VISUAL COGNITION
https://doi.org/10.1080/13506285.2023.2250505

Investigating the automaticity of links between body perception and trait concepts

Andrew Wildman [a] and Richard Ramsey [b]

[a]Wales Institute for Cognitive Neuroscience, School of Psychology, Bangor University, Bangor, UK; [b]Department of Psychology, Macquarie University, Sydney, Australia

ABSTRACT

Social cognition has been argued to rely on automatic mechanisms, but little is known about how automatically the processing of body shapes is linked to other social processes, such as trait inference. In three pre-registered experiments, we tested the automaticity of links between body shape perception and trait inference by manipulating cognitive load during a response-competition task. In Experiment 1 (*N* = 52), participants categorised body shapes in the context of compatible or incompatible trait words, under high and low cognitive load. Bayesian multi-level modelling of reaction times indicated that interference caused by the compatibility of trait cues was insensitive to concurrent demands placed on working memory resources. These findings indicate that the linking of body shapes and traits is resource-light and more "automatic" in this sense. In Experiment 2 (*N* = 39) and 3 (*N* = 70), we asked participants to categorise trait words in the context of task-irrelevant body shapes. Under these conditions, no evidence of interference was found, regardless of concurrent load. These results suggest that while body shapes and trait concepts can be linked in an automatic manner, such processes are sensitive to wider contextual factors, such as the order in which information is presented.

ARTICLE HISTORY
Received 14 January 2022
Accepted 7 June 2023

KEYWORDS
Social cognition; body perception; automaticity; trait inference; cognitive load

To develop mental representations of other people, we often link together various person features. For example, we may integrate what someone looks like, including their body size and shape, with inferences relating to their character, such as whether they are kind and generous. Although such feature-integration is commonplace, the type and nature of cognitive processes that underpin such binding of social information are far from clear. Indeed, it is unclear if such links rely on more automatic or more effortful operations with regards to the involvement of the central executive. Therefore, the current study shines new light on the cognitive processes that link perceptual and inferential aspects of social cognition together, by assessing the extent to which such processes rely on the availability of central executive resources.

While ample research has studied trait inferences from faces (Engell et al., 2007; Todorov et al., 2009; Todorov & Engell, 2008; Todorov & Uleman, 2003), bodies have received less attention. Nevertheless, bodies signal important social information (Aviezer et al., 2012; de Gelder et al., 2010). For example, body perception gives rise to stable inferences relating to aspects of people's character such as health and personality (Greven et al., 2018; Naumann et al., 2009; Puhl & Heuer, 2009; Wildman & Ramsey, 2021). Furthermore, inferences from bodies can be driven by expressive trait-implying behavior (e.g., overt body-language; de Gelder, 2006), as well as invariant features such as body size and shape (Hu et al., 2018). More broadly, with growing levels of obesity and body image dissatisfaction disorders emerging (Hehman et al., 2017; Wang et al., 2011; Zopf et al., 2016), the functional significance of body perception is of growing societal importance.

Of course, we do not solely rely on direct observation of visual appearance or behavior to form judgements of people's character. Trait-diagnostic information can also be gathered indirectly, such as

CONTACT Richard Ramsey richard.ramsey@hest.ethz.ch Department of Health Sciences and Technology and Department of Humanities, Social and Political Sciences, ETH Zürich, Zürich, Switzerland
Supplemental data for this article can be accessed online at https://doi.org/10.1080/13506285.2023.2250505.

when talking with a friend or when reading a book (Mitchell, 2009; Mitchell et al., 2006). Such indirect trait-inferences rely on the operations of the theory of mind network, which comprises neural regions that are largely distinct from networks involved in the visual processing of bodies (Ramsey, 2018). The theory of mind network spans the medial prefrontal cortex, temporoparietal cortex and the temporal poles (Frith & Frith, 1999; Saxe & Kanwisher, 2003; van Overwalle, 2009). In contrast, body image perception engages patches of occipitotemporal cortex in the ventral visual stream (Downing & Peelen, 2011). What this research makes clear is that social cognition in general, as well as body perception and social inferences more specifically, encompasses a distributed set of cognitive processes and neural networks.

Although past research on body perception and theory of mind has largely been conducted separately, more recent research has begun to probe the relationship between perceptual and inferential processes in body perception. For example, using fMRI, Greven and colleagues (2016, 2017a, 2017b) showed that links between body-shape perception and trait knowledge inferences involves functional coupling between distinct neural networks associated with body perception and theory of mind. Furthermore, behavioral research has shown that trait-inferences can bias later body-size judgments (Wildman & Ramsey, 2021). These studies, therefore, show that abstract trait concepts and visual depictions of body shapes encompass spaces with partially shared structure, which can lead to bias and influence between visual and inferential domains (Over & Cook, 2018; Ramsey, 2018).

The demonstration of reciprocal links between perceptual and inferential processes has opened new lines of research in understanding body perception, but many questions remain unanswered about the type and nature of such links. For example, to what extent are links between visual body shape representations and trait concepts reliant on more automatic or more deliberate cognitive processes?

One method to assess the automaticity of cognitive processes is provided by dual-task paradigms (Lavie, 2005; 2010). In one version of the paradigm, a concurrent working memory task, which heavily taxes central cognitive resources, is performed alongside the main task under investigation. For example, a high load condition may require holding six letters in memory, whereas a low load condition may only involve holding one letter in memory (Lavie et al., 2004). Limiting the availability of central resources by manipulating working memory load is argued to impair selective attention and reduce the ability to "filter out" distractor events competing for control over behavior.

As a result, in the context of a stimulus response compatibility paradigm, typical response patterns across differential load conditions show an increase in distractor interference under high load compared to low load for most classes of stimuli. For example, studies have shown this pattern of findings for the effects of task-irrelevant distractor letters on the identification of target letters (Konstantinou et al., 2014), and the effects of visual capture by singleton distractors during visual search (Lavie & De Fockert, 2005). Additionally, de Fockert et al. (2001) showed that task-irrelevant faces, which were either compatible or incompatible in terms of identity, interfered with the ability to classify written names as pop stars or politicians to a greater extent under high than low load. In exceptional cases, the level of interference is insensitive to concurrent demands placed on working memory (i.e., identical to low or no load), indicating that the processes relied on to complete task objectives are dissociable from attentional resources. Examples of such exceptions include interference from spatially incongruent visual and touch cues (Zimmer & Macaluso, 2007), counterproductive gaze-cues (Hayward & Ristic, 2013), and incongruent finger movements (Ramsey et al., 2019). What this research makes clear is that high cognitive load increases the likelihood of interference from task-irrelevant stimuli and can therefore reveal whether working memory resources play a regulatory role in this interference.

In the current work, if the process under study is relatively resource-intensive, a high compared to low cognitive load should increase interference in the main task. In contrast, minimal impact of load on the main task can be taken as an indicator of a relatively resource-light and efficient process that operates the same, irrespective of the level of load. In the present series of experiments, we used a dual-task paradigm to investigate the automaticity of links between body shapes and trait concepts. In a series of pilot studies, we first developed an interference task that was sensitive to a compatibility effect between body shapes and trait words. Subsequently, in three pre-registered experiments, we employed a dual-task paradigm to differentially tax working memory resources between high and low load conditions. In Experiment 1, participants categorized body shapes in the presence of compatible or incompatible trait terms. In Experiment 2 and 3, participants categorized the same trait words across different judgement dimensions in the presence compatible or incompatible body shapes. Across all experiments, if the linking process is relatively resource-intensive, then high load would increase the reaction time (RT) interference effect between bodies and traits. Alternatively, if the linking process is relatively resource-light, then high load would have minimal impact on RT interference. Because our Bayesian multi-level statistical analyses specify the probability of possible parameter values, including zero, our data analysis approach allows evidence of either possibility to be provided.

Experiment 1

Pilot experiments

Prior to conducting our main experimental task in which cognitive load would be manipulated, we sought to establish a response competition task sensitive to an interference effect between traits and bodies. Response competition tasks measure the difference in RT between responses made in the context of compatible or incompatible pairings of target and distractor stimuli. For present purposes, the basic logic of the task was based on evidence that heavier and slimmer bodies tend to be judged differently. For example, heavy individuals are typically rated as less healthy, less extraverted, and less conscientious than slimmer individuals (e.g., Greven et al., 2018; Wildman & Ramsey, 2021). In contrast, greater muscularity is associated with higher ratings of health and extraversion, as well as lower ratings of agreeableness. As a result, we expected differences in RT for body-trait pairings that were compatible (e.g., slim, muscular body shapes paired with extraverted or healthy traits), versus incompatible (e.g., heavy body shapes paired with extraverted or healthy traits). Such an interference effect would be indicative of links between body shapes and trait concepts.

We conducted a series of four pilot studies (Total $N = 79$), in which participants categorized bodies as either slim or heavy in the context of trait adjectives appearing as text on-screen. Several aspects of the design were adjusted between pilots, allowing us to refine the main experimental task, which was to be nested within our working memory load manipulation. A detailed report on the methods and findings of the pilot experiments is available in supplementary materials (see Supplementary Report).

A pooled analysis of the datasets from Pilot Experiments 3 and 4 (total $N = 40$) showed a small positive effect of body-trait compatibility on the speed of participants' responses. These findings suggest that the task parameters used in these last two pilot experiments brought about interference, which was detectable with a similar level of statistical power that we would be using in the main experiments. As a result, we used this version of the paradigm as a basis for our main experimental task, as well as changing some of the trait adjectives used in the pilots to improve clarity and concreteness. The pilot studies also confirmed that participants were able to categorize slim and heavy bodies accurately even when presented for only 30 ms. This constrained display interval for body stimuli was maintained for the main task, as it was likely to maximize the interference between trait adjectives and bodies.

Method

Pre-registration and open science statement

Across all three experiments, the research questions, hypotheses, planned analyses, and exclusion criteria were pre-registered. For Experiment 1, the pre-registration can be found at: https://aspredicted.org/uh5m5.pdf. We mention any deviations from the pre-registration in the text below. In addition, following open science initiatives (Munafò et al., 2017), the raw data, stimuli, and analysis code for each experiment are available online on the open science framework (https://osf.io/4en9f/). Based on the pilot studies, we had a basis to expect an interference effect that was detectable in a sample of around 50 participants. As a result, we aimed to collect a sample of 50 participants, and therefore set this as our stopping rule for data collection.

Participants

Fifty-six Bangor university students were recruited through Bangor University's student participation panel in exchange for course credit (17 male, 1 unspecified, $M_{age} = 19.20$, $SD_{age} = 0.91$). Following data preprocessing and exclusions, the final size of the sample subjected to our analyses was 52.

Stimuli

Eight computer-generated female bodies (four slim and four heavy) were produced in *MakeHuman* (version 1.1.1; www.makehumancommunity.org), a program for creating 3D human models. A slim body prototype was created using the muscular mesh, with the muscle parameter set to max and the weight parameter reduced slightly. A heavy prototype was produced using the default mesh, with the weight parameter set to max and the muscle parameter reduced considerably. Further changes were made to various parameters of the slim and heavy prototypes to maximize the salience of the overall size of the body. For the individual body identities, small variations were made to the body proportions and skin tone of these prototypes to produce four visually distinct bodies at each level of body size. Finally, these were rendered to 320 × 790 PNG images and cropped in *GIMP* to isolate the body (see Figure 1.).

Sixteen trait adjectives were selected for use in the current experiment, belonging to the categories of: healthy, unhealthy, extraverted and introverted (see Table 1). The pattern of compatibility was based on previous research regarding mappings between body shapes and trait inferences (e.g., Greven et al., 2018; Hu et al., 2018; Wildman & Ramsey, 2021). As a result, healthy and extraverted traits were coded as compatible with slim bodies, while unhealthy and introverted traits were coded as compatible with heavy bodies.

Tasks

Overview. All aspects of the experimental task were created and implemented in *MATLAB* 2015b using Psychtoolbox 3 (www.psychtoolbox.org). The experiment employed a dual-task paradigm in which participants were subjected to one of two load conditions across the duration of each experimental trial (working memory manipulation). During trials, participants were asked to categorize displayed bodies as slim or heavy as quickly and accurately as possible, in the context of compatible or incompatible trait cues (body categorization task). Participants were first asked to complete two separate practice blocks to familiarize themselves with the requirements of each task independently of the other. The first practice block consisted of 20 trials of the body categorization task, while the second consisted of 16 trials of the working memory task. Following this, participants completed the main experimental block of 256 trials, where the main body categorization task was nested within the secondary working memory manipulation. In this main experimental block, every combination of body and trait stimulus was presented in each condition of load and compatibility in a random order. Given the nature of our body stimuli, the best visual contrast for trait stimuli was achieved by placing them over the torso of the body rather than the true center of the screen. The locations of all stimuli were therefore standardized to this higher position by offsetting them by a fixed number of pixels vertically. However, for simplicity, this location is henceforth referred to as the center (see Figure 2).

Categorization task. The categorization task required participants to categorize a body stimulus as either slim or heavy as quickly and accurately as possible, in the presence of a task-irrelevant trait cue. The task began with a fixation cross that was displayed in the center of the screen for 1000 ms, followed by a trait adjective which remained in the center of the screen for 700 ms. A body was then displayed on-screen behind the trait adjective for 30 ms, before being backward masked. A unique mask was used for each of the eight body identities, each comprised of 10 pre-rendered images displayed serially for 10 consecutive frames. On each of these frames the original body image had been divided into a 6 × 10 grid, and the resulting rectangles rearranged randomly. These scrambled images were intended to limit any visual after-effects of the body which might otherwise bleed into the response phase of the task.

During and after the mask, a yellow question mark was displayed in the center of the screen, prompting participants to respond by pressing "K" (slim) or "M" (heavy). From the onset of the mask, participants had 1970ms to respond (making the overall time constraint 2000ms from the onset of the body). The body categorization task ended either when the participant responded, or when this time elapsed. At this point there was an interval of 1000 ms. Although the screen was usually blank during this interval, if the

Figure 1. Body stimuli used in all experiments.

participant had not responded to the trial the word "missed" would be displayed in the center of the screen. This was to indicate to participants that they had failed to respond or not done so quickly enough. RT was recorded as the amount of time that elapsed between the offset of the body stimulus and the participant's response. The keypress was also recorded, and later used to calculate participants' accuracy scores based on the proportion of correct responses (both overall and in each condition).

Working memory manipulation. The main task was nested within a secondary working memory manipulation. At the start of each trial, a set of six dots were displayed in a hexagonal array around the center of the screen for approximately 500 ms. One or all of these dots were then replaced by letters comprising a memory set, which consisted of either a single letter (low load condition) or a set of six letters (high load condition). The letter/s at each relevant position of the array were selected at random with equal probability from a set of 10 possibilities: F, H, K, L, M, T, V, W, Y and X. In the low load condition, the letter was always presented at the top of the array with the other five positions replaced by dots, whereas in the high load condition, all six positions were occupied by a unique letter. The letter array remained on screen for either 850 ms (low load) or 2000ms (high load), giving participants time to attend to each letter in the set. Participants were instructed to hold these letter/s in memory for the duration of the trial and were informed that a memory probe would test their retention at the end of the trial.

The memory probe consisted of a new letter displayed in yellow text in the center of the screen, which was equally likely to be one that had or had not appeared in the original memory set. On high load trials where the probe letter had been present at the start of the trial, this was equally likely to be drawn from any of the six locations of the original array. In cases where the probe letter did not match, this letter was equally likely to be any of the remaining letters from the original set of 10. To respond to the probe, participants were instructed to press "E" if the letter had been present in the initial memory set, or "D" if the letter had been absent. Participants had a maximum of 3000 ms to respond to the memory probe. The accuracy of this response was recorded as a manipulation check. Percentage accuracy was later calculated for each condition based on the proportion of correct responses in each condition. Mean differences across the sample, as well as their associated confidence intervals, were calculated based on the difference in accuracy scores between high and low load both on average and within each remaining condition of the design.

Table 1. Trait adjective stimuli.

Compatible with:	Slim bodies		Heavy bodies	
Category	Healthy	Extraverted	Unhealthy	Introverted
	Active	Outgoing	Inactive	Reserved
	Energetic	Talkative	Lethargic	Quiet
	Strong	Social	Weak	Unsocial
	Fit	Confident	Unfit	Shy

Trait probe

On a small subset of trials, a catch trial followed the ordinary sequence of events to probe participants

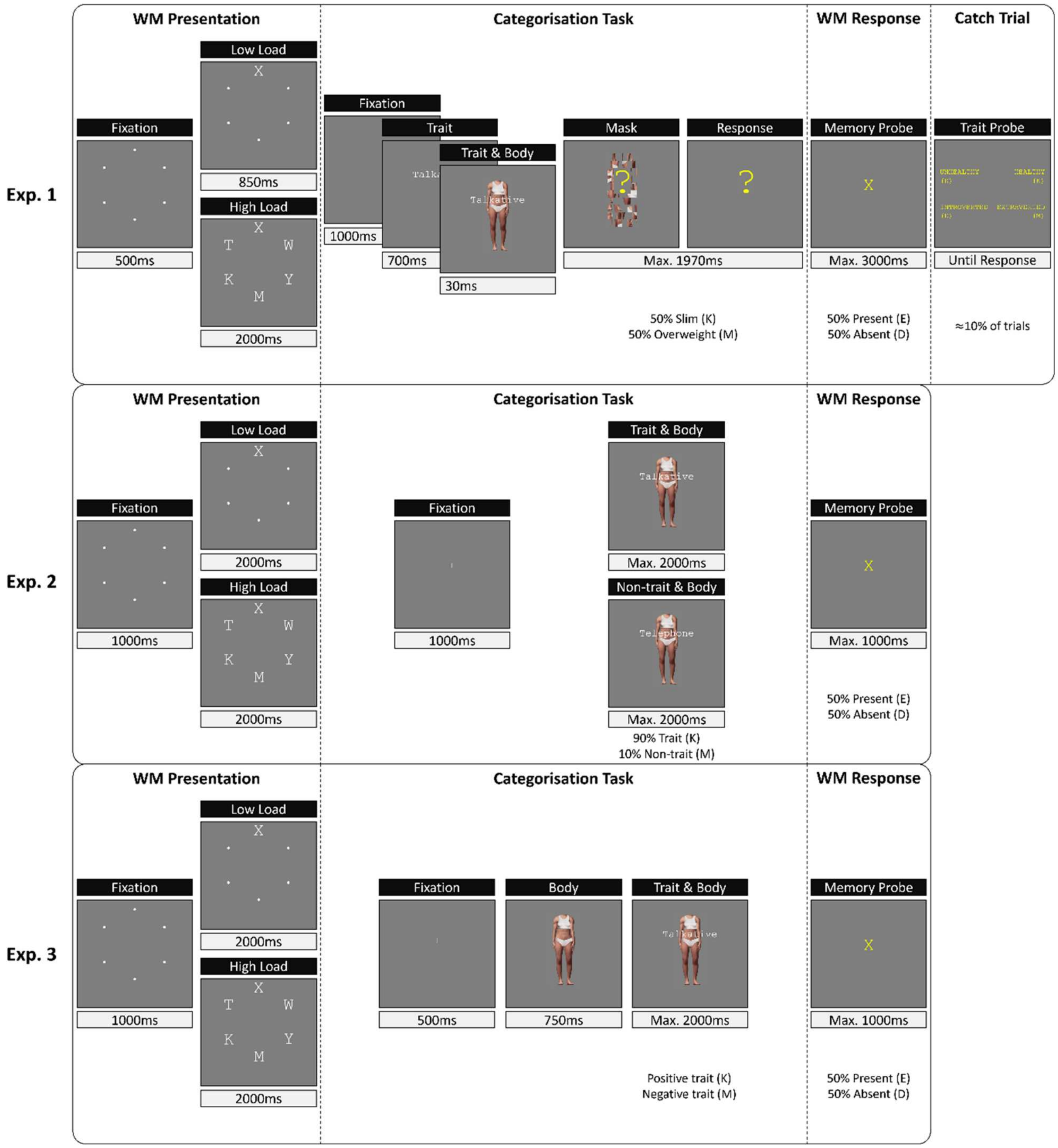

Figure 2. Trial diagram for all three experiments. The correct response to the memory probe in all cases would be E (present). The mask following presentation in Experiment 1 lasted for the first 10 frames of the response portion of the trial (approx. 200 ms).

knowledge of the trait cue. On these trials a question was displayed, asking participants to indicate which of four categories best described the trait adjective which had been displayed on the preceding trial (healthy, unhealthy, extraverted or introverted). 26 of these catch trials occurred at random points during the main block of the experiment, and there were no time constraints on responses. Accuracy on these catch trials was recorded to gauge engagement and for use as a possible filtering criterion for data analysis.

Procedure

Participants were invited into the testing lab and asked to complete a consent form. The experimental task was then described to them verbally before commencing the two practice blocks. In rare cases where participants had difficulty understanding a particular aspect of the task (either the body-categorization task or the working memory task), they were invited to repeat the relevant practice block. Following this, participants completed the main block of the experiment as described above. After this, participants also completed an Implicit Association Test (IAT) constructed using the same body and trait stimuli from the main experiment (see Greenwald et al., 1998). This additional task was not part of our primary research question, but was included for the benefit of exploratory analysis which may seek to investigate relationships between this and our other measures. Finally, participants completed a short demographic questionnaire before being debriefed and awarded course credits for their participation.

Design and data analysis

All conditions and comparisons were within-subjects. We indicated in the pre-registration for Experiment 1 that we would use a within-subjects factorial ANOVA as our primary inferential statistical test. However, since we submitted the pre-registration, we have begun to use a Bayesian multi-level estimation approach as a primary method of statistical inference. Given that multi-level models are a more comprehensive tool and avoid the unrealistic assumptions of ANOVA (Barr et al., 2013), we decided to report our original pre-registered ANOVA in supplementary materials (see Supplementary Analysis). In the main text we report a Bayesian model, which also mirrors our pre-registered analysis approach for Experiments 2 and 3. No meaningful differences existed between the inferences we drew from the Bayesian multi-level modeling, and the inferences we would have drawn from the ANOVA.

We followed a Bayesian estimation approach to multi-level modeling (McElreath, 2020), which our lab has adopted in recent papers (Bara et al., 2021; 2023i). The main goal was to estimate parameters of interest in multi-level models of varying complexity, and compare the performance of these models. Therefore, we used two approaches to guide our interpretation of the findings. First, we reported and discussed the posterior distribution of our key parameters of interest within the most complex model. Second, we performed model comparison via efficient approximate leave-one-out cross validation (LOO; Vehtari et al., 2017). LOO is a method of estimating how accurately the model in question can predict out-of-sample data. Therefore, we took all the models and compared how accurately they could predict out-of-sample data. In this way, we could estimate the extent to which an increase in model complexity corresponded to an increase in model accuracy.

More specifically, we followed a recent translation of McElreath's (2020) general principles into a different set of tools (Kurz, 2020), which use the Bayesian modeling package "brms" to build multi-level models (Bürkner, 2017, 2018). Additionally, our data wrangling approach follows the "tidyverse" principles (Wickham & Grolemund, 2016) and we generate plots using the associated data plotting package "ggplot2", as well as the "tidybayes' package (Kay, 2020).

Given that our primary dependent variable is RT, we modeled the data using a shifted log-normal model, which has previously been shown to be particularly well-suited to fitting the distribution of RT data (Haines et al., 2020). Following the "keep it maximal" approach to multi-level modeling (Barr et al., 2013), we included the maximal number of varying effects that the design permitted. As such, varying intercepts and effects of interest were estimated for participants and stimulus items when possible.

We computed 10 models, which built incrementally in complexity. We first computed three intercepts-only models, just so that we could compare subsequent models that included predictors of interest to models without any predictors. Model b0.1 included an overall intercept, model b0.2 additionally included varying intercepts for participants and stimulus items, and model b0.3 additionally included a varying non-decision time (ndt) or "shift" parameter per participant. We then added predictors for compatibility (b1) body size (b2) and load (b3). Two-way interactions between compatibility*body size (b4.1), compatibility*load (b4.2) and body size*load (b4.3) were then added in further models. Model b5 was the full model, which additionally included the three-way interaction between compatibility, body size and load.

Factors were coded according to a deviation coding style, where factors sum to zero and the intercept can then be interpreted as the grand mean and the main effects can be interpreted similarly to a conventional analysis of variance (http://talklab.psy.gla.

ac.uk/tvw/catpred/). As such, compatibility, body size and load were coded as −0.5 (compatible / slim / low) and 0.5 (incompatible / heavy / high).

We set priors using a weakly informative approach (Gelman, 2006). Weakly informative priors differ from uniform priors by placing a constrained distribution on expected results rather leaving all results to be equally likely. They also differ from specific informative priors, which are far more precisely specified, because we currently do not have sufficient knowledge to place more specific constraints on what we expect to find. Also, by using weakly informative priors, we allow for the possibility of larger effects, should they exist in the data (Gelman, 2006; Gelman et al., 2013; Gelman & Hill, 2006; Lemoine, 2019).

Moreover, a further advantage of weakly informative priors is that we would not expect the choice of prior, as long as it remained only weakly informative, to matter too much because the data would dominate the structure of the posterior distribution. The formula for the full model (model b5) is specified here:

rt ~ 1 + compatibility * body_size * load +

(1 + compatibility * body_size * load | pID) +

(1 + compatibility * load | word_stim) +

(1 + compatibility * load | body_stim),

ndt ~ (1 | pID)

Due to our use of a multi-level estimation approach, we sought to retain as much trial-level data as possible. That being said, we still sought to remove cases that were likely to reflect disengagement from the task on both the trial and participant level. As a result, before subjecting our data to analyses, we removed cases on the trial and participant level in accordance with our pre-registered criteria. On the trial level, we first removed all trials for which a response to the main task was not recorded (i.e., where the participant did not respond in time), and ensured that the shortest recorded RT was plausible (in this case the shortest RT was 64 ms). On the participant level, we excluded participants whose average RT in the main task was further than 2.5 SD from the group mean, as well as those whose accuracy in the main task was more than 2.5 SD below the group mean. We also excluded those whose overall accuracy in the working memory task was further than 2.5 SD from the group mean.

Finally, we specified in our pre-registration that those with lower than 80% accuracy in responses to the trait probe catch trials would be considered for exclusion, however upon processing the data we found that few participants performed this well on the trait probe. Moreover, further investigation demonstrated that performance on catch trials did not predict task performance or susceptibility to the interference effect, and therefore this criterion and threshold were unlikely to discriminate between participants who had or had not engaged with the task. As a result, we abandoned this exclusion criterion. Besides the removal of missed trials where no response was recorded, all our remaining filtering criteria on the trial and participant level removed around 7% of the total data. This included four participant exclusions.

Results

Manipulation check

To assess the efficacy of our cognitive load manipulation, we evaluated performance in the working memory task. When averaging across all other conditions, the overall mean difference in accuracy scores was 17.32% [14.59, 20.05], $d_z = 1.77$ [1.32, 2.20] (square brackets denote 95% confidence intervals for all statistics in the article). More specifically, the high load condition (71.09%, [69.18, 73.00]) was lower than the low load condition (88.41%, [86.50, 90.32]). This indicates that, as expected, our cognitive load manipulation substantially decreased performance on the working memory task. This difference was also insensitive to the contexts of compatibility or body size conditions, as indicated by the mean differences and confidence intervals per-condition (see Supplementary Figure S1).

Main task

Reaction time scores for correct responses in the main task are visualized in Figure 3. Visual inspection of the data indicates a small interference effect (slower responses to bodies in the context of incompatible traits, Cohen's $d_z = 0.27$, 95%CI[−0.01, 0.54]), which was largely insensitive to load and body size conditions.

Parameter estimates for the most complex model (Model b5), are shown in Figure 4 and Supplementary Table 1. The posterior distributions for our main predictors indicate a positive effect of compatibility

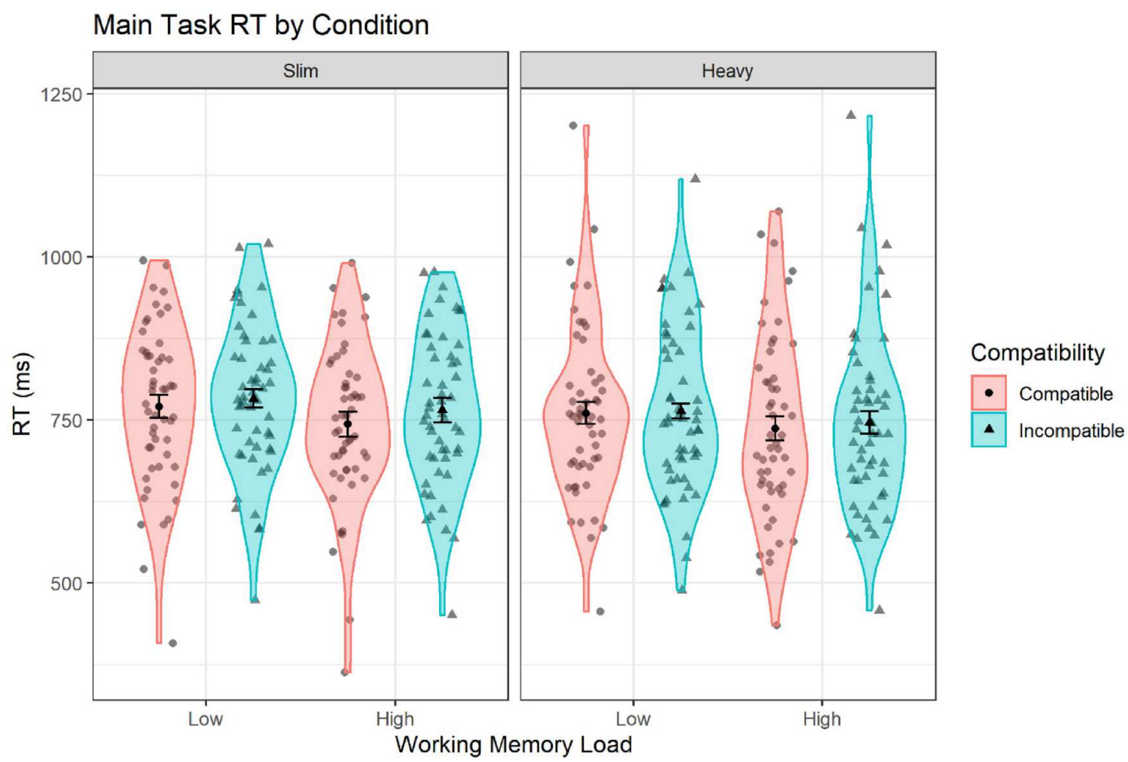

Figure 3. Mean RT scores per participant and across the sample, plotted by load and body size conditions. Error bars represent 95% confidence intervals.

condition (where incompatible trials elicited slower responses), and a negative effect of load condition (where low load trials elicited slower responses). Relating this back to the way our factors are coded, relative to compatible (−0.5), the property of incompatibility (+0.5) increases RT scores (i.e., slower responses). This can be thought of as the effect of one "unit" of incompatibility having a positive effect on RT scores. Similarly, comparing low (−0.5) to high (+0.5) cognitive load reveals that RT scores become lower (faster) when high load is present, thus higher (slower) when high load is absent. This negative effect of load condition is likely to be owed to the different durations of the memory set between high (2000ms) and low (850 ms) load conditions, allowing participants greater time to prepare for the onset of stimuli in the main task under high load. The duration was matched between these conditions in Experiment 2 to increase consistency. The distribution of parameter estimates for all remaining fixed effects, as well as interaction terms, centered around zero, supporting the conclusion that the main interference effect was not modulated by body size and/or load conditions.

Model comparison is visualized in Figure 5. The x-axis (expected log pointwise predictive density) reflects a measure of model performance, based on how well a model fitted with only part of the data predicts the remaining "left-out" data. Specifically, the value is the height of the probability distribution at the point of the left-out data, meaning that higher values correspond to better model performance. All models containing our fixed effects of interest outperformed the intercept-only model (Model b0.1), and the intercept plus varying item and participant intercepts model (Model b0.2). The addition of a varying non-decision time parameter per participant (Model b0.3) approximated the predictive accuracy of all subsequent models (see Figure 5). Error bars for all remaining models overlapped, suggesting similar performance in terms of out-of-sample predictive accuracy.

Accuracy scores showed broadly the same pattern as the RT data and are visualized in Supplementary Figure S3.

Discussion

The first experiment showed an effect in which trait-implying words interfered with the categorization of body stimuli, which was unaffected by the addition

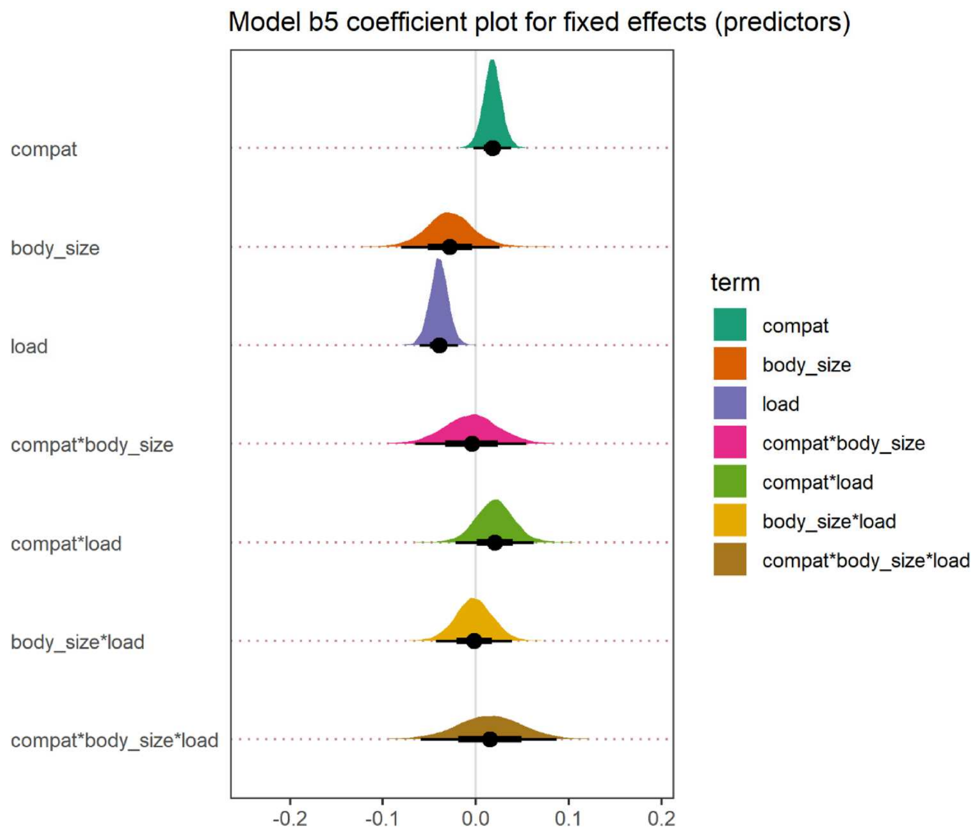

Figure 4. Parameter estimates for each predictor within Model b5.

Note: compat = compatible vs. incompatible; body_size = slim vs. heavy; load = high vs. low; x-axis = log(RT); point estimate = median; error bars represent 66% quantile intervals (thick black lines) and 95% quantile intervals (thin black lines). Interpretating these parameter estimates in terms of their original units is complex, as the shifted lognormal model is comprised of three components. To see estimates of these parameters in original units (milliseconds), please see Supplementary Figure S2.

of a demonstrably difficult secondary task. This pattern of findings suggests that the cognitive resources used by working memory are either entirely separate or only minimally involved in mediating interference in the main task. As such, the results suggest that cognitive processes involved in linking bodies with trait inferences are relatively automatic in the sense that they are

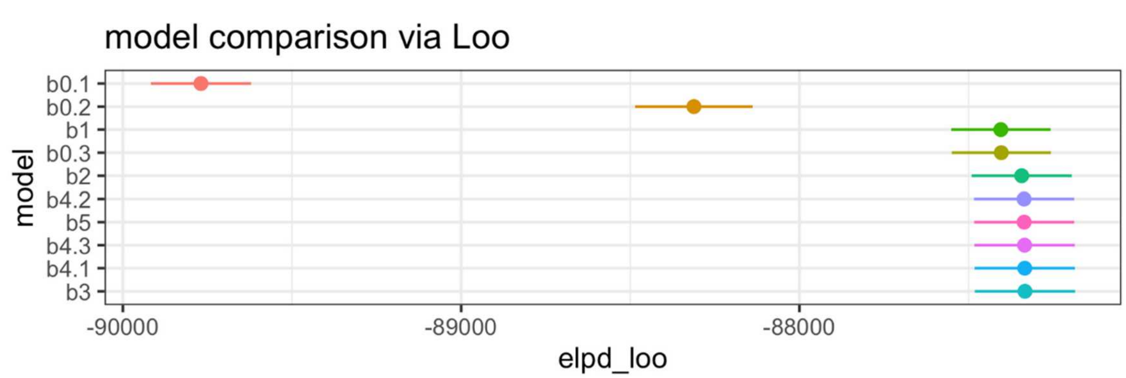

Figure 5. Model comparison (1–10 models).

Note: Model b0.1 included an overall intercept; Model b0.2 added varying intercepts for participants and stimulus items; Model b0.3 added a varying non-decision time parameter per participant; Model b1 included predictors for compatibility (compatible vs. incompatible); Model b2 included predictors for body size (slim vs. heavy); Model b3 included predictors for load (low vs. high); Model b4.1 included the interaction between compatibility and body size; Model b4.2 included the interaction between compatibility and load; Model b4.3 included the interaction between body size and load; Model b5 was the full model, and included the three-way interaction between compatibility, body size and load. Elpd_loo = estimate of the expected log pointwise predictive density; loo = leave-one-out estimated cross validation; error bars = standard error of the mean.

resource-light, and efficient, such that they operate in an unaffected manner when cognitive resources are taxed heavily.

To contextualize the interpretation of Experiment 1, we consider both the generalisability of the findings and one possible limitation, which together form the basis for a second follow-up experiment. First, we consider the question of whether the basic pattern of findings generalizes to a novel experimental context in which the roles of target and distractor stimuli are reversed. To do so, participants would respond to trait words while ignoring task-irrelevant body shapes. Assessing generalisability in this way will establish if the links between bodies and trait-words are also measurable when the information is presented differently and whether such links remain unaffected under cognitive load.

At the same time, reversing the roles of target and distractor stimuli in this manner would also address a possible limitation of Experiment 1, which relates to the consideration of data signal limits. While our experimental manipulation had a demonstrable effect on resource availability, data availability is also a critical factor in cognitive processing. Specifically, where a data signal is limited, the availability of additional processing capacity will not benefit performance outcomes that depend on this data (Norman & Bobrow, 1975). On this basis, insensitivity of interference to load could reflect a data signal limit, rather than processing efficiency.

As a result, by reversing the roles of target and distractor stimuli in the manner described above, we can pose the question of whether the effect of categorizing trait-implying words in the context of task-irrelevant bodies is similarly invariant to cognitive load. Doing so can address both issues because, with respect to the relevant processes under study, a body shape is likely to reflect a greater source of social information than a single trait-implying word (i.e., individuals may possess multiple traits, but not multiple bodies).

Experiment 2

To further probe the automaticity of links between body shapes and trait concepts, we created a second experiment in which the roles of the bodies and traits as target and distractor stimuli were switched. We also changed the dimension across which stimuli would be categorized, meaning the task assessed whether the purely incidental presence of compatible body shapes would automatically interfere with the processing of trait concepts. Following the same set of predictions as Experiment 1, we expected a RT interference effect between body shapes and trait concepts such that compatible pairings are categorized faster. An increase in the size of this effect under high load would be indicative of resource-intensive processing, whereas a more resource-light linking process would be typified by load having minimal impact on RT interference.

Method

Pre-registration and open science statement

As before, our research question, hypotheses, planned analyses, and exclusion criteria were pre-registered (https://aspredicted.org/hb9i4.pdf). Also, the raw data, stimuli, and analysis code are available online on the open science framework (https://osf.io/4en9f/). As our second experiment served primarily to replicate the same basic effect of Experiment 1 with an altered experimental task, we aimed to collect data from a similarly sized sample of around 50 participants.

Participants

Sixty-two participants were recruited online via Bangor University's student participation panel in exchange for course credit (9 males, $M_{age} = 22.2$, $SD_{age} = 4.9$). Following data pre-processing and exclusions, the final size of the sample subjected to our analyses was 39.

Stimuli

All experimental stimuli were identical to those used in Experiment 1, except for a set of non-trait words used as stimuli for infrequent catch-trials. Although the sizes and locations of stimuli were similar to those of Experiment 1, online data collection meant that participants completed the experiment on a wide variety of systems likely to have different monitor sizes and display settings. As a result, all stimuli sizes and locations were scaled relative to the height of the monitor on which they were displayed, meaning that the exact sizes of stimuli will have varied somewhat between participants.

Tasks

Overview. As data for our second experiment were collected online, the new task was created in *PsychoPy (2020.2)* (Peirce et al., 2019). This again consisted of a dual-task paradigm in which participants had to categorize stimuli in the presence of task-irrelevant distractors (see Figure 2.). Participants were automatically guided through three practice blocks consisting of just the trait-categorization task, just the working memory task, and finally a short block of the full task. Participants were also given the option of repeating the third practice block, however few chose to do so. The main experimental block consisted of 288 trials (256 experimental trials, 32 catch trials). As before, the locations of stimuli were offset vertically relative to the bodies to optimize readability. This is again referred to as the center for simplicity.

Trait categorization task. In the main task, participants had to decide whether a word written on-screen was a trait or non-trait. Traits were the same adjectives used in Experiment 1, and were displayed on around 90% of trials (256 out of 288 total trials). Non-traits were a set of nouns presented as catch trials on the other 10% of trials (32 trials). These were chosen to be of similar length and appearance to our trait stimuli to maximize the attention and engagement required to detect them (see Table 2.). Stimuli appeared in the center of the screen following a 1000 ms fixation cross, remaining for a maximum of 2000ms or until the participant responded. Participants were asked to press the "M" key if the word was a trait, or "K" if the word was a non-trait. Stimuli were presented in a fully randomized order, and response accuracy and RT were recorded.

Working memory manipulation. As in Experiment 1, the main task was nested within a secondary working memory task. Most aspects of this were identical to that of Experiment 1, although some minor differences existed due to practicalities of online data collection (see Figure 2). In particular, the memory set was always displayed for 2000ms, regardless of load condition, and written feedback followed the memory probe to compensate for the lack of a physical lab environment and experimenter. As a result, the memory probe ended either when the participant responded or after 1000 ms, at which point "Correct!," "Incorrect!," or "missed" was displayed based on how or whether they had responded. RT and accuracy were recorded and calculated in the same manner as Experiment 1.

Procedure

Following recruitment, participants were directed to a Qualtrics survey which functioned as the consent form. Here participants were advised of the nature of the experiment, their right to withdraw and how to do so. At the end of this survey, they were automatically redirected to the experiment hosted on Pavolvia. Upon completion, participants were redirected to another Qualtrics survey which collected demographic information before presenting the debrief form. Here they were also asked two brief questions regarding strategies used during the experiment, the data from which are available for the purpose of exploratory analyses. Upon submitting this final survey, participants were redirected to the automatic credit-granting system.

Design and data analysis

Per our pre-registration, we used a Bayesian multilevel estimation approach to evaluate our data, which was identical to that used in Experiment 1. As we also outlined in our pre-registration, we applied several criteria to judge exclusions with an overall aim of reducing or eliminating cases which reflected disengagement from the experimental task. This was especially relevant in the context of an online experiment where experimenter control was minimized. On the trial-level, we first excluded all non-trait trials, as these were unrelated to our hypotheses. We also excluded trials on which the

Table 2. Traits and non-traits used in Experiment 3.

Trait Adjective (same as Experiment 1)	Non-trait Counterpart
Outgoing	Outhouse
Talkative	Telephone
Social	Socks
Confident	Confetti
Active	Acorn
Energetic	Engine
Strong	Stream
Fit	Fur
Reserved	Restaurant
Quiet	Quilt
Unsocial	Unicycle
Shy	Sea
Inactive	Insect
Lethargic	Lettuce
Weak	Wood
Unfit	Unicorn

response was incorrect, as these were unlikely to yield valid effects of compatibility or interference. Finally, we excluded trials with implausibly short RTs, as these were likely to reflect an accidental keypress.

On the participant-level we sought to exclude participants with high error rates likely to reflect disengagement. Our intention was to exclude participants with less than half the maximum possible trials per cell after excluding errors, however upon calculation it was found that this threshold would exclude a very high proportion of participants. Instead, we applied a threshold of a quarter of maximum possible trials for our main model. As trait trials represented the majority of trials across the experiment, we evaluated performance on non-trait trials as another measure of engagement. Participants with below chance performance on non-trait trials were excluded, as well as those whose overall accuracy in the working memory task was below 55%. In total, our filtering criteria excluded 22 participants. Although our revised exclusion criteria involve a relaxation of our preregistered criteria, we believe this reflects a balance between retaining potentially meaningful data compared to excluding a much greater proportion of our total sample than originally anticipated.

Results

Manipulation check

As before, we first sought to evaluate the efficacy of our cognitive load manipulation. Having removed non-trait trials, we calculated mean differences and confidence intervals for participants' accuracy (calculated in the same manner as in Experiment 1) and RT scores on average and at each level of the design. The size of this mean difference for accuracy scores was similar to that of Experiment 1: mean difference = 18.97% [15.78, 22.15], $d_z = 2.11$ [1.49, 2.72]. Again, accuracy was lower for high load (66.87%, [64.62, 69.12]), than low load (85.84%, [83.59, 88.09]). Participants' RT scores also followed a similar pattern: mean difference = 68.10 ms [50.66, 85.53], $d_z = 1.39$ [0.90, 1.86], with responses faster for low load (534.66 ms, [522.33, 546.99]), than high load (602.76, [590.43, 615.09]). Both of these patterns were consistent across all remaining levels of the design (see Supplementary Figures S4 and S5). This reflects the anticipated impact of load, yielding large and unambiguous effects on both RT and accuracy in responses to the working memory probe.

Main task

Reaction time scores for correct responses in the main task are visualized in Figure 6. Visual inspection of the data does not show the predicted interference effect. Instead, RTs follow the category of trait word displayed. Responses were consistently faster for traits associated with slimmer body shapes (extraverted and healthy), and slower for those associated with heavier body shapes (introverted and unhealthy), irrespective of compatibility. In our figures and analyses, this is reflected by an interaction effect between compatibility and body size. However, this pattern simply signifies faster categorization of extraverted and healthy trait words regardless of condition. As such, it is possible that this effect is accounted for by the greater processing speed typically associated with positively-valenced stimuli (e.g., Kauschke et al., 2019; Kuperman et al., 2014).

Parameter estimates for the most complex model (Model 5), are shown in Figure 7 and Supplementary Table 1. The posterior distributions for our main predictors indicate a negative effect for the compatibility*body size interaction term. This reflects the faster categorization of extraverted and healthy trait words, irrespective of compatibility (as discussed above). The distribution of parameter estimates for all remaining fixed effects centered around zero, indicating no systematic effects of our manipulations.

Model comparison is visualized in Figure 8. All models containing our fixed effects of interest outperformed the intercept-only model (Model b0.1), and the intercept plus varying intercepts for participants and items model (Model b0.2). Again, the model which included a varying non-decision time parameter per participant (Model b0.3) had approximately equal predictive accuracy to all subsequent models (see Figure 8). Error bars for all remaining models overlapped, suggesting similar performance in terms of out-of-sample predictive accuracy.

Accuracy scores, calculated in the same way as before, showed broadly the same pattern as the RT data, and are visualized in Supplementary Figure S7.

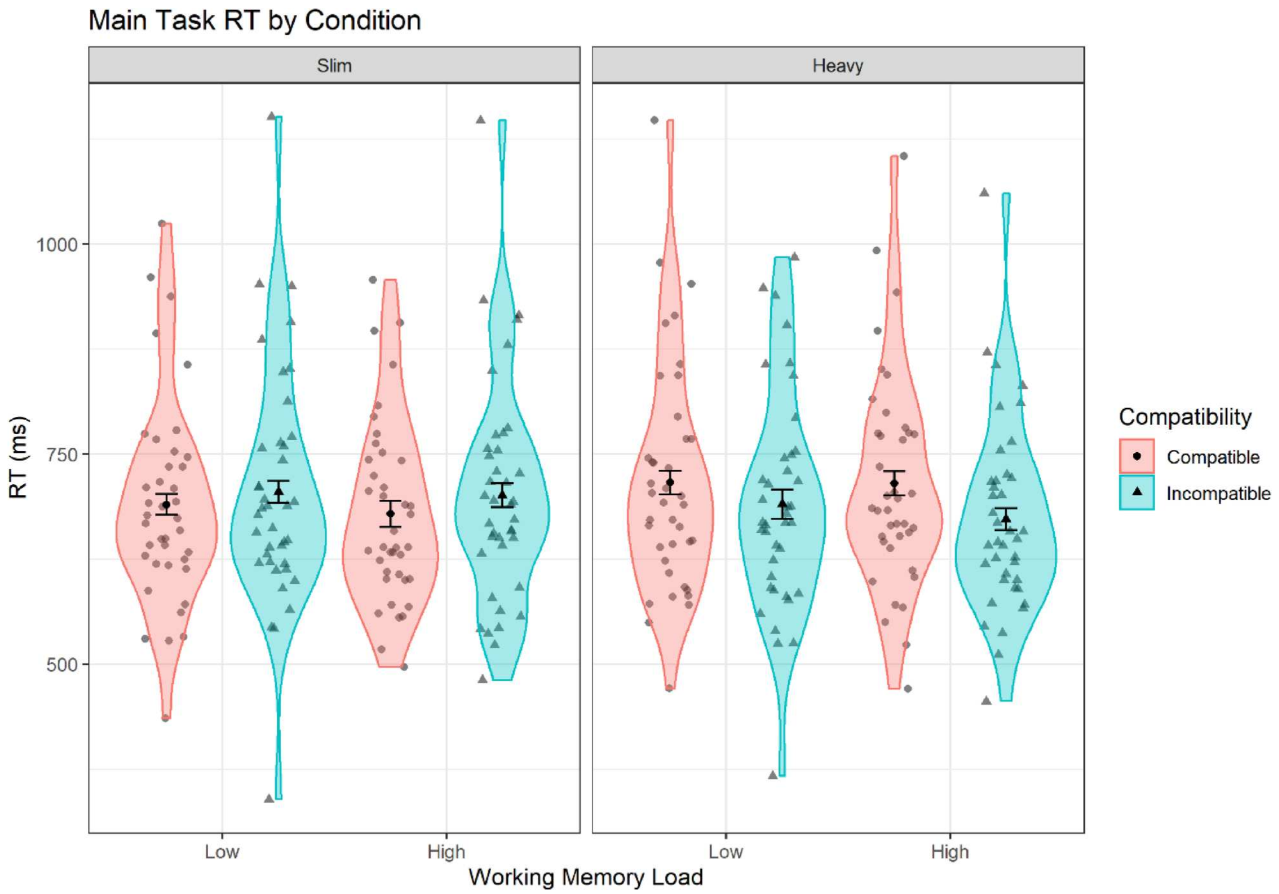

Figure 6. Mean RT scores per participant and across the sample, plotted by load and body size conditions. Error bars represent 95% confidence intervals.

Discussion

Experiment 2 did not yield the same pattern of interference as found in Experiment 1. Manipulating the compatibility of task-irrelevant bodies did not interfere with the performance of a trait-based word categorization task, even in the presence of a demonstrably difficult secondary task. Instead, the speed at which traits were categorized was explained by the trait concepts themselves. We therefore show that the findings of Experiment 1, which support the notion that links between body shapes and trait concepts can be considered largely automatic, do not generalize to a context in which distractor bodies are present during the categorization of trait-based words. As a consequence, our findings cannot speak to the potential for data-limited processing as described earlier in the Discussion of Experiment 1, and as such leave open the possibility that forming links between other types of social information during body perception may place greater demands on executive resources.

Given the ambiguous outcome of Experiment 2, we sought to further contextualize our findings with a third experiment. The design of Experiment 2 differed in comparison to Experiment 1 in two main ways. Firstly, in contrast to Experiment 1, the task used in Experiment 2 did not require a judgement across a social dimension, instead requiring categorization of the trait word itself. Secondly, the trait words in Experiment 1 were displayed for a brief period before the bodies appeared, whereas both the target and distractor stimuli appeared at the same time in Experiment 2. Although this was designed to ensure participants had time to read the words (which was irrelevant when words became the target stimulus in Experiment 2), it is possible that this had a priming effect which is critical to the experimental outcomes seen in Experiment 1. These differences leave open the possibility that the outcomes of Experiment 2 reflect the activity of separate processes than those measured in Experiment 1, which may be agnostic to the high-level social connotations of trait-implying words. If so,

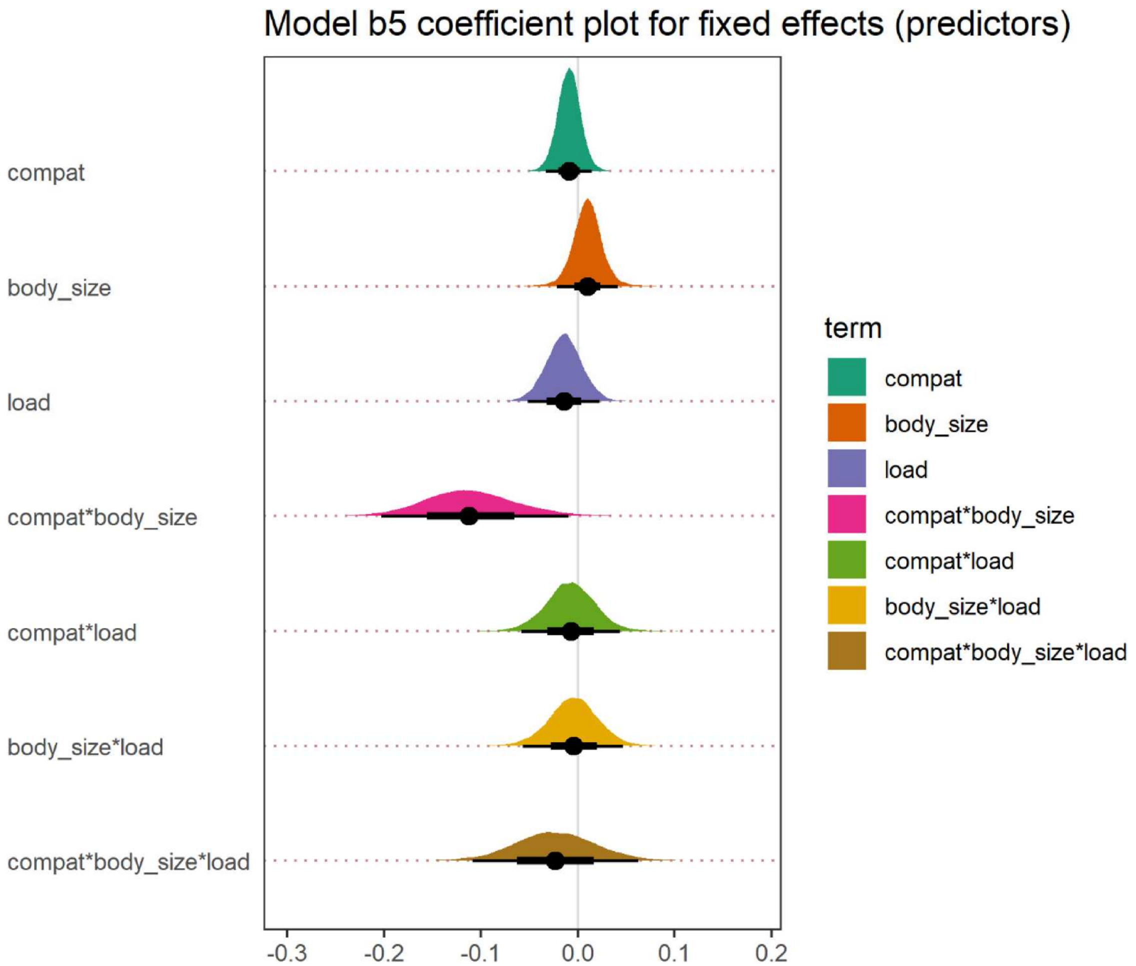

Figure 7. Parameter estimates for each predictor within Model 5.

Note: compat = compatible vs. incompatible; body_size = slim vs. heavy; load = high vs. low; x-axis = log(RT); point estimate = median; error bars represent 66% quantile intervals (thick black lines) and 95% quantile intervals (thin black lines). Interpretating these parameter estimates in terms of their original units is complex, as the shifted lognormal model is comprised of three components. To see estimates of these parameters in original units (milliseconds), please see Supplementary Figure S6.

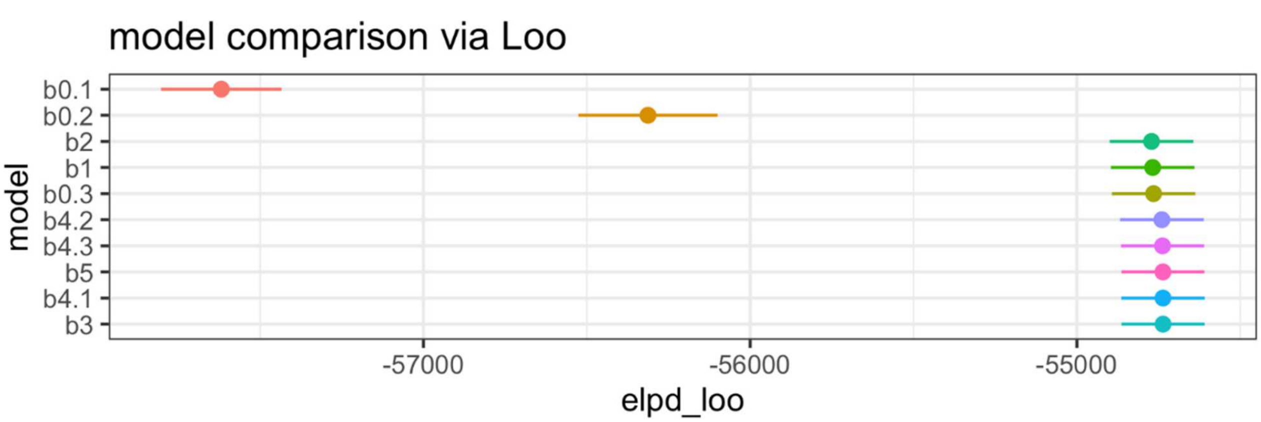

Figure 8. Model comparison (1–10 models).

Note: Model b0.1 included an overall intercept; Model b0.2 added varying intercepts for participants and stimulus items; Model b0.3 added a varying non-decision time parameter per participant; Model b1 included predictors for compatibility (compatible vs. incompatible); Model b2 included predictors for body size (slim vs. heavy); Model b3 included predictors for load (low vs. high); Model 4.1 included the interaction between compatibility and body size; Model 4.2 included the interaction between compatibility and load; Model 4.3 included the interaction between body size and load; Model b5 was the full model, and included the three-way interaction between compatibility, body size and load. Elpd_loo = estimate of the expected log pointwise predictive density; loo = leave-one-out estimated cross validation; error bars = standard error of the mean.

this makes it difficult to draw meaningful comparisons between Experiments 1 and 2.

Experiment 3

Our third experiment tested whether the presence of bodies would interfere with judgements about the valence of trait-implying words. As such, we addressed the same research question as Experiment 2 with a task that more closely resembled the design used in Experiment 1. To this end, using the same body stimuli and trait words as both previous experiments, participants were tasked with making speeded valence judgements (positive or negative) about visually presented words after being briefly primed with slim or heavy bodies. Our predictions for how RTs would vary across compatibility and load conditions were identical to the two previous experiments. Specifically, we anticipated a RT interference effect based on our compatibility factor, which would either increase or remain unaffected under high versus low cognitive load.

Method

Pre-registration and open science statement

As in our previous two experiments, our research question, hypotheses, planned analyses, and exclusion criteria were pre-registered (https://aspredicted.org/bn2qq.pdf). Raw data, stimuli, and analysis code are available online on the open science framework (https://osf.io/4en9f/). As data collection for our third experiment was carried out online, we anticipated a potentially high exclusion rate like Experiment 2 (which was also carried out online). As a result, we collected data from 75 participants, to increase the likelihood of having at least 50 sets of useable data.

Participants

75 participants were recruited online via Prolific in exchange for payment (75 males, $M_{age} = 30.90$, $SD_{age} = 5.81$). Following data pre-processing and exclusions, the final sample subjected to our analysis consisted of 70 participants.

Stimuli

All experimental stimuli were identical to those used in the previous two experiments.

Tasks

Overview. The experimental task for our third experiment was created in *PsychoPy (2022.2.1)* (Peirce et al., 2019), and again required participants to categorize trait-implying words in the context of task-irrelevant bodies (see Figure 2.). As in previous experiments, participants were given the opportunity to practice each task separately before practicing both together. This time, all participants were required to complete the final dual-task practice block twice and were optionally allowed to complete it a third time. The main experimental block consisted of 256 trials.

Trait categorization task. In the main task, participants were asked to judge whether a trait-implying word displayed on-screen was positively or negatively valenced, in the context of compatible and incompatible body shapes. Each trial began with a fixation cross, displayed in the center of the screen for 1000 ms. A slim or heavy body then appeared for 750 ms, before the target trait word appeared on screen until the participant responded (up to a maximum of 2000ms). Participants were asked to press "K" for positive trait words and "M" for negative trait words, and to respond as quickly and accurately as possible. Trials were presented in a fully randomized order, and RTs were recorded.

Working memory manipulation. The secondary working memory task was identical to that of Experiment 2.

Procedure

Participants recruited through Prolific were directed to the experiment hosted on Pavlovia. Consent and debrief forms were built into the experiment, which automatically redirected them back to Prolific and granted payment upon completion.

Design and data analysis

In line with our preregistration, and the analysis pipeline of the previous two experiments, we used Bayesian multi-level estimation to examine and draw inferences from our data. All models and parameters are defined identically to those used in the previous two experiments. Regarding exclusions, we again sought to retain as much trial-level data as possible while removing any cases likely to reflect disengagement from the task. As a result, we applied a set of performance-based exclusion criteria similar to those used in the previous two experiments, as well as attention checks which prevented disengaged participants from commencing the main experimental block of the task.

Attention checks appeared amongst a short set of general questions about the experiment, presented after the practice blocks, which were responded to on a 5-point likert scale. Two of these contained explicit instructions to select a specific response, and participants were allowed to proceed if they responded to at least one of these checks correctly. Following data collection, we evaluated performance on both the group and individual level to further filter out cases in line with our pre-registered criteria. Specifically, we excluded all trial-level responses with a RT of less than 20 ms, participants with valid responses to fewer than a quarter of the maximum possible trials for any cell of the design, and participants who appeared to have disengaged with the working memory task. To assess working memory responses, we identified participants who had achieved less than 55% accuracy in the working memory task and assessed their trial-by-trial response patterns. By visualizing their responses in chronological order, we were able to identify and remove participants who had responded to the task with the same keypress for large portions of the experiment. Two participants were removed based on these criteria. We also excluded trials on which (a) for a given word stimulus, a given participant had responded in the opposite manner to 80% of other trials containing that word stimulus, and (b) the RT for that trial was more than one standard deviation above that participants mean RT. In total, our exclusion criteria removed 5.46% of the total experimental data collected, including whole data for 5 participants.

Results

Manipulation check

The efficacy of our cognitive load manipulation was assessed in the same manner as our previous experiments. The mean difference in accuracy scores between load conditions was 21.03% [18.89, 23.18], $d_z = 2.34$ [1.90, 2.81]. As expected, accuracy was lower for high load (64.36%, [59.77, 82.16]), than low load (85.40%, [68.96, 88.63]). RT data also followed the expected pattern: mean difference = 72.60 ms [61.02, 84.15], $d_z = 1.50$ [1.16, 1.85], with responses faster for low load (567.73 ms, [548.81, 586.65]) than high load (640.31 ms, [618.06, 662.57]). Again, these effects were consistent across all remaining levels of the design (see Supplementary Figures S8 and S9).

Main task

Reaction times for the main task are visualized in Figure 9. Visual inspection of the data does not indicate the predicted pattern of interference, but instead follows that found in Experiment 2. Responses were always fastest for traits implying health and extraversion, regardless of the compatibility of the body shape stimulus. This pattern of results suggests that the results of Experiment 2 are unlikely to be due to the judgement dimension used by the task (deciding if a word is a trait or non-trait), as they translate to a judgement task which more closely resembles that used in Experiment 1.

Parameter estimates for the most complex model (Model 5), are shown in Figure 10 and Supplementary Table 1. Posterior distributions for most predictors overlap zero, except for load which indicates higher scores (i.e., longer reaction times) for high load. The compatibility*body size interaction term shows a similar trend to that seen in Experiment 2, again reflecting the pattern of faster responses to positively valenced trait words.

Model comparison is visualized in Figure 11. Like before, the models containing only fixed and varying intercepts (Model b0.1 and b0.2) underperformed all subsequent models, most of which overlapped in terms of out-of-sample predictive accuracy.

General discussion

In the present work, we tested the extent to which links between body perception and trait inferences are automatic, in the sense of being unaffected by a demonstrably difficult secondary task. Taken together, our findings support two novel conclusions. First, Experiment 1 shows that there are circumstances in which body shapes and trait concepts can be relatively automatically linked, in the sense that such links remain unaffected by a demonstrably difficult secondary task. These results suggest that linking trait concepts like “outgoing” and “lethargic” to visual body shape representations relies only minimally on cognitive

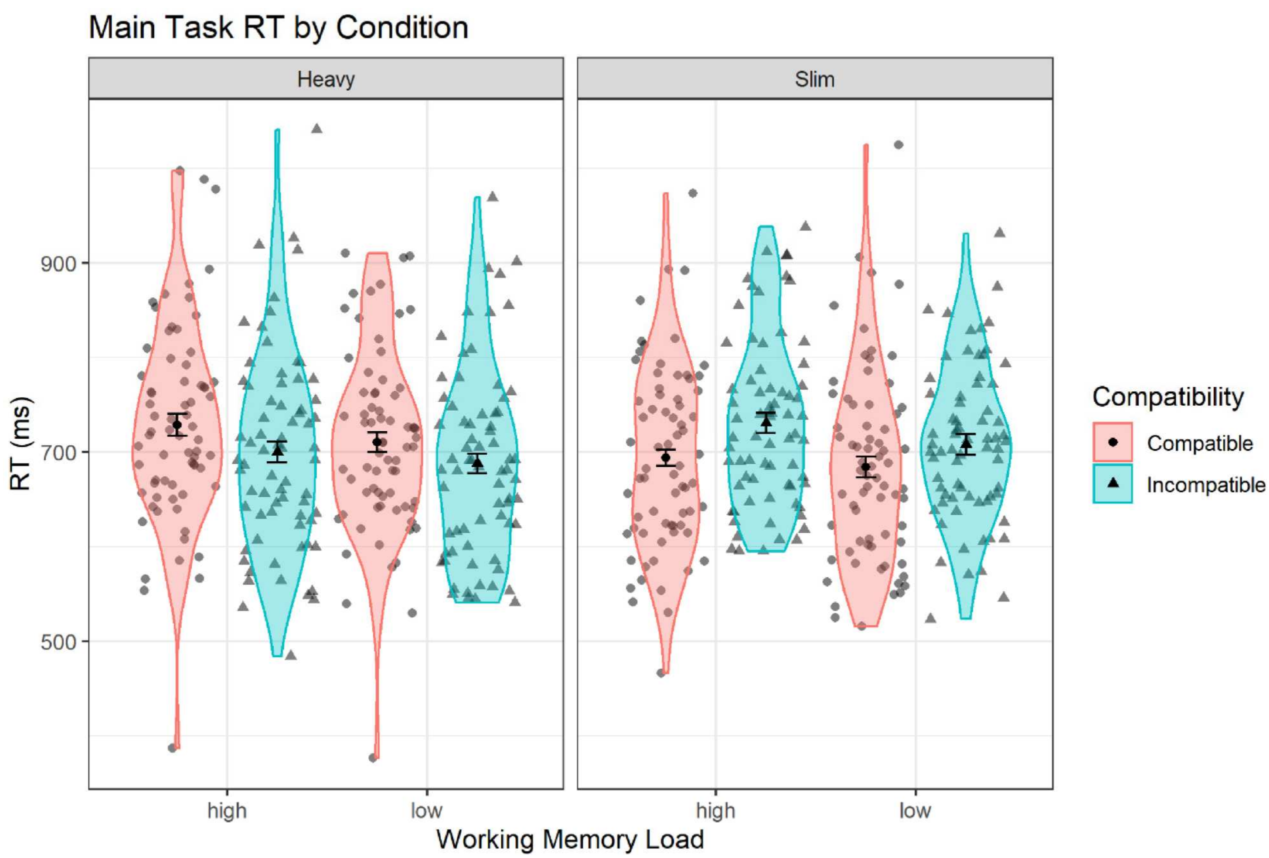

Figure 9. Mean RT scores per participant across the sample, plotted by load and body size conditions. Error bars represent 95% confidence intervals

control mechanisms within the central executive. On the other hand, Experiment 2 suggested that such resource-efficient integration between body perception and trait inference knowledge is not obligatory and may not always occur in every context. Experiment 3 supported this interpretation by repeating the null effect of compatibility in a task which relied on similar judgements to Experiment 1, making it unlikely that the findings of Experiment 2 can be explained by differences in the judgement dimension of the task. Overall, these findings update our understanding of the cognitive mechanisms that link different aspects of social perception and cognition together.

The present work extends previous findings regarding the relationship between perceptual and trait inferential processes in social cognition. Specifically, previous behavioral and neuroimaging studies have demonstrated links between body shape and inference processing when participants explicitly form trait inferences (Greven et al., 2016; Greven & Ramsey, 2017a, 2017b; Hu et al., 2018; Wildman & Ramsey, 2021). Here we add that, in some circumstances, such links occur spontaneously and do not rely on resource-intensive executive control. Instead, they appear to be linked in a manner that is largely invariant to the presence of a demanding secondary task that taxes executive resources. In this sense, we provide evidence that such links are efficient, and can therefore be considered more automatic with regards to this dimension of automaticity (Bargh, 1994).

On a theoretical level, evidence of automatic links between body shape and trait processing has implications for models of impression formation. Over and Cook (2018) present a model in which stereotypes and trait inferences rely on partially learned mappings between points or regions in distinct multidimensional spaces that encompass variation in visual (e.g., face/body space), and trait representations (trait space). Here, we argue that such mappings between body space and trait space largely bypass the central executive or only have a light exchange in terms of the consumption of resources.

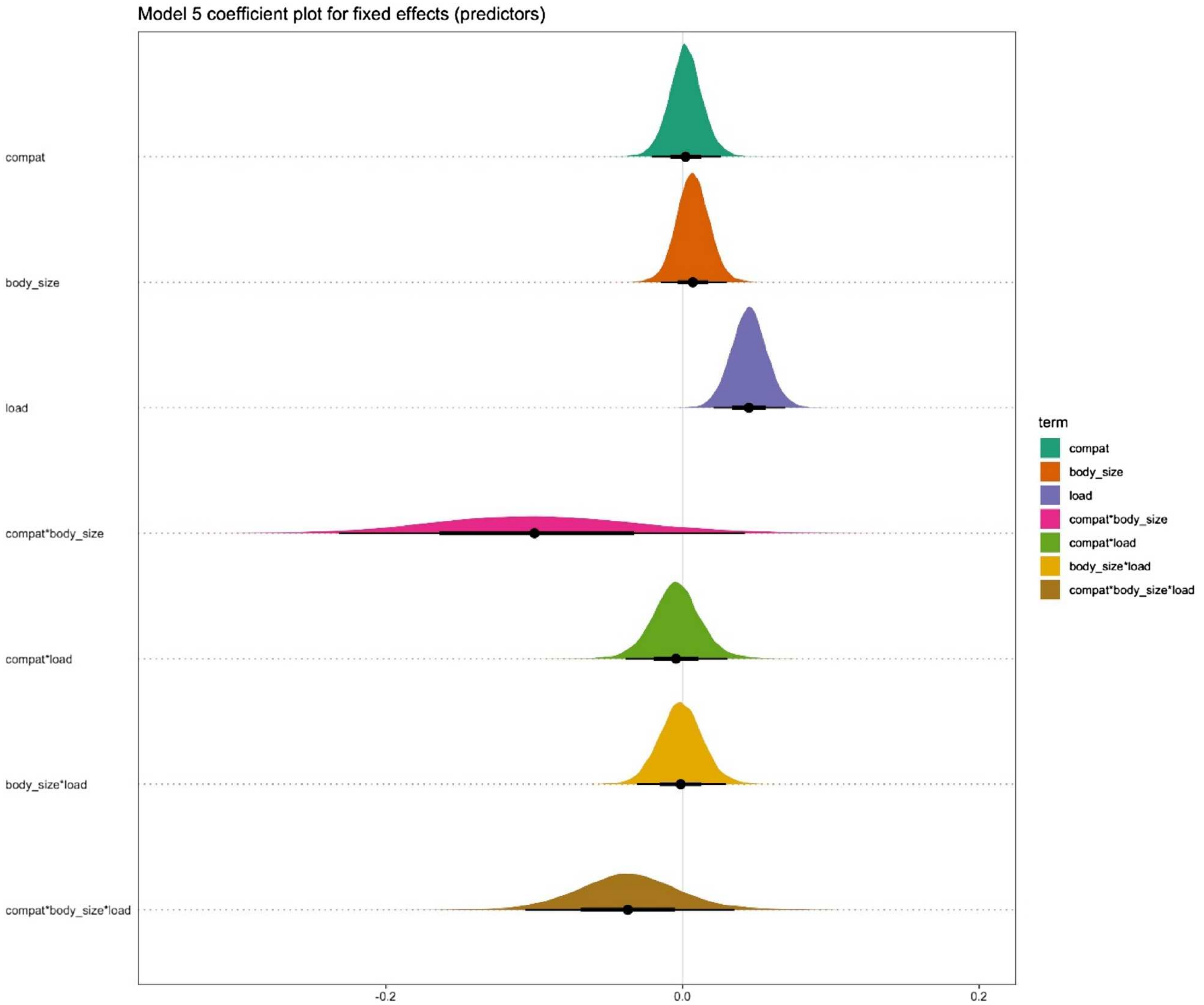

Figure 10. Parameter estimates for each predictor within Model 5.

Note: compat = compatible vs. incompatible; body_size = slim vs. heavy; load = high vs. low; x-axis = log(RT); point estimate = median; error bars represent 66% quantile intervals (thick black lines) and 95% quantile intervals (thin black lines). Interpretating these parameter estimates in terms of their original units is complex, as the shifted lognormal model is comprised of three components. To see estimates of these parameters in original units (milliseconds), please see Supplementary Figure S10.

Constraints on generality

Our findings demonstrate that some, but not all, circumstances trigger spontaneous links between body shape and trait representations. The results of Experiment 2 and 3 showed no indication of an interference effect between bodies and traits, suggesting that such links were either not formed or were not measured by our design. Assuming that links were not formed at all, it is possible that the underlying process is directional such that spontaneous and resource-light links occur in one direction (between traits and bodies) but not the other (between bodies and traits). One possible theoretical interpretation of this pattern is that the perception and evaluation of bodies necessarily involves the extraction of trait information, and therefore the pattern of interference found in Experiment 1, whereas evaluating trait concepts does not require the perceiver to generate body shape representations. This lack of symmetry makes sense from the perspective that traits can be considered attributes of body shapes, but not vice versa.

Limitations

While our main task in Experiment 1 required participants to discriminate between two distinct body sizes, the constrained presentation of bodies may

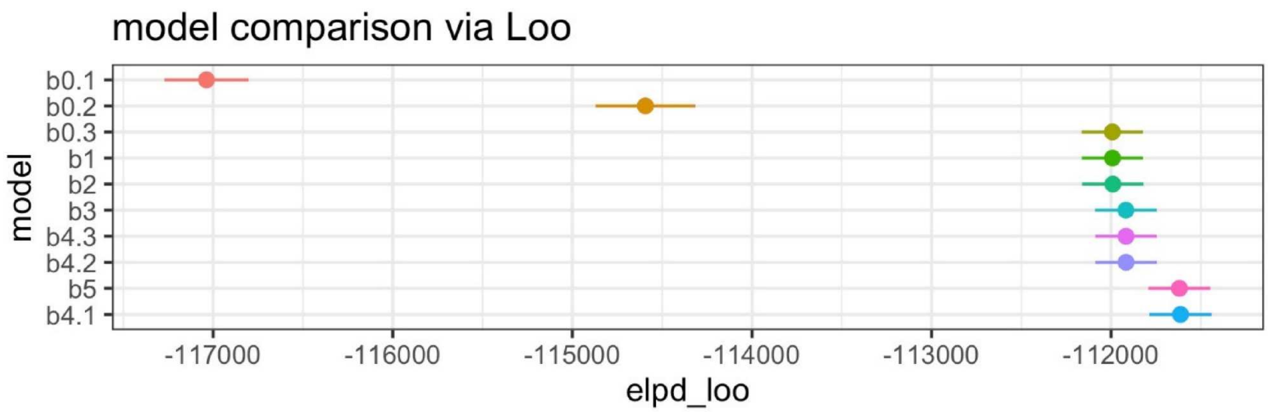

Figure 11. Model comparison (1–10 models).

Note: Model b0.1 included an overall intercept; Model b0.2 added varying intercepts for participants and stimulus items; Model b0.3 added a varying non-decision time parameter per participant; Model b1 included predictors for compatibility (compatible vs. incompatible); Model b2 included predictors for body size (slim vs. heavy); Model b3 included predictors for load (low vs. high); Model 4.1 included the interaction between compatibility and body size; Model 4.2 included the interaction between compatibility and load; Model 4.3 included the interaction between body size and load; Model b5 was the full model, and included the three-way interaction between compatibility, body size and load. Elpd_loo = estimate of the expected log pointwise predictive density; loo = leave-one-out estimated cross validation; error bars = standard error of the mean.

have limited the processing of body features beyond those directly subserving a decision based on overall body weight. As a result, our findings cannot directly support claims regarding the spontaneity of mappings between other dimensions of body shape and trait inference. Also, the format of catch trials in Experiment 1 is likely to have directed some degree of attention to our trait stimuli, and it is possible that this had some influence on our findings.

The findings of Experiments 2 and 3 present a clear case that the effect found in Experiment 1 does not transfer to a similar experimental context in which the roles of target and distractor are reversed. However, this may reflect a directionality specific to the experimental task rather than the underlying cognitive processes which produce the effect. For example, the types of stimuli used across the experiments meant that judgements directed at body size concerned specific individual on-screen, whereas judgements of traits concerned either the linguistic function or affective qualities of broad concepts denoted by written words. Therefore, it is possible that the asymmetry in outcomes between experiments is not based on an inherent directionality in the modalities and system of information processing involved, but in the degree to which the task prompted person-judgements. Another account of the asymmetry in outcomes concerns the modalities of stimuli and responses in the tasks, and the degree to which different tasks required participants to "translate" information from one modality to another. For example, categorizing a body as "slim" may demand that shape information is translated into a word which codes for that shape, exposing the process to competing information from other word-based information. In contrast, the tasks in Experiments 2 and 3 required participants to categorize words into word-based categories, which could have occurred without translating the bodies into word-based codes.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Andrew Wildman http://orcid.org/0000-0002-9433-1120
Richard Ramsey http://orcid.org/0000-0002-0329-2112

References

Aviezer, H., Trope, Y., & Todorov, A. (2012). Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science (New York, N.Y.)*, *338*(6111), 1225–1229. https://doi.org/10.1126/science.1224313

Bara, I., Binney, R. J. (2023). Investigating the role of working memory resources across aesthetic and nonaesthetic judgements. *Quarterly Journal of Experimental Psychology*, *76*(5), 1026–1044.

Bara, I., Darda, K. M., Kurz, A. S. (2021). Functional specificity and neural integration in the aesthetic appreciation of artworks with implied motion. *European Journal of Neuroscience*, *54*(9), 7231–7259.

Bargh, J. (1994). The four horsemen of automaticity: Awareness, efficiency, intentions and control. In R. Wyer & T. Srull (Eds.), *Handbook of social cognition* (pp. 1040). Lawrence Erlbaum.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bürkner, P.-C. (2017). Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, *10*(1), 395–411. https://doi.org/10.32614/RJ-2018-017

de Fockert, J. W., de Rees, G., Frith, C. D., & Lavie, N. (2001). The role of working memory in visual selective attention. *Science*, *291*, 1803–1806. https://doi.org/10.1126/science.1056496

de Gelder, B. (2006). Towards the neurobiology of emotional body language. *Nature Reviews Neuroscience*, *7*(3), 242–249. https://doi.org/10.1038/nrn1872

de Gelder, B., Van den Stock, J., Meeren, H. K. M., Sinke, C. B. A., Kret, M. E., & Tamietto, M. (2010). Standing up for the body. Recent progress in uncovering the networks involved in the perception of bodies and bodily expressions. *Neuroscience & Biobehavioral Reviews*, *34*(4), 513–527. https://doi.org/10.1016/j.neubiorev.2009.10.008

Downing, P. E., & Peelen, M. V. (2011). How might occipitotemporal body-selective regions interact with other brain areas to support person perception? *Cognitive Neuroscience*, *2*(3-4), 216–226. https://doi.org/10.1080/17588928.2011.613987

Engell, A. D., Haxby, J. V., & Todorov, A. (2007). Implicit trustworthiness decisions: Automatic coding of face properties in the human amygdala. *Journal of Cognitive Neuroscience*, *19*(9), 1508–1519. https://doi.org/10.1162/jocn.2007.19.9.1508

Frith, C. D., & Frith, U. (1999). Interacting minds—A biological basis. *Science*, *286*(5445), 1692–1695. https://doi.org/10.1126/science.286.5445.1692

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, *1*(3), 515–534. https://doi.org/10.1214/06-BA117A

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. Chapman and Hall/CRC. https://doi.org/10.1201/b16018

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464.–1480 https://doi.org/10.1037/0022-3514.74.6.1464

Greven, I. M., Downing, P. E., & Ramsey, R. (2016). Linking person perception and person knowledge in the human brain. *Social Cognitive and Affective Neuroscience*, *11*(4), 641–651. https://doi.org/10.1093/scan/nsv148

Greven, I. M., Downing, P. E., & Ramsey, R. (2018). Neural networks supporting social evaluation of bodies based on body shape. *Social Neuroscience*, *14*(3), 328–344. https://doi.org/10.1080/17470919.2018.1448888

Greven, I. M., & Ramsey, R. (2017a). Person perception involves functional integration between the extrastriate body area and temporal pole. *Neuropsychologia*, *96*, 52–60. https://doi.org/10.1016/j.neuropsychologia.2017.01.003

Greven, I. M., & Ramsey, R. (2017b). Neural network integration during the perception of in-group and out-group members. *Neuropsychologia*, *106*, 225–235. https://doi.org/10.1016/j.neuropsychologia.2017.09.036

Haines, N., Kvam, P. D., Irving, L. H., Smith, C., Beauchaine, T. P., Pitt, M. A., Ahn, W.-Y., & Turner, B. M. (2020). Theoretically informed generative models can advance the psychological and brain sciences: Lessons from the reliability paradox. *PsyArXiv*. https://doi.org/10.31234/osf.io/xr7y3

Hayward, D. A., & Ristic, J. (2013). The uniqueness of social attention revisited: Working memory load interferes with endogenous but not social orienting. *Experimental Brain Research*, *231*(4), 405–414. https://doi.org/10.1007/s00221-013-3705-z

Hehman, E., Sutherland, C. A. M., Flake, J. K., & Slepian, M. L. (2017). The unique contributions of perceiver and target characteristics in person perception. *Journal of Personality and Social Psychology*, *113*(4), 513–529. https://doi.org/10.1037/pspa0000090

Hu, Y., Parde, C. J., Hill, M. Q., Mahmood, N., & O'Toole, A. J. (2018). First impressions of personality traits from body shapes. *Psychological Science*, *29*(12), 1969–1983. https://doi.org/10.1177/0956797618799300

Kauschke, C., Bahn, D., Vesker, M., & Schwarzer, G. (2019). The role of emotional valence for the processing of facial and verbal stimuli—positivity or negativity bias? *Frontiers in Psychology*, *10*, 1654. https://doi.org/10.3389/fpsyg.2019.01654

Kay, M. (2020). *tidybayes: Tidy data and 'geoms' for Bayesian models*. http://mjskay.github.io/tidybayes/

Konstantinou, N., Beal, E., King, J.-R., & Lavie, N. (2014). Working memory load and distraction: Dissociable effects of visual maintenance and cognitive control. *Attention, Perception, & Psychophysics*, *76*(7), 1985–1997. https://doi.org/10.3758/s13414-014-0742-z

Kuperman, V., Estes, Z., Brysbaert, M., & Warriner, A. B. (2014). Emotion and language: Valence and arousal affect word recognition. *Journal of Experimental Psychology: General*, *143*(3), 1065–1081 https://doi.org/10.1037/a0035669

Kurz, A. S. (2020). *Statistical Rethinking with brms, ggplot2, and the tidyverse*. https://bookdown.org/ajkurz/Statistical_Rethinking_recoded/

Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in Cognitive Sciences*, *9*(2), 75–82. https://doi.org/10.1016/j.tics.2004.12.004

Lavie, N. (2010). Attention, distraction, and cognitive control under load. *Current Directions in Psychological Science*, *19*(3), 143–148. https://doi.org/10.1177/0963721410370295

Lavie, N., & De Fockert, J. (2005). The role of working memory in attentional capture. *Psychonomic Bulletin & Review*, *12*(4), 669–674. https://doi.org/10.3758/BF03196756

Lavie, N., Hirst, A., de Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General*, *133*(3), 339–354.https://doi.org/10.1037/0096-3445.133.3.339

Lemoine, N. P. (2019). Moving beyond noninformative priors: Why and how to choose weakly informative priors in Bayesian analyses. *Oikos*, *128*(7), 912–928. https://doi.org/10.1111/oik.05985

McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and stan* (2nd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9780429029608.

Mitchell, J. P. (2009). Inferences about mental states. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364* (1521), 1309–1316. https://doi.org/10.1098/rstb.2008.0318

Mitchell, J. P., Cloutier, J., Banaji, M. R., & Macrae, C. N. (2006). Medial prefrontal dissociations during processing of trait diagnostic and nondiagnostic person information. *Social Cognitive and Affective Neuroscience*, *1*(1), 49–55. https://doi.org/10.1093/scan/nsl007

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(1), 1–9. https://doi.org/10.1038/s41562-016-0021

Naumann, L., Vazire, S., Rentfrow, P., & Gosling, S. (2009). Personality judgments based on physical appearance. *Personality and Social Psychology Bulletin*, *35*(12), 1661–1671. https://doi.org/10.1177/0146167209346309

Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology*, *7*(1), 44–64. https://doi.org/10.1016/0010-0285(75)90004-3

Over, H., & Cook, R. (2018). Where do spontaneous first impressions of faces come from? *Cognition*, *170*, 190–200. https://doi.org/10.1016/j.cognition.2017.10.002

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Puhl, R. M., & Heuer, C. A. (2009). The stigma of obesity: A review and update. *Obesity*, *17*(5), 941–964. https://doi.org/10.1038/oby.2008.636

Ramsey, R. (2018). Neural integration in body perception. *Journal of Cognitive Neuroscience*, *30*(10), 1442–1451. https://doi.org/10.1162/jocn_a_01299

Ramsey, R., Darda, K. M., & Downing, P. E. (2019). Automatic imitation remains unaffected under cognitive load. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(5), 601–615. https://doi.org/10.1037/xhp0000632

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in "theory of mind". *NeuroImage*, *19*(4), 1835–1842. https://doi.org/10.1016/S1053-8119(03)00230-1

Todorov, A., & Engell, A. D. (2008). The role of the amygdala in implicit evaluation of emotionally neutral faces. *Social Cognitive and Affective Neuroscience*, *3*(4), 303–312. https://doi.org/10.1093/scan/nsn033

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition*, *27*(6), 813–833. https://doi.org/10.1521/soco.2009.27.6.813

Todorov, A., & Uleman, J. S. (2003). The efficiency of binding spontaneous trait inferences to actors' faces. *Journal of Experimental Social Psychology*, *39*(6), 549–562. https://doi.org/10.1016/S0022-1031(03)00059-3

van Overwalle, F. (2009). Social cognition and the brain: A meta-analysis. *Human Brain Mapping*, *30*(3), 829–858. https://doi.org/10.1002/hbm.20547

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

Wang, Y. C., McPherson, K., Marsh, T., Gortmaker, S. L., & Brown, M. (2011). Health and economic burden of the projected obesity trends in the USA and the UK. *Lancet (London, England)*, *378*(9793), 815–825. https://doi.org/10.1016/S0140-6736(11)60814-3

Wickham, H., & Grolemund, G. (2016). *R for data science*. O'Reilly Media.

Wildman, A., & Ramsey, R. (2021). Estimating the effects of trait knowledge on social perception. *Quarterly Journal of Experimental Psychology*, *75*, 969–987. https://doi.org/10.1177/17470218211047447

Zimmer, U., & Macaluso, E. (2007). Processing of multisensory spatial congruency can be dissociated from working memory and visuo-spatial attention. *European Journal of Neuroscience*, *26*(6), 1681–1691. https://doi.org/10.1111/j.1460-9568.2007.05784.x

Zopf, R., Contini, E., Fowler, C., Mondraty, N., & Williams, M. A. (2016). Body distortions in Anorexia Nervosa: Evidence for changed processing of multisensory bodily signals. *Psychiatry Research*, *245*, 473–481. https://doi.org/10.1016/j.psychres.2016.09.003