numbers: 1947, 7281, 7545, 7160, 5711, 7340). These images are more ambiguous and susceptible to various interpretations and personal associations (see Figure S2). Fourth, images of IKEA furniture against a white background, from their online catalog, were presented as *neutral images* (www.ikea.com).

Fifth, we let participants taste different flavors of hard *candy* (The Hershey Company's Jolly Rancher, https://en.wikipedia.org/wiki/Jolly_Rancher) to provide a pleasurable experience in another sense modality. Jolly Rancher hard candy is popular in the USA and comes in 6 fruit-like flavors: green apple, grape, cherry, blue raspberry, watermelon, and fruit punch. The flavor is a bit sweeter and more intense than that of comparable European candies, such as "nimm2". Participants were asked to unpack the candy and to prepare a cup of water before each candy trial. At trial onset, they put the candy in their mouth and sucked on it for 30 s, after which they were told to spit the candy into a second cup and take a sip of water. To provide yet another stimulus modality, Experiment 1B included a tactile stimulus: touching a very soft alpaca *teddy bear* ("Tiny Teddy Bears" from Peruvian Link Alpaca Collection, made from 100% baby alpaca, http://www.peruvianlink.com/product-p/16401.htm). Participants were asked to reach their non-dominant hand into a plain cotton pillowcase "to touch the teddy bear," for a non-visual tactile pleasure.

We deliberately incorporated diverse kinds of stimuli in our experiments, including several modalities and ranges of evoked response. Our interest in the current study is not in what makes an object beautiful. That is subject to taste. Rather, our interest is in the apparently universal experience of beauty. If we each look at our thing of beauty, what do our experiences have in common? A Q-mode principal components analysis (PCA) was used to assess whether responses to each stimulus kind are consistent across participants. As opposed to standard (R-mode) PCA, Q-mode evaluates whether there are systematic differences in response between subgroups of participants. Both the scree plot and acceleration factor of Q-mode PCA of beauty ratings across all participants in all experiments — excluding teddy bear trials (20 participants of Experiment 1B) — revealed a one-factor structure which indicates that our stimulus kinds elicited consistent responses across observers. The rating differences between observers were dominated by differences between stimulus kinds.

Procedures.—Pleasure ratings were obtained through the web-based app emotiontracker.com on an iPad [16]. Participants were asked to place the index and middle finger of their dominant hand on the iPad screen and continuously rate the amount of pleasure they were experiencing by adjusting the spread of their two fingers. (Outside our lab, people typically use their thumb and index finger for two-finger gestures on trackpads, but in preliminary studies we found that the index and middle fingers are better for our rating task because they have a more consistent resting state—corresponding to no effort and no pleasure—which results in a more stable baseline rating.) This method of measuring pleasure has been shown to be tightly linked to physiological responses, i.e. facial muscle [25]. We calibrated each participant's manual response scale before the first trial of the experiment. Participants were asked to indicate "maximum pleasure" by spreading their two fingers as far apart as could be comfortably maintained, and "minimum pleasure" by relaxing their fingers to whatever spread felt most natural and required least effort. Subsequent finger spread measurements were linearly mapped to this pleasure scale, from "minimum" (1) to "maximum" (10). Participants were asked to continuously rate the pleasure they were experiencing from the stimulus at each moment from the beginning of stimulus presentation until the app announced that the trial was "done" (60 s after the

stimulus offset). The finger spread, sampled at 1 Hz, was linearly converted to a pleasure rating by reference to the participant's initial calibration settings of spread for minimum (1) and maximum (10) pleasure. We collected about 90 pleasure ratings over the 90 s duration of each trial.

After the pleasure rating, the app asked participants "During this trial, did you get the feeling of beauty from the object?", referring to the object which had been present during the first 30 s of the trial. They could choose between the answers "Yes, definitely" (scored as 3), "Perhaps yes" (2), "Perhaps no" (1) and "Definitely no" (0). We ask participants to judge whether they "felt beauty" without defining beauty. We refrained from providing participants with a Kantian definition of beauty, as it includes the requirement of thought, and thus our hypothesis, which we did not want to reveal to participants. Definitions of beauty differ widely in both philosophy and experimental aesthetics [2–10], but these scholars are not trying to select one among several competing candidate phenomena for "beauty." The definitions generally agree that there is a single well-known basic perceptual phenomenon, but differ in how they formally describe it. These technical definitions are confusing to naive participants, so we operationally define the experience of beauty as the participant judgements given in response to our "feeling of beauty" question.

At the end of each trial, the emotiontracker.com app emails to the experimenter an Excel spreadsheet with the calibrations, the 90 s record of finger spread sampled at 1 Hz, and the final beauty judgement. Our measures of pleasure and beauty are both recorded in each trial from the participant, but one is a dynamic analog finger-spread rating of pleasure over time throughout the trial, while the other is a 4-level beauty rating at the end. The differences between these methods was intended to keep the measures independent. It turns out that beauty and pleasure are correlated, which implies a common cause or that one influences the other.

Participants in each experiment did two trials for each stimulus kind, one with and one without an added task that limited the availability of cognitive resources. Within each experiment, there was one trial per condition (stimulus kind ✕ with or without added task), in random order. The experiment required sustained attention, so we kept it under an hour. Recall that we collected 90 pleasure ratings (1 per s) during each trial, so that one trial per condition and participant and 20 participants per experiment were enough to assess our hypotheses.

In Experiments 1A and 1B, participants performed an auditory 2-back task. This task demands a broad set of executive functions, including working memory and attentional control. It is thus ideally suited to interfere with Kant's postulated "harmonious interplay" of sensation and cognition [2, p.102], since updating and manipulating working memory content seems likely to demand the same resources as continually binding and updating sensory input with higher-level cognitions. The 2-back task is frequently used as a secondary task in psychological experiments, often to measure executive function in general, including working memory and attention control [26]. Participants had to decide whether each spoken letter in a sequence matches the letter they heard two letters ago. Experiments 1A and 1B differed solely in response modality. Participants responded to each match by pressing a key

on a keyboard (Experiment 1A) or verbally (Experiment 1B). Using a verbal response in Experiment 1B freed a hand, allowing us to introduce the teddy bear tactile stimulus. The free software BrainWorkshop (http://brainworkshop.sourceforge.net/ [27]) was used to present stimuli and determine the proportion of correct trials. Further details of the N-back are described elsewhere [28].

In Experiment 2 (*N*=22), the added task was a simple number-span memory task. On each of these trials, the experimenter first read a random sequence of digits to the participant. Participants had to remember these digits through the whole trial and repeat them at the end of the trial, after they finished the pleasure and beauty judgments. This is like remembering a long phone number. Their answers were recorded on paper by the experimenter. To determine the participants' digit span individually, participants did 14 digit-span trials prior to the experiment. Testing began with a list of three digits, which increased or decreased, according to a 1-up 2-down staircase, i.e., the subsequent list length was increased by one digit following a correct response and decreased by one digit after two incorrect responses [28]. The highest number of digits presented to them (i.e. the smallest number of digits that they were consistently unable to remember correctly) was used throughout the experiment.

In all experiments, participants were told that it was important to do their best. For the 2-back task, participants needed to correctly detect 50% of 2-back trials. When participants did not reach that goal, another trial of the same condition (stimulus kind, with or without added task) was appended at the end of the experiment. For the digit-span task, we always chose a length beyond their ability, so we merely exhorted them to perform as well as they could, and used all the trials regardless of accuracy of recall.

Upon arriving at the laboratory, participants were asked to provide informed consent. Subsequently, the experiment started with either 2-back practice trials (Experiments 1A and 1B) or the number-span staircase (Experiment 2). For the 2-back tasks, participants were required to complete two continuous 2-back sequences of 88 seconds with an accuracy of at least 50% before proceeding to the main experiment. Following this, the experimenter explained the pleasure-rating procedure and handed over the iPad with the EmotionTracker.com app running. There was one practice trial for pleasure ratings per stimulus modality. Participants practiced by rating the pleasure they experienced from a mid-valence IAPS image, a teddy bear, and a candy. The same image, teddy bear, and candy flavor were not used for the actual experiment later. (We have several flavors of candy and slightly different models of baby alpaca teddy bear.) After the practice trials, the main data collection began.

Stimuli were presented in a pre-randomized sequence of conditions (stimulus kind with or without added task). Each stimulus was presented once without, and once with, the added cognitive task. Images were presented to participants on a 21.5" iMac display using PowerPoint. For trials not involving pictures, a blank screen was presented instead of an image. In image trials, participants were instructed to start rating their pleasure as soon as the image appeared on the screen, and stop rating when the app said "Done!". For the other trials, they were instructed to simultaneously start rating and sucking the candy (or touching the teddy bear) as soon as the fixation cross on the screen disappeared. Each stimulus was

presented for 30 seconds. (Post hoc analysis revealed that the effective duration of the candy stimulus was 38 s.) After that, the images were replaced by a fixation cross, or the participant was asked to spit out the candy and take a sip of water, or to stop touching the teddy bear and take his or her hand out of the pillowcase, respectively. As mentioned above, when participants failed to reach the 2-back performance criterion of > 50% on a given trial, or if any disruptions occurred during the trial, a trial of the same condition was appended at the end of the experiment. Finally, participants were thanked and — if they were not participating for course credit — paid.

In analysis, we found that all the pleasure curves dropped steeply at stimulus offset, except the candy curves, which dropped 8 s later than the nominal offset time, which is when we asked the participant to spit out the candy and take a sip of water. Since spitting and sipping take time, we made the post hoc correction of estimating the effective offset time of candy as 8 s after the nominal offset time, and used that effective offset time for candy in all figures and analyses.

QUANTIFICATION AND STATISTICAL ANALYSIS

Data processing and analyses were conducted with MATLAB (version R2015a), SPSS (version 20), and R (version 3.2.3). We fit the average (across participants) curves (pleasure vs. time) per condition (separate averages per stimulus kind and per beauty category) using the model developed in our lab[28] and specified in Eqs. 1–3.

We used MATLAB's fminsearch function to find optimal parameter values. As initial values for fitting, we used $r_{initial} = 1$, $r_{steady} = 5$, $r_{final} = 1$, $\tau_{short} = 1$ s, $\tau_{long} = 60$ s, and $w_{short} = 0.5$. The time for stimulus offset was generally defined as 30 s after trial start, but was set to 38 s for candy trials, as spitting out the candy and rinsing its taste away with water required additional time. The full model was fit, simultaneously, to 12 average pleasure curves, one per condition (6 stimulus kinds ✗ with and without added task). The parameters $r_{initial}$, r_{final}, τ_{short}, τ_{long}, and w_{short} were optimized independent of condition (stimulus kind, and presence of added task), while r_{steady} was allowed a different value for each condition. In all subsequent fitting, only r_{steady} was allowed to vary and the values of the rest of the parameters were locked to their values in the initial fit ($r_{initial} = 1.13$; $r_{final} = 1.01$; $\tau_{short} = 3.05$; $\tau_{long} = 104$; $w_{short} = 0.215$). Fits for each participant are shown in Table S1. This reduced model has only one degree of freedom, r_{steady}. Eq. 4 is the analytic solution for the best fit of the reduced model. For each trial response $R(t)$, the best RMSE fit $\hat{R}(t)$ by the single-parameter model has steady-state response

$$r_{steady} = \frac{\sum_t (R(t) - f(t))g(t)}{\sum_t g^2(t)} \quad (4)$$

where

$$f(t) = \alpha_{on}(t)r_{initial} + (1 - \alpha_{on}(t))(1 - \alpha_{off}(t))r_{final} \quad (5)$$

and

$$g(t) = \left(1 - \alpha_{\text{on}}(t)\right)\alpha_{\text{off}}(t) \quad (6)$$

Eqs. 5 and 6 are plotted in Figure S4. These trial-wise r_{steady} values (Eq. 4) were used for all further analyses of pleasure.

To assess differences in parameter estimates for r_{steady}, repeated measures ANOVAs were conducted. As a measure of effect size for significant differences between conditions, Cohen's d with Hedge's correction was calculated.

DATA AND SOFTWARE AVAILABILITY

All data and main analyses files for this article are accessible on GitHub: https://github.com/aenneb/beautyRequiresThought.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

1. Kant I (1764/2004). Observations on the Feeling of the Beautiful and Sublime. (Berkeley: University of California Press).

2. Kant I (1790/2000). Critique of the Power of Judgment. (New York: Cambridge University Press).

3. Armstrong T and Detweiler-Bedell B (2008). Beauty as an emotion: The exhilarating prospect of mastering a challenging world. Rev. Gen. Psychol 12(4), 305–329

4. Graf LKM and Landwehr JR (2015). A dual-process perspective on fluency-based aesthetics: The pleasure-interest model of aesthetic liking. Personal. Soc. Psychol. Rev 19, 395–410.

5. Leder H (2013). Next steps in neuroaesthetics: Which processes and processing stages to study? Psychology of Aesthetics, Creativity, and the Arts 7, 27–37.

6. Chatterjee A Neuroaesthetics: A coming of age story. (2011). J. Cogn. Neurosci 23, 53–62. [PubMed: 20175677]

7. Conway BR and Rehding A (2013). Neuroaesthetics and the Trouble with Beauty. PLoS Biol. 11, 1–5.

8. Santayana G (1896). The sense of beauty: Being the outline of aesthetic theory. (New York: Charles Scribner's sons).

9. Wittgenstein L (1966). Lectures and Conversations on Aesthetics, Psychology, and Religious Belief. (Berkeley: University of California Press).

10. Reber R, Schwarz N, and Winkielman P (2004). Processing Fluency and Aesthetic Pleasure: Is Beauty in the Perceiver's Processing Experience? Personality and Social Psychology Review, 8, 364–382. [PubMed: 15582859]

11. Brown S, Gao X, Tisdelle L, Eickhoff SB and Liotti M (2011). Naturalizing aesthetics: brain areas for aesthetic appraisal across sensory modalities. Neuroimage 58, 250–8. [PubMed: 21699987]

12. Guyer P (2005) Values of Beauty: Historical Essays in Aesthetics. (Berkeley: Cambridge University Press).

13. Chatterjee A and Vartanian O (2014). Neuroaesthetics. Trends Cogn. Sci 18, 370–375. [PubMed: 24768244]

14. Vessel E. a., Starr GG and Rubin N (2012). The brain on art: intense aesthetic experience activates the default mode network. Front. Hum. Neurosci 6, 1–17. [PubMed: 22279433]

15. Lang PJ, Bradley MM and Cuthbert BN (2008). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. (Technical Report A-8, University of Florida).

16. Brielmann AA, Vale L, and Pelli DG (under review). Beauty at a glance. Experience and pleasure of beauty does not depend on exposure duration. Journal of Vision.

17. Martin P (2008). Sex, Drugs and Chocolate: The Science of Pleasure. (London: Forth Estate).

18. Gallace A, & Spence C (2011). Tactile aesthetics: towards a definition of its characteristics and neural correlates. Social Semiotics, 21(4), 569–589.

19. Plato. in Hippias major. (390 BCE). at <http://www.ellopos.net/elpenor/greek-texts/ancient-greece/plato/plato-hippias-major.asp>

20. Homer. (2004). Iliad of Homer. 1st World Publishing Retrieved from: https://books.google.com/books?id=DcBjCAAAQBAJ&lpg=PA5&ots=HTD59ADKUt&dq=homer's%20illiad&lr&pg=PA5#v=onepage&q&f=false

21. Wood L Research and Markets: Global cosmetics market 2015–2020: Market was $460 billion in 2014 and is estimated to reach $675 billion by 2020. Available at: http://www.businesswire.com/news/home/20150727005524/en/Research-Markets-Global-Cosmetics-Market-2015-2020-Market.

22. Hardy GH (1941). A mathematician's apology. (Cambridge: Cambridge University Press).

23. Green DM, and Swets JA (1966). Signal Detection Theory and Psychophysics. (New York: John Wiley & Sons, Inc.).

24. Lau B MatlabAUC. (2014). at <https://github.com/brian-lau/MatlabAUC>

25. Vale L, Gerger G, Leder H, and Pelli DG (submitted). The time course of experiencing beauty: Comparing perceptual and physiological measures of aesthetic experience. Frontiers in Human Neuroscience.

26. Carter RM, Hofstotter C, Tsuchiya N and Koch C (2003). Working memory and fear conditioning. Proc. Natl. Acad. Sci. U. S. A 100, 1399–1404. [PubMed: 12552137]

27. Jaeggi SM, Buschkuehl M, Perrig WJ and Meier B (2010). The concurrent validity of the N-back task as a working memory measure. Memory 18, 394–412. [PubMed: 20408039]

28. Kane MJ, Conway ARA, Miura TK and Colflesh GJH (2007). Working memory, attention control, and the N-back task: a question of construct validity. J Exp Psychol Learn Mem Cogn 33, 615–622. [PubMed: 17470009]
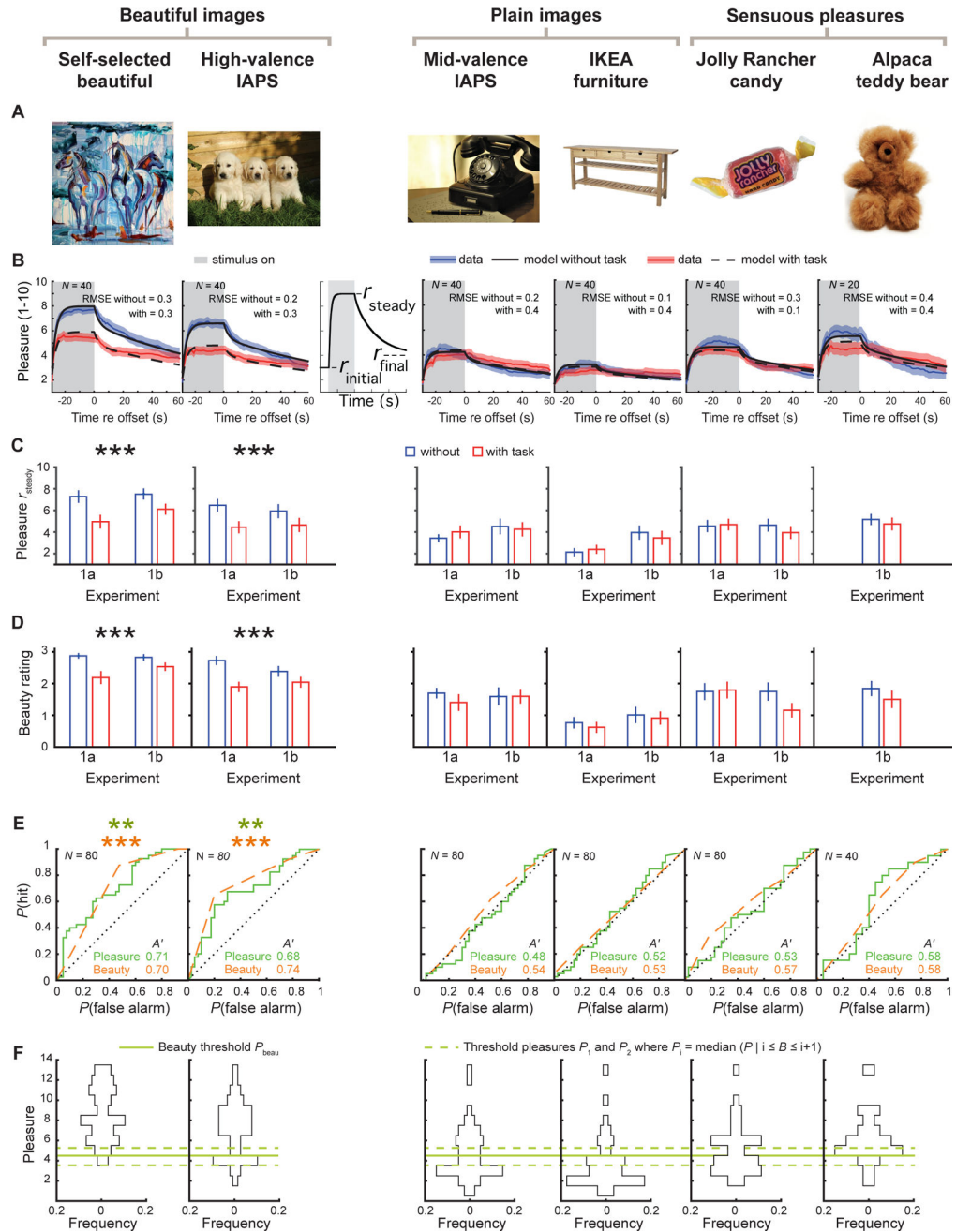
Figure 1. Main results of Experiment 1.

(**A**) Examples of the six kinds of stimulus. (**B**) Time course of the pleasure rating (colored) and model fit (black) in Experiment 1 for each stimulus kind for trials without (blue and solid-black) and with (red and dashed-black) the added task requiring thought. The grey shaded area indicates the stimulus duration. A schematic of the model (Eq. 1) used to fit pleasure ratings is shown in the gap between beautiful images and the remaining stimuli. Colored lines with shaded areas represent mean ± 1 SE of the data. RMSE = root mean square error of the model fit across the entire 90 s trial duration; N = number of trials (and participants) per curve. Experiment 1 combines 1A and 1B, which each had 20 participants,

for a total $N = 40$. (**C**) Mean estimated steady-state response r_{steady} for each kind of stimulus for trials without (blue) and with (red) an added task. Error bars represent mean ± SE. Across Experiments 1A and 1B, average r_{steady} estimates were 2.1 and 1.8 (on the $1 - 10$ pleasure scale) lower in trials with an added task for self-selected beautiful and high-valence IAPS images, both $p < 0.001$, task × stimulus interaction $F(4,152) = 7.99$, $p < 0.001$. r_{steady} pleasure values for all other stimulus categories were unaffected, all p 0.407. There was no three-way interaction, $F(4,152) = 1.86$, $p = 0.120$. Interactions were tested with a $2 \times 2 \times 5$ (task × experiment × stimulus kind) repeated measures ANOVA (rmANOVA) that excluded the teddy bear stimulus only presented in experiment 1B. A separate rmANOVA confirmed that there was no effect of added task on teddy bear trials, $F(1,19) = 0.44$, $p = 0.513$. Asterisks designate significant differences, without vs. with added task, according to post-hoc pairwise comparisons: *** $p < 0.001$. (**D**) Average final beauty judgment in Experiment 1 for each kind of stimulus for trials without (blue) and with (red) an added task. Error bars represent mean ± SE. Despite the three-way interaction of task, stimulus, and experiment, $F(4,152) = 2.64$, $p = 0.031$, separate repeated-measures ANOVAs per stimulus kind showed the same pattern of results as for pleasure r_{steady}: Average final beauty judgments were 0.5 and 0.6 (on our $0 - 3$ scale) lower in trials with an added task for self-selected beautiful and high-valence IAPS images, both $p < 0.001$. Beauty judgments for all other stimulus categories were generally unaffected, all p 0.081. The only interaction of task and experiment was observed for candy, $F(4,152) = 4.49$, $p = 0.041$, all other p 0.098. For candy, the added task decreased beauty judgments in Experiment 1B by 0.6, $p = 0.009$, but the similar 0.5 reduction in Experiment 1A was not significant, $p = 0.819$. Asterisks designate significant differences according to post-hoc pairwise comparisons between trials with and without added task: ** $p < 0.01$, *** $p < 0.001$. (**E**) Receiver-operator-characteristic (ROC) curves [23] for classification of trials as with or without added task based on final beauty judgments (orange) and average pleasure during stimulus exposure (green) in Experiment 1. An area under the curve (A') greater than 0.5 indicates better-than-chance decoding of task presence; asterisks (orange for beauty and green for pleasure) designate significant deviations of A' from 0.5 based on bootstrap tests [24]: ** $p < 0.01$, *** $p < 0.001$. (**F**) Vertical histograms (binned scatter dot plots) of pleasure per stimulus kind for trials without task. The dashed lines indicate the estimated threshold pleasures P_1 and P_2 for perhaps and definitely feeling beauty, where P_i = median $(P | i B i+1)$, calculated as the median, across stimulus kinds, of the median pleasure for trials with beauty rating in the range [1 2] for P_1 or [2 3] for P_2. Between the dashed lines, the green line indicates the threshold pleasure $P_{beau} = 4.29$ for effect of added task (see Figure 2). See also Figure S1–4 and Table S1–2.
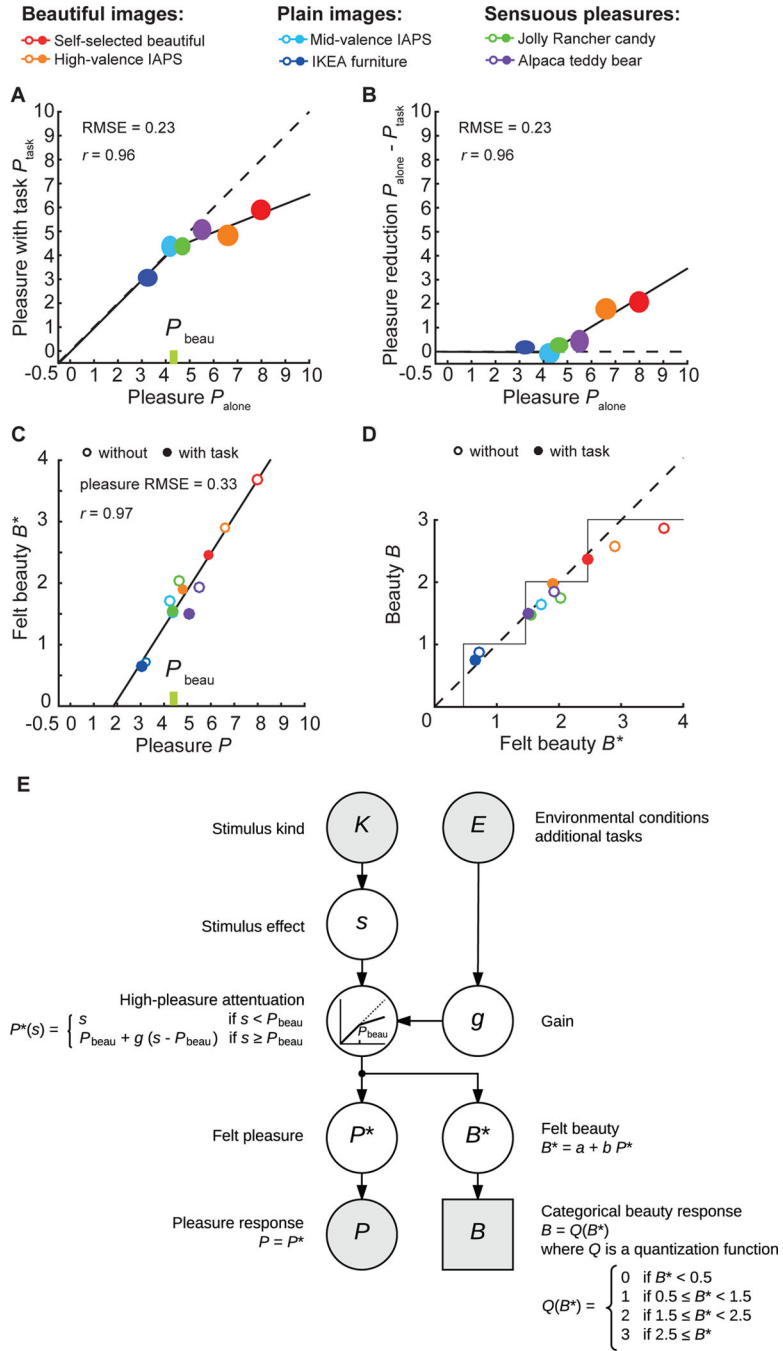
Figure 2. Relation between beauty and pleasure in Experiment 1.

(**A**) Effect of task on pleasure: average pleasure with vs. without added task. The *all-pleasure attenuation model* (not shown) for pleasure is $P_{task} = g\,P_{alone}$, where P_{alone} is pleasure without task, P_{task} is pleasure with task, and g is the task-dependent gain. Though not shown, it fits our data moderately well: 0.62 RMSE with 0.82 gain. However, looking at these data suggests that only high pleasure, exceeding some value P_{beau} (green tick mark), is attenuated. The black line indicates the prediction of this *high-pleasure attenuation model*. This is like retained income after a progressive tax, with a higher tax on income exceeding a

certain level. It predicts $P_{\text{task}} = P_{\text{alone}}$ if $P_{\text{alone}} < P_{\text{beau}}$, and $P_{\text{task}} = P_{\text{beau}} + g\,(P_{\text{alone}} - P_{\text{beau}})$ if $P_{\text{alone}} \geq P_{\text{beau}}$. The high-pleasure attenuation model fit has 0.23 RMSE with a high-pleasure gain of 0.61 and a high-pleasure threshold $P_{\text{beau}} = 4.29$. The RMSE favors the high-pleasure attenuation model. Note that P_{beau} lies between the pleasure thresholds P_1 and P_2 estimated for perhaps and definitely feeling beauty (green between dashed lines in Figure 1F). (**B**) Difference between average pleasure without and with task. The black line represents the predictions of the high-pleasure attenuation model. (**A-B**) Ellipses indicate mean ± SE. (**C**) Predicted felt beauty B^* vs. pleasure for every stimulus kind (color), without (hollow ellipses) and with (filled ellipses) the 2-back task. (**D**) Average beauty B versus felt beauty B^*, for each stimulus kind (color) without (hollow ellipses) and with (filled ellipses) the 2-back task. The observed B response is quantized (0 … 3), but our model assumes that the internal feeling B^* is continuous, with a fixed-variance normal distribution. In this plot, the value of felt beauty is a maximum likelihood estimate of the mean, based on the histogram of quantized beauty responses. The dashed line represents equality. The solid line represents the quantization function $B = Q(B^*)$ (see panel **E**). (**E**) Generative model for *mean* beauty and pleasure responses. Only gray nodes are observed. Round nodes are continuous; square nodes are categorical. Arrows indicate a causal relation. We presume that each kind of stimulus K has a (mean) stimulus effect s. We model the effect of added task by high-pleasure gain g. High-pleasure gain is set to 1 for the (baseline) condition without task. (The high-pleasure gain with task is a degree of freedom.) Felt pleasure and beauty are determined by nonlinear compression of the stimulus effect s by high-pleasure gain g, where $P^* = s$ if $s \leq P_{\text{beau}}$, and $P^* = P_{\text{beau}} + g\,(s - P_{\text{beau}})$ if $s \geq P_{\text{beau}}$, and $B^* = a + bP^*$. The continuous feeling of beauty B^* is quantized to produce the categorical response B to the beauty question at the end of the trial. The pleasure response P (measured as r_{steady}) is the felt pleasure P^*.
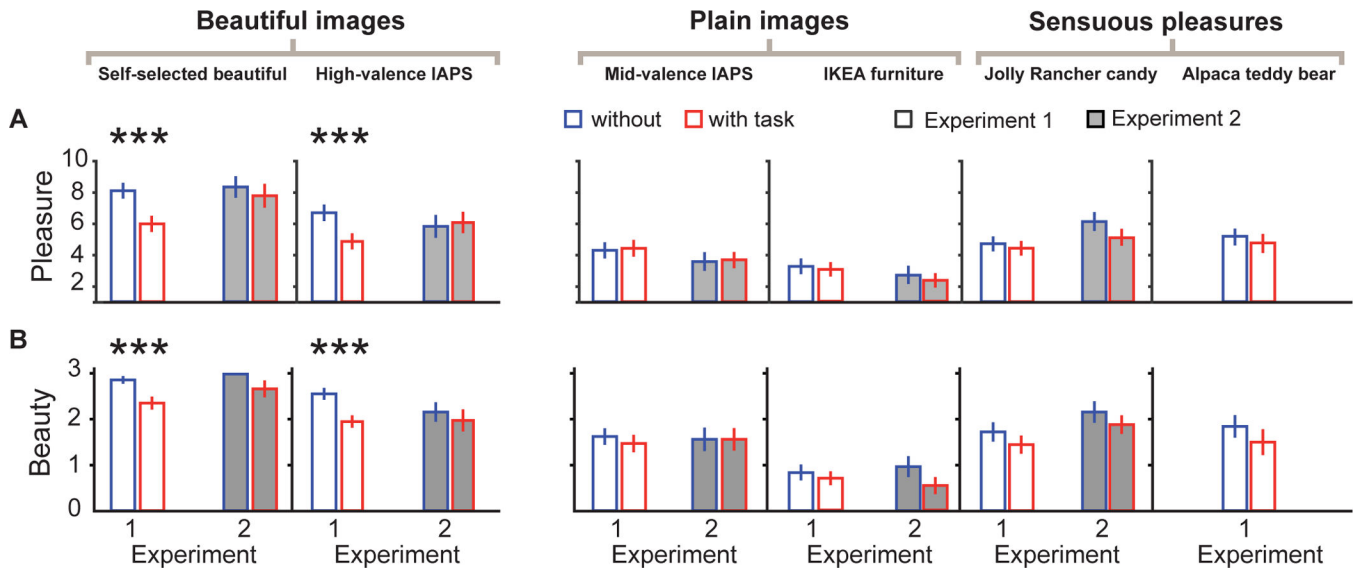
Figure 3. Main results of Experiment 2.
(**A**) Average steady–state pleasure r_{steady} and (**B**) final beauty judgments for each experiment without (blue) and with (red) an added task. The added task was 2-back (Expt 1, left, white bars) or digit-span (Expt 2, right, gray bars). Error bars represent mean ± 1 SE. Asterisks designate significant differences, with vs. without added task, according to post-hoc pairwise comparisons: *** $p < 0.001$. See also Figure S2.

Table 1.

Distribution of beauty ratings in Experiments 1A,B (N = 20 each) and Experiment 2 (N =22).

For each stimulus category with or without an added task. Beauty ratings: "definitely not" ($B = 0$); "perhaps no" (1); perhaps yes (2); "definitely yes" (3). The teddy bear was used only in Experiment 1B. See also Figure S4.

Stimulus Kind	Count without Task				%	Count with Task				%
Beauty Rating	0	1	2	3		0	1	2	3	
Experiment 1A[a]										
Self-selected beautiful	0	0	2	18	90	1	2	9	8	40
High-valence IAPS	0	1	3	16	80	0	5	12	3	15
Mid-valence IAPS	0	8	10	2	10	5	6	5	4	20
Neutral IKEA	8	9	3	0	0	10	8	2	0	0
Candy	4	3	7	6	30	4	2	8	6	30
Experiment 1B[a]										
Self-selected beautiful	0	0	3	17	85	0	0	9	11	55
High-valence IAPS	0	2	8	10	50	0	4	11	5	25
Mid-valence IAPS	6	1	8	5	25	3	5	9	3	15
Neutral IKEA	9	4	5	2	10	8	6	6	0	0
Candy	5	2	6	7	35	6	6	7	1	5
Teddy bear	2	5	7	6	30	5	5	5	5	25
Experiment 2[b]										
Self-selected beautiful	0	0	0	22	100	1	0	4	17	77
High-valence IAPS	1	3	9	9	41	2	4	8	8	36
Mid-valence IAPS	4	6	7	5	23	4	5	9	4	18
Neutral IKEA	9	4	9	0	0	12	7	3	0	0
Candy	2	2	8	10	45	1	5	11	5	23

For each stimulus, category is with or without an added task. Beauty ratings: "definitely not" ($B = 0$); "perhaps no" (1); perhaps yes (2); "definitely yes" (3). The teddy bear was used only in experiment 1B.

See also Figure S4.

[a]N = 20 each.

[b]N = 22.