FRONT MATTER

Title

- Full title: Human Ageing is Associated with More Rigid Concept Spaces
- Short title: Concept Change in Older Adults

Authors

Sean Devine [1, 2]*, Cassandra Neumann [1], David Levari [3], Robert C. Wilson [4, 5, 6], Ben Eppinger [1, 7, 8]

* seandamiandevine@gmail.com

Affiliations

[1] Department of Psychology, Concordia University, Montreal, Canada

[2] Department of Psychology, McGill University, Montreal, Canada

[3] Department of Psychology, Harvard University, Cambridge, USA

[4] Department of Psychology, University of Arizona, Tucson, USA

[5] Cognitive Science Program, University of Arizona, Tucson, USA

[6] Evelyn F. McKnight Brain Research Foundation, University of Arizona, Tucson, USA

[7] PERFORM Center, Concordia University, Montreal, Canada

[8] Lifespan Developmental Neuroscience Department, Technische Universität Dresden, Germany

Abstract

Prevalence-induced concept change describes a cognitive mechanism by which someone's definition of a concept shifts as the prevalence of instances of that concept changes. The phenomenon has real-world implications because this sensitivity to environmental characteristics may lead to substantial biases in judgements. While prevalence-induced concept change has been established in young adults, it is unclear how it changes as a function of human ageing. In this cross-sectional study, we explore how prevalence-induced concept change affects older adults' lower-level, perceptual, and higher-order, ethical, judgements. We find that older adults are less sensitive to prevalence-induced concept change than younger adults across domains. Using a combination of computational and experimental approaches, we demonstrate that these changes in judgements are sensitive to the pace with which the stimuli occur in the environment and are affected by the effort that subjects invest in order to make accurate decisions. Based on findings from three experiments we argue that older adults' concept spaces are more rigid than those of younger adults. However, what appear as an age-related cognitive "deficit" may turn out to be beneficial because it makes older adults less susceptible to biases in judgments.

"*The more things change, the more they stay the same.*"

— Jean-Baptiste Alphonse Karr and, later, Jon Bon Jovi

By 2068, almost 30% of the North-American population will be 65 years or older (Statistics Canada, 2019; U.S. Census Bureau, 2018). As adults age, their judgments and decisions will affect society more than ever before and will largely influence our collective future. As such, it is critical to understand how the cognitive and motivational processes underlying judgement and decision-making change with age and how these changes may affect real-world decisions.

In this study, we explore how changes in one's environment affect concept formation and judgements in younger and older adults. Specifically, we consider how judgements about perceptual and ethical concepts are affected by the prevalence of instances of them in the environment. This phenomenon has been referred to as **prevalence-induced concept change** (Levari et al., 2018). Prevalence-induced concept change describes the empirical observation that as the numbers of instances of a given concept change in the environment, so do the boundaries for that concept, such that they come to include instances that they would otherwise exclude. For example, one task that measures prevalence-induced concept change requires participants to serially judge whether individual dots that vary on a spectrum between blue and purple are in fact blue or purple. When the relative frequency of objectively coloured dots in the environment is consistent across the task (50% blue 50% purple dots), peoples' judgements are relatively stable: blue dots are judged as blue and purple dots as purple. However, if the number of blue dots changes, dots initially judged as purple are later (after the prevalence changes) categorised as blue. Interestingly these changes do not only occur on the perceptual level, but also arise in higher-order social and ethical judgements. (Levari et al., 2018).

What these findings suggest is that, from judgement to judgement, people adjust their concepts to environmental characteristics. In line with this view, recent work suggests that the cognitive mechanisms underlying prevalence-induced concept change can be captured using computational modeling (Wilson, 2018). In this model concept boundaries arise from competition between the effect of the past stimulus and the effect of the past response on the judgement; the former increasing prevalence-induced concept and the latter reducing it.

Why would we expect ageing-related changes in this regard then? Many cognitive changes that come with healthy ageing offer good reason to suspect that older adults might be differentially affected by prevalence-induced concept change compared to younger adults. Several empirical findings suggest that older adults differ from young adults in terms of motivation, postponement of gratification, and to what degree they value desired outcomes (Mather & Harley, 2016; Samanez-Larkin & Knutson, 2015; Eppinger et al., 2011). These differences in decision-making processes can in part be explained by age-related differences in cognitive ability, such as changes in executive function (Mayr, Spieler, & Kliegl, 2001), memory (Nyberg et al., 2012), and processing speed (Salthouse, 1992; 1996). Specifically, two lines of research paint opposing pictures of how older adults might make different concept judgements than younger adults when the prevalence of instances of a concept in the environment changes.

On the one hand, previous work suggests that older adults have more difficulty learning from uncertain outcomes compared to younger adults (Nassar et al., 2016). This difficulty manifests as perseverative behaviour, whereby older adults have a tendency to repeat previous responses despite changes in the environment (Bruckner et al., 2020; Eppinger, Walter, Heekeren, & Li, 2013). With respect to the current study, these findings may suggest a decreased sensitivity to prevalence-induced concept change in older adults, because the repetition of past choices makes it less likely that a rarer category will be chosen after a shift in prevalence. The same can also be expressed computationally, where perseverative behaviour is reflected by a higher influence of past choice, which has previously shown to reduce prevalence-induced concept change (Wilson, 2018).

On the other hand, results from several recent studies suggest that older adults have difficulty converging on an accurate representation of the current state, particularly if these states are latent (not directly observable) and need to be inferred from experience (Hämmerer et al., 2019; Hämmerer, Müller, & Li, 2014; Eppinger, Heekeren, & Li, 2015). To help compensate for this difficulty in distinguishing task states, older adults may outsource control to the environment rather than relying on (sometimes inaccurate) internal representations (Mayr, Spieler, & Hutcheon, 2015; Spieler, Mayr, & LaGrone, 2006; Lindenberger & Mayr, 2015). In the case of prevalence-induced concept change, this outsourcing of control is likely to lead to increased concept change. From a computational perspective, this tendency to outsource also represents an increased sensitivity to the effect of previous stimulus, which has similarly been shown to increase prevalence-induced concept change (Wilson, 2018).

Taken together, the results of the aforementioned studies point to two opposing hypotheses: Hypothesis 1 suggests that older adults are **less** sensitive to prevalence-induced concept change than younger adults, whereas Hypothesis 2 suggests that older adults are **more** sensitive to prevalence-induced concept change than younger adults. To visualize our predictions we used the computational model by Wilson (2018) to simulate data for each of the hypotheses (see Figure 1).

To tease these hypotheses apart and gain a better understanding of the cognitive mechanisms underlying age-related changes in prevalence-induced concept change we ran three experiments. In the first experiment, we use an age-comparative study design and a computational model to investigate age differences in prevalence-induced concept change in lower-level, perceptual, and higher-level, moral, judgements. We show that, across domains, older adults are less susceptible to prevalence-induced concept change than younger adults. In the second step, we explore two potential explanations for these age-related changes in prevalence-induced concept change by manipulating concept formation in younger adults. In Experiment 2, we vary the intertrial intervals (ITI) between occurrences of stimuli and show that a greater spacing of stimuli reduces prevalence-induced concept change in younger adults (akin to the findings observed in older adults). In Experiment 3, we provide incentives for consistent judgments and show that this manipulation also shifts younger adults' judgments towards the behavior observed in older adults. Across experiments, our findings suggest that age-related changes in prevalence-induced concept change may reflect a combination of changes in the timing between stimuli and higher-level changes in the motivation for accurate judgements.

Experiment 1

Method

Participants

We recruited 160 participants from the community and the university participation pool, 80 of which were older adults (M_{age} = 70.10; s_{age} = 5.55) and 80 of which were younger adults (M_{age} = 21.85; s_{age} = 2.27). All participants were English-speaking, free of neurological or psychiatric disorders, and free of any cognitive, motor, visual, or other condition(s) that would impede their performance, including but not limited to a history of head trauma with loss of consciousness, organic brain disorders, seizures, or neurosurgical intervention, to sensory deficits (i.e. deafness, blindness, colour blindness, intellectual disability), or self-reported cognitive impairment, and to a recent history of substance abuse. In each age group, 40 participants were randomly assigned to either the decreasing prevalence condition or the stable prevalence condition, in a counterbalanced order. In the former, they experienced a decreasing prevalence of instances of the concept in both tasks detailed below. In the latter, the prevalence remained the same throughout the entire experiment. All participants were compensated $20 CAN or 2 participation pool credits for participating in the study. The study protocol was approved by the Concordia Human Research Ethics Committee (certification number 30011191). In this experiment and the two others detailed below, sample sizes were based on Levari et al. (2018) data and the stopping rule for data collection was to collect until the end of the semester and group sizes were even.

Materials

The Dots Task. In the Dots Task, participants had to judge the colour of an individual dot presented on the screen. The task began with a series of instruction screens explaining the task to the participant. These instructions were followed by a practice block consisting of 10 trials, in which participants could familiarize themselves with the task. These trials were identical to trials in the real task and consisted of 50% purple dots and 50% blue dots. Data from practice trials were not analysed.

After the practice block, participants performed a total of 800 trials, divided into 16 blocks of 50 trials each. In the decreasing prevalence condition, the number of blue dots in the environment decreased as the number of blocks increased in a predetermined fashion. Specifically, the proportion of blue relative to purple dots, was as follows for each of the 16 blocks: .50, .50, .50, .50, .40, .28, .14, .06, .06, .06, .06, .06, .06, .06, .06, .06. In the stable prevalence condition, the proportion of blue dots in the environment remained the same (.50) across the experiment. In both cases, blue dots were defined as any dot who’s RGB value was between [0, 0, 254] and [49, 0, 205]. Purple dots were defined as any dot who’s RGB value was between [50, 0, 204] and [99, 0, 155]. Dot colours were randomly chosen for each trial based on the number of trials per block (50) and the frequency with which blue and purple dots should appear in a given block.

In each trial, participants judged the color of the dots as being either blue or purple by pressing the ‘A’ or ‘L’ key on the keyboard. All stimuli were presented against a dark grey background. Each trial started with a dot presented on the screen for 500 ms, followed by a question mark that appeared on the screen until participants made a choice, and finally a blank screen appeared for 500 ms as an ITI. Thus, all timing was fixed across participants, except that which would arise from differences in response times. After each block text appeared that indicated that the block was finished, which block the participant was now at, and offering them a short break should they choose to take one.

The Ethics Task. In the Ethics Task, participants had to take on the role of a member of an Ethics Review Board and judge whether fictitious research proposals were ethical or not (phrased as whether they would allow these research studies to be conducted or not). All research proposals were norm tested by Levari et al. (2018, see Supporting Online Material) to produce scores depicting how ethical people found the 273 proposals. These scores were used to bin proposals as unethical (80 proposals), ethical (113 proposals), or ambiguous (80 proposals). These bins were used to calculate the proportion of proposals that appeared in each block (including the practice trial). Just as in the Dots Task, participants were first presented with instruction screens explaining the task to them. Following the instructions, participants completed a practice trial in which they judged a research proposal using the keyboard keys. In this task, they pressed 'A' when they would not allow a study to be conducted and 'L' when they would.

Following the practice trial, participants began the test trials. All proposals in the experiment were presented in black text against a dark grey background. The task consisted of 240 trials broken into 10 blocks. In the decreasing prevalence condition, the proportion of unethical, ethical, and ambiguous proposals varied across blocks. Specifically, for the 10 blocks of the study, the proportion of unethical proposals relative to ethical and ambiguous proposals were as follows: .33, .33, .33, .33, .25, .17, .08, .04, .04, .04 (rounded to the nearest 2^{nd} decimal). In the stable prevalence condition, the proportion between the three types of proposals was the same throughout the task: .33.

Each trial, participants read a proposal and pressed 'A' or 'L' on the keyboard indicating whether they thought that the research should be allowed to be conducted on people or not. There was no time limit on this choice. Following the choice, a fixation cross appeared on the screen for 500 ms, followed by the next proposal. Between each block, text appeared that indicated that the block was finished, which block the participant was now at, and offering them a short break should they choose to take one.

Both the Dots and Ethics Tasks described above were taken from Levari et al. (2018). Both tasks were programmed in Python using the PsychoPy libraries.

Procedure

Participants were recruited from the community through online or paper advertisements or from Concordia's participation pool. Participants were contacted by telephone or email and were asked basic demographic information to determine initial eligibility. If eligible at this stage, they were invited for a single two-hour session in the lab.

Once at the lab, participants were asked to fill out a consent form and complete the Richmond HRR pseudoisochromatic test for colour vision (Cole, Lian, & Lakkis, 2016; see Supplement for more details). Participants were then asked to complete the Dots Task and Ethics Task, back-to-back. The order of these tasks was counterbalanced across participants. They were told that they would be free to take short breaks during the tasks (between blocks) and a longer break between the tasks, should they choose to. After completing both tasks, participants were debriefed and paid $20 for participating or were given their participation credits.

Computational Model

We used a sequential decision-making model based upon logistic regression to explore prevalence-induced concept change on a trial-by-trial basis (Wilson, 2018):

$$p_t = 1 - \frac{1}{1 + exp\left(\beta_0 + \beta_f f_t + \beta_F F_t + \beta_c C_t\right)} \qquad \text{(Eq. 1)}$$

In this model, p_t is the probability of classifying the current stimulus as blue or unethical, β_0 captures the overall bias for classifying the stimulus as blue or unethical, β_f captures the effect of the current stimulus, β_F captures the effect of the past stimulus, and β_c captures the effect of past response. F_t and C_t represent the exponentially weighted sum of past stimuli and past response respectively. These parameters are controlled by two other parameters, λ_F and λ_C, which dictate the rate of decay of the exponential weighting with larger values corresponding to slower decay. They follow the following update rules:

$$F_t = \lambda_F F_t + f_t \qquad \text{(Eq. 2)}$$

$$C_t = \lambda_C\, C_t + c_t \qquad \text{(Eq. 3)}$$

This leaves six free parameters to be estimated (β_0, β_f, β_F, β_C, λ_F and λ_C). These parameters can be estimated using a standard maximum likelihood approach, where the parameters that produce the maximum sum of log scores is taken (in practice, the minimum negative logged sum; see Daw, 2011). Of these parameters, β_F and β_C are of the most theoretical interest. As β_F decreases, the effect of previous stimuli becomes stronger and biases choice behaviour to the opposite of the previous stimulus (e.g., if the previous dot was purple, the subsequent choice would be blue). On the other hand, as β_C increases, responses become biased to match the previous response (e.g., if one responded that the previous dot was blue, they would do so again on their subsequent choice). Thus, according to Wilson (2018), the degree to which one is sensitive to prevalence-induced concept change arises as a function of the relative strengths of these opponent parameter weights.

Simulation of Predicted Results

Based on our hypotheses, we simulated data using the computational model detailed above. Specifically, we simulated data for three scenarios. First, H0 predicted that older and younger adults would not differ in their sensitivity to prevalence-induced concept change. To simulate this scenario, we imputed similar parameters as found in healthy young adults into the computational model (parameters from Wilson, 2018). Second, H1 predicted that older adults would be less sensitive to prevalence-induced concept change than young adults.

Computationally, we simulated this scenario by inputting greater β_C values into the model (greater effect of past response). Finally, H2 predicted that older adults would be more sensitive to prevalence-induced concept change than young adults, which we simulated by inputting lower β_F values into the model. The results of these simulations are summarised in Figure 1.

Qualitatively, these simulations demonstrate the pattern of results we expect in accordance with each hypothesis, as well as a rough estimate of the parameter values we think would underlie participants' behaviour in each scenario.

Statistical Analysis

All data were analysed in R (version 3.6.1). For the Ethics task normed scores were reversed, to make the plots in the same direction as the Dots Task, such that lower normed scores now represented more ethical scenarios. The main analysis consisted of six general binomial mixed-effects models that were implemented and fit using the *lme4* package (Bates, Maechler, Bolker, & Walker, 2015). The main models in each task predicted response using age group (young adult or older adult), condition (stable prevalence or decreasing prevalence), trial number, and stimulus strength (colour in the Dots Task and normed ethicality scores in the Ethics Task) as fixed effects, a random slope of trial, and a random intercept for each participant. All main effects and interactions were explored. Two follow-up models were conducted in both tasks using the same predictors, split by age group. In all statistical models, trial and stimulus strength were put on a scale between zero and one. Model weights were estimated using the *nlminb* compared across age groups and conditions using a between-groups 2x2 ANOVA.

Finally, we analysed response times using two between-groups 2x2 (age group × condition) ANOVA, one for each task. In line with best practices regarding HARKing (Hollenback & Wright, 2017), we wish to disclose that these analyses were exploratory and not based on original hypotheses. To supplement these exploratory analyses, we conduct two follow-up studies, presented in more detail below.

Results

Choice Data

Statistically-speaking, prevalence-induced concept change is reflected in a three-way interaction between condition, trial, and stimulus strength, predicting responses. The effect size of this interaction reflects the degree to which a participant's choice to categorize a given exemplar (dot or research proposal) as one concept or another is influenced by a combined effect of three factors: (a) the prevalence of instances in the environment (i.e., the effect of condition), (b) the amount of time that has past (i.e., the effect of trial), and (c) the strength of the stimulus (i.e., blueness or ethicality). Thus, if younger and older adults differ in their sensitivity to prevalence-induced concept change, we would expect to see a four-way interaction between these three terms above and age group, as well as different effect sizes for the three-way effect within each of the age groups.

Indeed, this is exactly what we observe. Results from mixed-effects regressions are represented in Figure 2. In both tasks, there was a four-way interaction between age group, condition, trial, and stimulus strength (In the Dots Task: $\beta = 4.32$, SE = 0.70, $p < .0001$, 95% CI = [3.22, 5.42]; In the Ethics Task: $\beta = 1.13$, SE = 0.21, $p < .0001$, 95% CI = [0.72, 1.54]). Furthermore, separate regression analyses for the two age groups revealed that the effect of prevalence-induced concept change was stronger in younger adults ($\beta = 22.54$, SE = 0.42, $p < .0001$, 95% CI = [21.73, 23.36]) than older adults ($\beta = 18.62$, SE = 0.28, $p < .0001$, 95% CI = [18.06, 19.18]). In the Ethics Task the three-way interaction was significant in younger adults ($\beta_{\text{Young Adults}} = 1.34$, SE = 0.17, $p < .0001$, 95% CI = [1.00, 1.69]), but not in older adults ($\beta_{\text{Older Adults}} = 0.13$, SE = 0.12, $p = .2754$, 95% CI = [-0.11, 0.37])). Overall, these results suggest that older adults are less sensitive to prevalence-induced concept change than younger adults.

Response Times

Response time data across age groups are presented in Figure 3. Two 2x2 ANOVA (age group × condition) were conducted on each subjects' mean response time data. These analyses revealed a significant main effect of age group on response time in both tasks (Dots Task: $F(1, 116) = 87.33$, $p < .0001$, 95% CI = [0.23, 0.40], $\text{difference}_{\text{Older - Young}} = 0.28$ seconds; Ethics Task: $F(1, 116) = 28.19$, $p < .0001$, 95% CI = [0.86, 3.30], $\text{difference}_{\text{Older - Young}} = 2.32$ seconds), but no statistically significant main effect of condition (Dots Task: $F(1, 116) = 0.12$, $p = .7529$, 95% CI = [-0.06, 0.11]; Ethics Task: $F(1, 116) = 0.37$, $p = .5432$, 95% CI = [-1.72, 0.71]) or interaction between age group or condition (Dots Task: $F(1, 116) = 1.20$, $p = ..2768$, 95% CI = [-0.19, 0.05]; Ethics Task: $F(1, 116) = 0.30$, $p = .5837$, 95% CI = [-1.24, 2.21]). These findings suggest that older adults made slower responses in both tasks, but that neither group differed with regards to response speed between conditions.

Computational Modeling Results

We conducted a 2x2 ANOVA (age group × condition) on each free parameter (β_0, β_f, β_F, β_C, λ_F and λ_C). We found no statistically significant effect of age group on any of the estimated parameters. However, both decay parameters, λ_F and λ_C, in the Dots Task showed a significant interaction between age group and condition (λ_F: $F(1, 116) = 9.94$, $p = .0019$, $\hat{\eta}^2_G = 0.06$, 95% CI = [-0.42, -0.10]; λ_c: $F(1, 116) = 11.92$, $p = .0007$, $\hat{\eta}^2_G = 0.07$, 95% CI = [-0.55, -0.15]). In both cases, this interaction suggests that older adults have a slower decay parameter in the stable condition, but quicker decay in the decreasing condition. No such interactions were found in the Ethics Task.

Overall, these results did not substantiate our original prediction that older adults' decreased sensitivity to prevalence-induced concept would be reflected by an increased influence of the β_C parameter (greater influence of past choice; increased perseveration). We provide an alternative explanation for the behavioural results we observe below.

Discussion

In this experiment, we demonstrate that older adults are less sensitive to prevalence-induced concept change than young adults. The observed age differences are independent of the complexity of the judgements (colour of dots versus ethical of research proposals) and may indicate that concept spaces in older adults are less susceptible to environmental changes.

But, where do these age differences come from? We hypothesized that if age differences existed, they could be explained by differences between young and older adults in estimated parameter values using a computational model (Figure 1; Wilson, 2018). This hypothesis was not substantiated, however. We did not find meaningful age differences in any of the estimated parameter weights. While the model replicated Wilson's (2018) overall pattern of results, neither parameter accounted for age differences in concept change.

Apart from the reduction of prevalence-induced concept change in older adults, one major difference in the behavior of younger and older adults is in their response times. As shown in Figure 3, older adults responded much slower than younger adults. This age-related slowing is an expected and well-documented feature of healthy cognitive ageing (Salthouse, 1992). However, the consequence of the slower response times is that the interval between stimuli increases, which may lead to a reduced impact of the previous stimulus on the current judgement. As noted by Wilson (2018), increases in this ITI may lead to a reduction in prevalence-induced concept change, given that the weakest effects observed in Levari's and colleagues' (2018) results were in the task with the longest space between stimuli (the Ethics Task). To account for the potential effects of prolonged responses on prevalence-induced concept change, we modified the original computational model such that it accounts for response time differences (see Supplement). Simulated data from this model (Figure 4) qualitatively resemble the empirical findings in Figure 2 and suggest a reduction of prevalence-induced concept change in older adults.

However, there is also an alternative explanation for why longer response times in older adults may have led to reduced prevalence-induced concept change. Several findings suggest that older adults engage in speed-accuracy trade-offs when performing cognitive tasks (Starns & Ratcliffe, 2010; 2012; Forstmann et al., 2011). That is, they slow down in order to respond more accurately. From this perspective, it wouldn't be the case that older adults are necessarily limited in their ability to respond quickly *per se*, but rather prioritise accuracy over response speed. Thus, in the context of our tasks, older adults might spend more time judging each exemplar in a bid to maximise "accuracy" (or internal consistency in the case of the Ethics Task). Were this the case, it would suggest that the effects of prevalence-induced concept change might be constrained by a more deliberate decision-making strategy rather than a processing limitation in older adults.

To tease these hypotheses apart, we conducted two follow-up experiments in younger adults. The aim of these two experiments was to shift the behaviour of younger adults to resemble that of older adults. In Experiment 2, we manipulated the ITI from 500 ms (in Experiment 1) to 2000 ms to determine the effect of spacing the stimuli out more without changing participants' decision-making strategies. In Experiment 3, we did the opposite: we maintained the same ITI of 500 ms, but offered younger participants reward for correct responses to shift decision-making strategies (i.e., speed-accuracy trade-offs).

Experiment 2

To explore whether a longer spacing between stimuli might incur reductions in prevalence-induced concept change, we asked younger participants to complete a modified version of the Dots Task where the time between trials was greater than in Experiment 1. This version of the task was used to simulate the experience of slower responses, without modifying the strategy participants used to complete the task. We hypothesized that younger adults in this sample would experience reduced prevalence-induced concept change compared to those in Experiment 1.

Method

Participants

We recruited 36 young adults (M_{age} = 20.89; s_{age} = 2.09) from the community and the university participation pool. All participants met the inclusion criteria mentioned in the Method section for Experiment 1. The participants were randomly assigned to either the decreasing prevalence condition or the stable prevalence condition, in a counterbalanced order. All participants were compensated $20 CAN or 2 participation pool credits for participating in the study.

Materials and Procedure

ITI-Modified Dots Task. This version of the Dots Task is the same as the original version used in Experiment 1 in all respects except for one: the ITI between stimuli was increased from 500ms to 2000ms. This time of 2000 ms was chosen as a way to test whether simple timing differences between stimuli affected sensitivity to prevalence-induced concept change (a) between the age groups (i.e., to make the case that longer responses on the part of older adults may affect sensitivity to the phenomenon) and (b) between the tasks (i.e., if reduced prevalence-induced concept change in the Ethics Task might be driven by longer response times across both age groups). The procedure was the same as in Experiment 1.

Analysis

As before, the main analysis consisted of two general binomial mixed-effects models. The first model compared young adults in both experiments, predicting response using study (Experiment 1 or Experiment 2), condition (stable prevalence or decreasing prevalence), trial number, and colour strength as fixed effects. All main effects and interactions were explored. The follow-up model used the same predictors, but in Experiment 2 only. In all models, trial and stimulus strength were converted to scores between zero and one.

Results

Choice Data

Choice data for Experiment 2 are presented in Figure 5A. We found a significant four-way interaction between study, condition, trial, and colour strength ($\beta = -27.00$, SE = 1.13, $p < .0001$, 95% CI = [-29.24, -24.76]), such that the effect of prevalence-induced concept change was significantly smaller for young adults in Experiment 2 compared to those in Experiment 1. A follow-up analysis revealed a still significant, but dramatically smaller, interaction between condition, trial, and colour on response for participants in Experiment 2 ($\beta = -4.41$, SE = 1.05, $p < .0001$, 95% CI = [-6.93, -2.39]).

Discussion

In this experiment, we explored whether the degree of prevalence-induced concept change depends on the spacing between stimuli. Our results support this hypothesis and demonstrate a significant reduction in prevalence-induced concept in younger adults from Experiment 1 to Experiment 2. It is worth noting though that, even in the long-ITI condition in Experiment 2, the prevalence-induced concept change effect was still present and statistically significant. These results support the possibility that a reduced sensitivity to prevalence-induced concept change in older adults could be due to general slowing and the consequential increased spacing between stimuli. It is important to keep in mind, however, that even in the task with the longest ITI (the ethics task in experiment 1) the effect of prevalence-induced concept change remained present and statistically significant. Therefore, our evidence suggests that while prevalence-induced concept change is sensitive to the spacing between stimuli, it is not dependent on it as a cognitive phenomenon.

Experiment 3

In Experiment 3, we tested whether reduced concept change in older adults in Experiment 1 was due to a speed-accuracy trade-off, where they sacrificed faster responses for more accurate ones. If this was the case, the difference in response times we observed between young adults and older adults might not be due to a processing limitation (i.e., general slowing), but rather to a more deliberate and cautious approach to the task that prioritized accuracy (or perhaps consistency) over response speed.

To test whether speed-accuracy trade-offs of this kind could affect sensitivity to prevalence-induced concept change, we tested young adults on a modified version of the Dots Task, where we provided additional credit for correct responding and block-to-block performance feedback. We hypothesized that participants would slow down their responses to prioritize accuracy in order to avoid losing credit and that this strategy would result in reduced prevalence-induced concept change.

Method

Participants

We recruited 50 young adults (M_{age} = 21.27; s_{age} = 1.87) from the university participation pool. All participants met the inclusion criteria mentioned in the Method section for Experiment 1 and Experiment 2 and half of the participants were randomly assigned to either the decreasing prevalence condition or the stable prevalence condition, in a counterbalanced order.

Materials and Procedure

Rewarded Dots Task. This version of the Dots Task offered participants rewards in the form of class credits for correctly identifying dots as blue or purple. The task was the same as Experiment 1, with an ITI of 500ms. Because presenting performance feedback on each trial would increase the space between stimuli, effectively increasing the ITI, feedback was provided at the end of each block.

Participants began each session with 4 credits. With each incorrect response, participants lost 0.0025 credits (conveyed to them by saying 20 mistakes would cost them 0.05 credits). The minimum number of credits they could receive was 2 credits, to ensure that all participants were fairly compensated for their time. Between each block, text appeared that indicated that the block was finished, which block the participant was now at, the number of errors they made, and the amount of credits they had remaining. The procedure for this Experiment was the same as in Experiment 1 and 2, except that all participants were compensated with the participation pool credits they earned during the Rewarded Dots Task for participating in the study.

Analysis

The main analysis consisted of two general binomial mixed-effects models. The first model compared young adults in Experiment 1 and Experiment 3, predicting response using study (Experiment 1 or Experiment 3), condition (stable prevalence or decreasing prevalence), trial number, and colour strength as fixed effects, a random slope of trial, and a random intercept for each participant. All main effects and interactions were explored. The follow-up model used the same predictors, but in Experiment 3 only. In all models, trial and stimulus strength were put on a scale between one and zero.

Two one-way ANOVA were also conducted comparing response times between young and older adults in the Dots Task in Experiment 1 vs. the Rewarded Dots task in Experiment 3, as a check to demonstrate that our experimental manipulation affected response strategies, using response time as a proxy.

Results

Choice Data

Choice data for Experiment 3 are presented in Figure 5B. We found a significant four-way interaction between study, condition, trial, and colour strength (β = -7.24, SE = 1.98, p = .0002, 95% CI = [-11.10, -3.37]), such that the effect of prevalence-induced concept change was significantly smaller for young adults in Experiment 3 compared to those in Experiment 1. A follow-up analysis revealed a significant interaction between condition, trial, and colour on response for participants in Experiment 3 (β = 16.41, SE = 1.16, p < .0001, 95% CI = [14.10, 18.70]). However, the size of this effect was smaller than the effect size observed in young adults during Experiment 1 (β = 22.54 in Experiment 1 vs. β = 16.41 in Experiment 3).

Response Times

We found a significant main effect of experiment on response times for both sets of young adults (F(1, 128) = 32.87, p < .0001, 95% CI = [0.08, 0.16], Mean $\text{difference}_{\text{YoungExp3- YoungExp1}}$ = 0.12 seconds), such that younger adults in Experiment 3 took more time responding than younger adults in Experiment 1. Furthermore, we found a significant main effect of age group between young adults in Experiment 3 and older adults in Experiment 1 (F(1, 128) = 17.21, p < .0001, 95% CI = [0.08, 0.23], Mean $\text{difference}_{\text{OldExp1- YoungExp3}}$ = 0.16 seconds), such that older adults still responded slowest overall across experiments.

Discussion

In this experiment, we aimed to demonstrate that participants who engaged in a speed-accuracy trade-off would demonstrate reduced prevalence-induced concept change (as observed in older adults in Experiment 1). Our results support this view. Young adults in Experiment 3 responded more slowly than young adults in Experiment 1 and, in turn, were less sensitive to prevalence-induced concept change (i.e., they made more accurate judgements). Importantly, these differences were not confounded by differences in the spacing of stimuli.

These findings suggest that when participants have a vested interest in maximizing accurate responding, they can reduce—though not eliminate—prevalence-induced concept change. The results line up with the older adults' finding in Experiment 1 and indicate that reduced sensitivity to prevalence-induced concept change in older adults may not only result from general slowing, but may be due to a difference in how older adults approached the task; namely, by trading speed for accuracy in their responses.

General Discussion

The purpose of this study was to investigate how prevalence-induced concept change affects judgements of older adults across two conceptual domains: perception and ethics. Based on previous findings (e.g., Nassar et al., 2016; Lindenberger & Mayr, 2015), we hypothesized that older adults would either be less sensitive (H1) or more sensitive (H2) to prevalence-induced concept change than younger adults. Our results support H1, demonstrating that older adults were less sensitive to prevalence-induced concept change in their judgements about the colours of dots and not significantly affected by the phenomenon in their ethical judgements about fictitious research proposals. We furthermore predicted that this reduction would be explained by increased perseveration on the part of older adults, reflected in estimated weights in a computational model (Wilson, 2018). However, we failed to find meaningful age differences in model parameters. Thus, we conducted exploratory analyses to test the possibility that the observed behavioural differences in sensitivity to prevalence-induced concept change might be due to differences in the spacing between stimuli between young and older adults. If this were the case, we hypothesized that these differences would be due to general slowing (Verhaeghen & Cerella, 2002) and/or a speed-accuracy trade-off (Starns & Ratcliffe, 2010; 2012; Salthouse, 1979) on the part of older adults. To test these hypotheses we conducted two follow-up experiments: In Experiment 2, we changed task and increased the time between stimuli without changing participants motivation; In

Experiment 3, we kept the spacing between stimuli constant but manipulated the decision process by incentivizing accurate responses. In both cases, we found notable reductions in prevalence-induced concept change in younger adults.

There are three important messages that can be taken from these experiments. The results of the first experiment suggest that older adults' concepts seem to be more stable than those of younger adults when faced with a changing task environment. This finding dovetails nicely with a body of research demonstrating that older adults have greater difficulty than younger adults updating behaviour despite changes in the environment (Eppinger, Hämmerer, & Li, 2011; Nassar et al., 2016; Hämmerer et al., 2019). As Wilson (2018) has suggested, the types of serial judgements where prevalence-induced concept change may affect judgments can be thought of as a form of implicit learning, where the underlying state of a stimuli (e.g., the average blueness or ethicality) is implicitly estimated based on recently seen instances of the concept (cf. Cleeremans, Destrebecqz, & Maud Boyer, 1998; Nassar et al., 2012; McGuire, Nassar, Gold, & Kable, 2014; Wilson & Niv, 2012). From this perspective, older adults may have more difficulty learning these latent states of stimuli and default to their original responses (Nassar et al., 2016; Bruckner et al., 2020). Notably, in most task environments these impairments in inferences about latent states are associated with performance deficits. Even outside the lab, it is not difficulty to see how a difficulty in learning from uncertain environment may pose significant problems in day-to-day life. In contrast, in the current task, older adults' reduced sensitivity to environmental statistics was *protective* against some of the negative consequences that could be associated with prevalence-induced concept change (e.g., claiming something is ethical, when you previously said it was not). Thus, while cognitive ageing has negative consequences in many contexts (learning, memory, etc.), it may have unexpected benefits for judgement and decision-making, such as in the case of prevalence-induced concept change.

Second, prevalence-induced concept change is sensitive to the spacing between stimuli. As demonstrated in Experiment 2, prevalence-induced concept change can be meaningfully reduced by increasing the ITI from 500 ms (in Experiment 1) to 2000 ms. As such, it is possible that the reduction in prevalence-induced concept change observed in older adults might be a byproduct of general slowing: As people age, they respond more slowly and, in tasks where the spacing of stimuli affects the prevalence-induced effect, they experience less prevalence-induced concept change. On the one hand, this makes intuitive sense when we consider that in real life instances of categories often change prevalence at a rate of days or months and not seconds. As such, we would expect that the magnitude of the prevalence-induced effect would decrease in accordance with the time it has to take effect (i.e., a smaller effect that carries out over days nevertheless affects our judgements). In the same sense then, this finding also highlights the automatic nature of concept change as a process whose effect is greatest when information is processed rapidly. On the other, more work is needed to determine the degree to which the spacing of stimuli affects prevalence-induced concept change in the lab and in the real world. In the current study, despite reductions in prevalence-induced concept change, changes in ITI did not affect the statistical significance of the effect. As such, it may prove beneficial to design an experimental paradigm that is less sensitive to subtle differences in stimuli spacing and that more readily mimics real-world instances of prevalence-induced concept change (cf. Yarkoni, 2019).

Third, prevalence-induced concept change can be reduced if participants are motivated to respond accurately. This point fits within Kahneman's (2003; cf. 2011) dual-system framework, where inconsistencies in our judgments can be tapered if we engage in what he calls "System 2 Thinking" (slower, more effortful, thinking). That is, when young adults in Experiment 3—a population that is not affected by general slowing—responded more slowly relative to young adults in the Dots Task in Experiment 1, they also experienced less prevalence-induced concept change and were more consistent and accurate in their judgements overall. This finding provides support for the view that older adults engaged in a speed-accuracy trade-off in Experiment 1 that led to a reduction in their sensitivity to prevalence-induced concept change. If this is the case, it is particularly interesting that older adults engaged in this trade-off without external incentive, while young adults required substantial motivation to do so (i.e., to be offered double course-credits for their participation).

Future Directions

The results of the current study answer our initial research question and go beyond it to describe two general factors that may influence one's sensitivity to prevalence-induced concept change. However, this work is but a starting point in describing the intricacies of how and why concept change occurs.

Future work should focus on better understanding the relationship between deliberation and prevalence-induced concept change. In this paper, we have referred to a certain style of decision-making as a speed-accuracy trade-off: one where participants traded fast responses for more accurate ones. However, the history of this term within the ageing literature refers more narrowly to the observation that older adults show disproportionately longer response times and favour accuracy when presented with feedback (see Salthouse, 1979). This process is likely different than the one used by older adults in our sample, who responded more slowly and more accurately, without feedback. The same can be said of young adults in Experiment 3, who received unspecific feedback after long delayed fashion (at the end of each block). As such, it remains unclear *how* participants use these slower response times to adapt behaviour: Do longer response times simply extend the space between stimuli and thus reduce concept change, in line with our results from Experiment 2, or do participants use this extra time to make more confident choices that increase accuracy when motivated to do so?

Taken together, these points emphasize the need for future work that focuses on studying prevalence-induced concept change in more ecologically valid settings, where the temporal spacing of stimuli mimics that which we would expect outside of the lab (e.g., hours, days, weeks) and people's motivations are taken into account.

Conclusion

The current study shows that as we age, our judgements about concepts become more rigid as we face a changing world. While older adults are still generally susceptible to prevalence-induced concept change in basic perceptual tasks, they seem to resist it entirely in higher-order concept judgements about ethics. To shed light more on the cognitive and motivational mechanisms underlying these age-related changes we performed two follow-up experiments in younger adults. The results demonstrate that prevalence-induced concept change is automatic in nature, occurring most prominently when information is processed quickly (Experiment 1). Furthermore, we show that concept change is reduced when subjects are motivated to make slow and deliberate judgements. Taken together, our results suggest that when making judgements older adults engage more in this slow and effortful response mode. Doing so results in a conceptual rigidity that can be beneficial in curtailing biases in judgement that result from changes in the prevalence of events in the environment.

The current findings have real-world relevance when considering the degree to which older adults' use of concepts will come to affect the future direction of our society. As we age, it seems our concepts remain more stable, even if the world around us presents us with continued reason to change them. It is in this sense that the quote at the beginning of this paper earns its relevance: the more things (our age and our environment) change, the more they (our concepts) stay the same.

Significance

The current research is significant for at least three reasons. First, we directly replicate the prevalence-induced concept change effect in young adults across two conceptual levels (Levari et al., 2018). Second, we demonstrate how older adults' concept space is more resilient to concept change that that of younger adults, a fact worth bearing in mind given that older adults' judgements risk to become increasingly important in coming years. Third, we show that prevalence-induced concept change is automatic in nature, occurring most prominently when information is processed quickly, whereas concept change is reduced when people are motivated to make slow and deliberate judgements. In line with this point, we demonstrate that older adults engage in such slow, deliberate, decision-making and, by doing so, demonstrate a conceptual rigidity that can be beneficial in curtailing biases in judgement that result from changes in the prevalence of events in the environment.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bruckner, R., Nassar, M., Li, S.C., & Eppinger, B. (2020). Default beliefs guide learning under uncertainty in children and older adults. https://psyarxiv.com/nh9bq/

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: News from the front. *Trends in Cognitive Sciences*, *2*(10), 406–416. https://doi.org/10.1016/S1364-6613(98)01232-7

Cole, B. L., Lian, K. Y., & Lakkis, C. (2006). The new Richmond HRR pseudoisochromatic test for colour vision is better than the Ishihara test. *Clinical and Experimental Optometry, 89*(2), 73-80. https://doi.org/10.1111/j.1444-0938.2006.00015.x

Daw, N. D. (2011). Trial-by-trial data analysis using computational models. *Decision Making, Affect, and Learning: Attention and Performance XXIII*, 1–26. https://doi.org/10.1093/acprof:oso/9780199600434.003.0001

Eppinger, B., Hämmerer, D., & Li, S. C. (2011). Neuromodulation of reward-based learning and decision making in human aging. *Annals of the New York Academy of Sciences, 1235*, 1. https://doi.org/10.1111/j.1749-6632.2011.06230.x

Eppinger, B., Heekeren, H. R., & Li, S. C. (2015). Age-related prefrontal impairments implicate deficient prediction of future reward in older adults. *Neurobiology of Aging*, *36*(8), 2380–2390. https://doi.org/10.1016/j.neurobiolaging.2015.04.010

Eppinger, B., Nystrom, L. E., & Cohen, J. D. (2012). Reduced sensitivity to immediate reward during decision-making in older than younger adults. *PLoS ONE*, *7*(5), 1–10. https://doi.org/10.1371/journal.pone.0036953

Eppinger, B., Walter, M., Heekeren, H. R., & Li, S. C. (2013). Of goals and habits: Age-related and individual differences in goal-directed decision-making. *Frontiers in Neuroscience*, *7*(7 DEC), 1–14. https://doi.org/10.3389/fnins.2013.00253

Forstmann, B. U., Tittgemeyer, M., Wagenmakers, E. J., Derrfuss, J., Imperati, D., & Brown, S. (2011). The speed-accuracy tradeoff in the elderly brain: a structural model-based approach. *Journal of Neuroscience, 31*(47), 17242-17249. https://doi.org/10.1523/JNEUROSCI.0309-11.2011

Hämmerer, D., Müller, V., & Li, S. C. (2014). Performance monitoring across the lifespan: Still maturing post-conflict regulation in children and declining task-set monitoring in older adults. *Neuroscience and Biobehavioral Reviews*, *46*(P1), 105–123. https://doi.org/10.1016/j.neubiorev.2014.06.008

Hämmerer, D., Schwartenbeck, P., Gallagher, M., FitzGerald, T. H. B., Düzel, E., & Dolan, R. J. (2019). Older adults fail to form stable task representations during model-based reversal inference. *Neurobiology of Aging*, *74*, 90–100. https://doi.org/10.1016/j.neurobiolaging.2018.10.009

Harrison XA, Donaldson L, Correa-Cano ME, Evans J, Fisher DN, Goodwin CED, Robinson BS, Hodgson DJ, Inger R. (2018). A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ 6,* e4794 https://doi.org/10.7717/peerj.4794

Hollenbeck, J. R., & Wright, P. M. (2017). Harking, Sharking, and Tharking: Making the Case for Post Hoc Analysis of Scientific Data. *Journal of Management*, *43*(1), 5–18. https://doi.org/10.1177/0149206316679487

Kahneman, D. (2011). *Thinking Fast and Slow*. Farrar, Straus and Giroux.

Kahneman, D. (2003). Maps of Bounded Rationality: Psychology for Behavioral Economics. *American Economic Review*, *93*(5): 1449-1475. https://doi.org/10.1257/000282803322655392

Levari, D. E., Gilbert, D. T., Wilson, T. D., Sievers, B., Amodio, D. M., & Wheatley, T. (2018). Prevalence-induced concept change in human judgment. *Science*, *360*(6396), 1465–1467. https://doi.org/10.1126/science.aap8731

Li, S. C., Lindenberger, U., & Sikström, S. (2001). Aging cognition: from neuromodulation to representation. *Trends in cognitive sciences*, *5*(11), 479-486. https://doi.org/10.1016/S1364-6613(00)01769-1

Lindenberger, U., & Mayr, U. (2015). Cognitive Aging : Is There a Dark Side to Environmental, *18*(1), 7–15. https://doi.org/10.1016/j.tics.2013.10.006.Cognitive

Mayr, U., Spieler, D. H., & Hutcheon, T. G. (2015). When and why do old adults outsource control to the environment? *Psychology and Aging*, *30*(3), 624–633. https://doi.org/10.1037/a0039466

Mayr, U., Spieler, D. H., & Kliegl, R. (2001). *Ageing and Executive Control*. Psychology Press.

McGuire, J. T., Nassar, M. R., Gold, J. I., & Kable, J. W. (2014). Functionally dissociable influences on learning rate in a dynamic environment. *Neuron*, *84*(4), 870-881. https://doi.org/10.1016/j.neuron.2014.10.013

Nassar, M. R., Bruckner, R., Gold, J. I., Li, S. C., Heekeren, H. R., & Eppinger, B. (2016). Age differences in learning emerge from an insufficient representation of uncertainty in older adults. *Nature Communications*, *7*(May 2015). https://doi.org/10.1038/ncomms11609

Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., & Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature neuroscience*, *15*(7), 1040. https://doi.org/10.1038/nn.3130

Nyberg, L., Lövdén, M., Riklund, K., Lindenberger, U., & Bäckman, L. (2012). Memory aging and brain maintenance. *Trends in cognitive sciences*, *16*(5), 292-305 .https://doi.org/10.1016/j.tics.2012.04.005

Salthouse, T. A. (1979). Adult age and the speed-accuracy trade-off. *Ergonomics*, *22*(7), 811–821. https://doi.org/10.1080/00140137908924659

Salthouse, T. A. (1992). Influence of processing speed on adult age differences in working memory. *Acta psychologica*, *79*(2), 155-170. https://doi.org/10.1016/0001-6918(92)90030-H

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological review, 103*(3), 403. https://doi.org/10.1037/0033-295X.103.3.403

Samanez-Larkin, G. R., & Knutson, B. (2015). Decision making in the ageing brain: changes in affective and motivational circuits. *Nature Reviews Neuroscience, 16*(5), 278-289. https://doi.org/10.1038/nrn3917

Mather, M., & Harley, C. W. (2016). The locus coeruleus: Essential for maintaining cognitive function and the aging brain. *Trends in cognitive sciences*, *20*(3), 214-226. https://doi.org/10.1016/j.tics.2016.01.001

Spieler, D. H., Mayr, U., & LaGrone, S. (2006). Outsourcing cognitive control to the environment: Adult age differences in the use of task cues. *Psychonomic Bulletin and Review*, *13*(5), 787–793. https://doi.org/10.3758/BF03193998

Starns, J. J., & Ratcliff, R. (2010). The effects of aging on the speed-accuracy compromise: Boundary optimality in the diffusion model. *Psychology and Aging*, *25*(2), 377–390. https://doi.org/10.1037/a0018022

Starns, J. J., & Ratcliff, R. (2012). Age-related differences in diffusion model boundary optimality with both trial-limited and time-limited tasks. *Psychonomic bulletin & review*, *19*(1), 139-145. https://doi.org/10.3758/s13423-011-0189-3

Statistics Canada. (2019). *What will the population of Canada look like in 2068?* Retrieved from: https://www150.statcan.gc.ca/n1/pub/11-627-m/11-627-m2019050-eng.htm.

U.S. Census Bureau (2018). The 2017 National Population Projections. Retrieved from https://www.census.gov/programs-surveys/popproj.html

Verhaeghen, P., & Cerella, J. (2002). Aging, executive control, and attention: a review of meta-analyses. *Neuroscience and Biobehavioral Reviews*, *26*, 1–9.

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & cognition*, *32*(7), 1206-1220. https://doi.org/10.3758/BF03196893

Wilson, R. C., & Niv, Y. (2012). Inferring relevance in a changing world. *Frontiers in human neuroscience*, *5,* 189. https://doi.org/10.3389/fnhum.2011.00189

Acknowledgments

Funding: This work was supported by a grant from the Social Sciences and Humanities Research Council of Canada (SSHRC) [CGS-M 2019] awarded to S.D. B.E. received support from the German Research Foundation (DFG; SFB 940/2 B7), as well as the Natural Sciences and Engineering Research Council of Canada (NSERC; N01882). Furthermore, this research was undertaken, in part, thanks to funding from the Canada Research Chairs program.

Author contributions: S.D. conceived of the three experiments, oversaw testing, analysed the data, and composed the initial version of the manuscript. C.N. tested participants and provided critical comments on the manuscript in its early stages. D.L. provided guidance on the project and commented on the manuscript. R.W. helped implement the computational models and provided comments on the manuscript. B.E. supervised the project and offered crucial revisions.

Competing interests: The authors declare no conflict of interest.

Data and materials availability: All task code, raw data, and analysis scripts are available at https://osf.io/pbzv3/.

Figures and Tables

Figure 1 Simulations

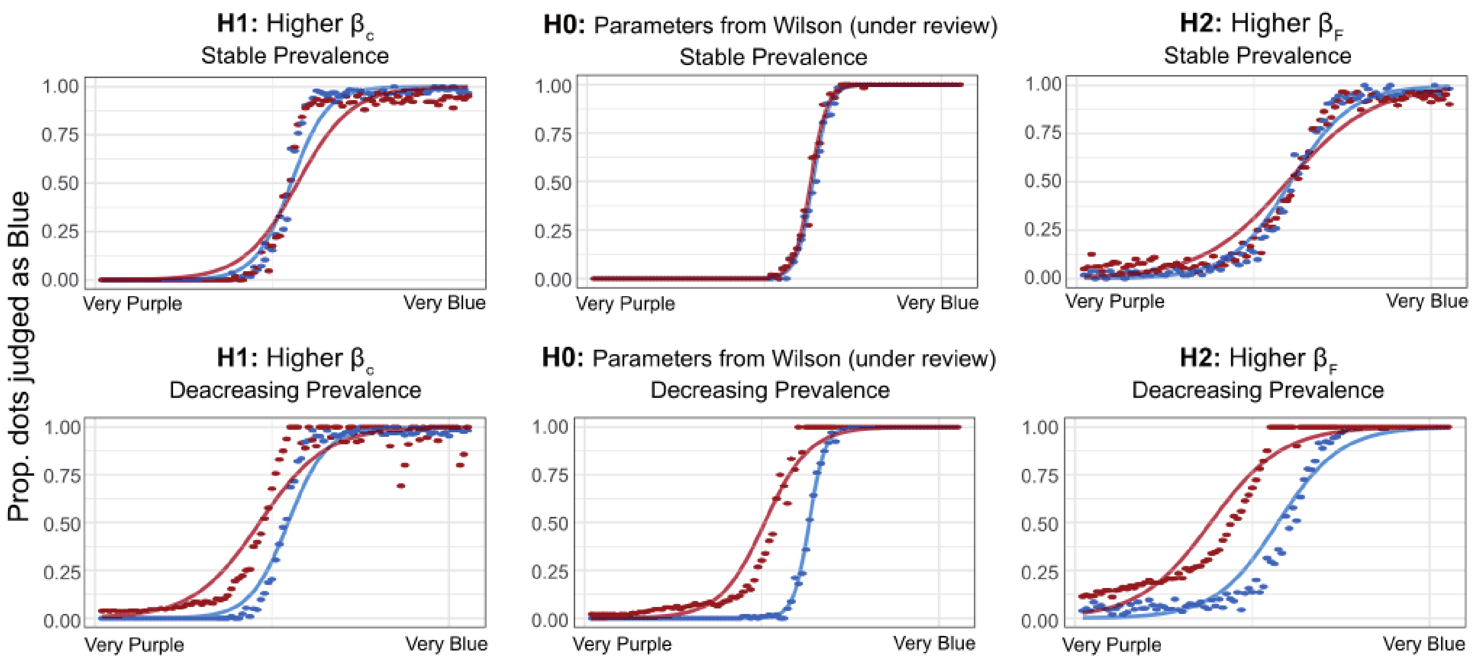

Fig. 1. **Simulated data representing three hypotheses**. H1 (far left) shows decreased prevalence-induced concept change if older adults have a higher average β_C, representing a greater effect of past response on current response (i.e., more perseverance). H2 (far right) shows increased prevalence-induced concept change if older adults have higher absolute β_F values (negative in reality), representing a greater effect of past stimuli on current response (i.e., more outsourcing). Finally, H0 (center) represents a scenario where older and younger adults do not differ in their sensitivity to prevalence-induced concept change.

Figure 2 *Experiment 1*

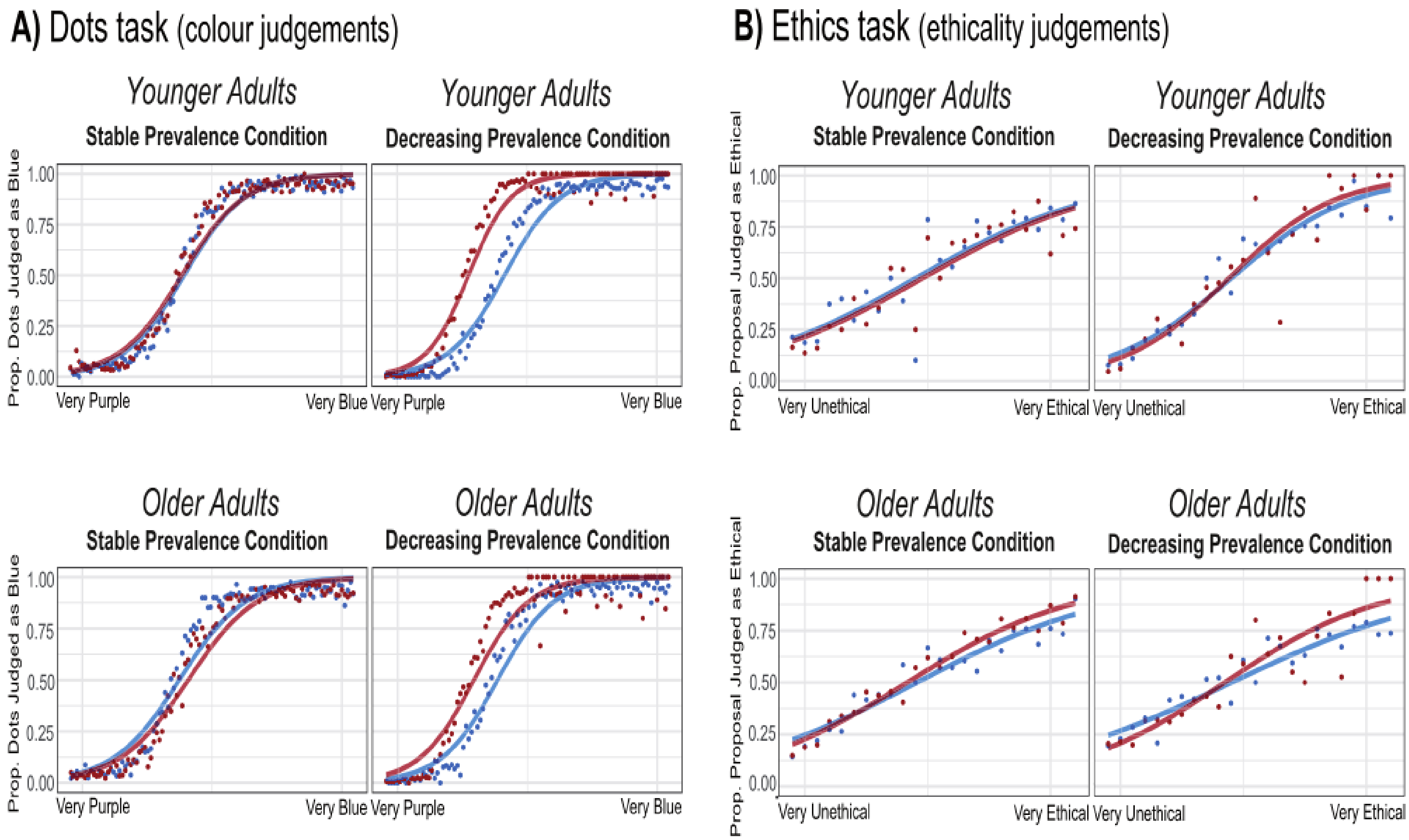

Fig 2. Concept judgements in (A) the Dots Task and (B) the Ethics Task. The x-axis represents stimulus strength: blueness in the Dots Task and ethicality in the Ethics task. The y-axis represents the percentage of dots/proposals judged as blue/ethical. Points represent the percent of choices for the corresponding stimulus strength, averaged across subjects within that cell. Curves represent fitted binomial regression curves.

Figure 3

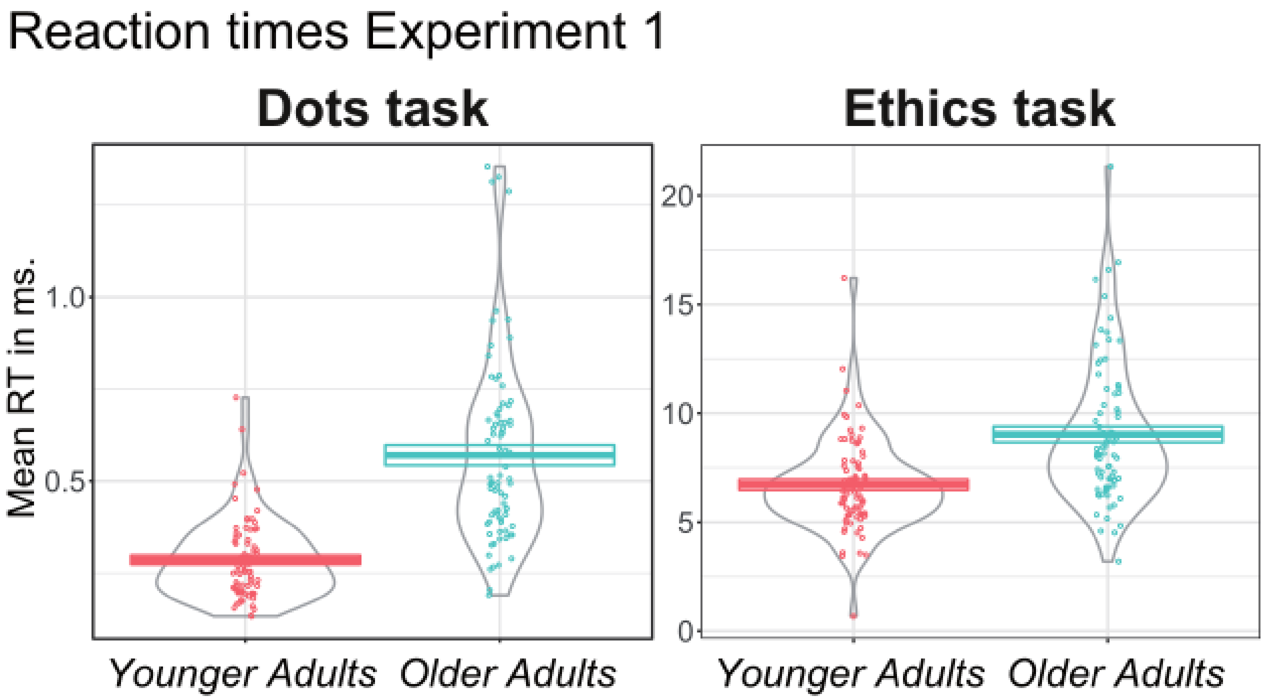

Fig. 3. Pirate plots of mean response times in both age groups across both tasks. X-axis is the age group. Y-axis is the mean response time in seconds per participant (note the difference in scale across tasks). Each point represents an individual participant's mean response time. Boxes represent standard error and horizontal lines represent group means.

Figure 4

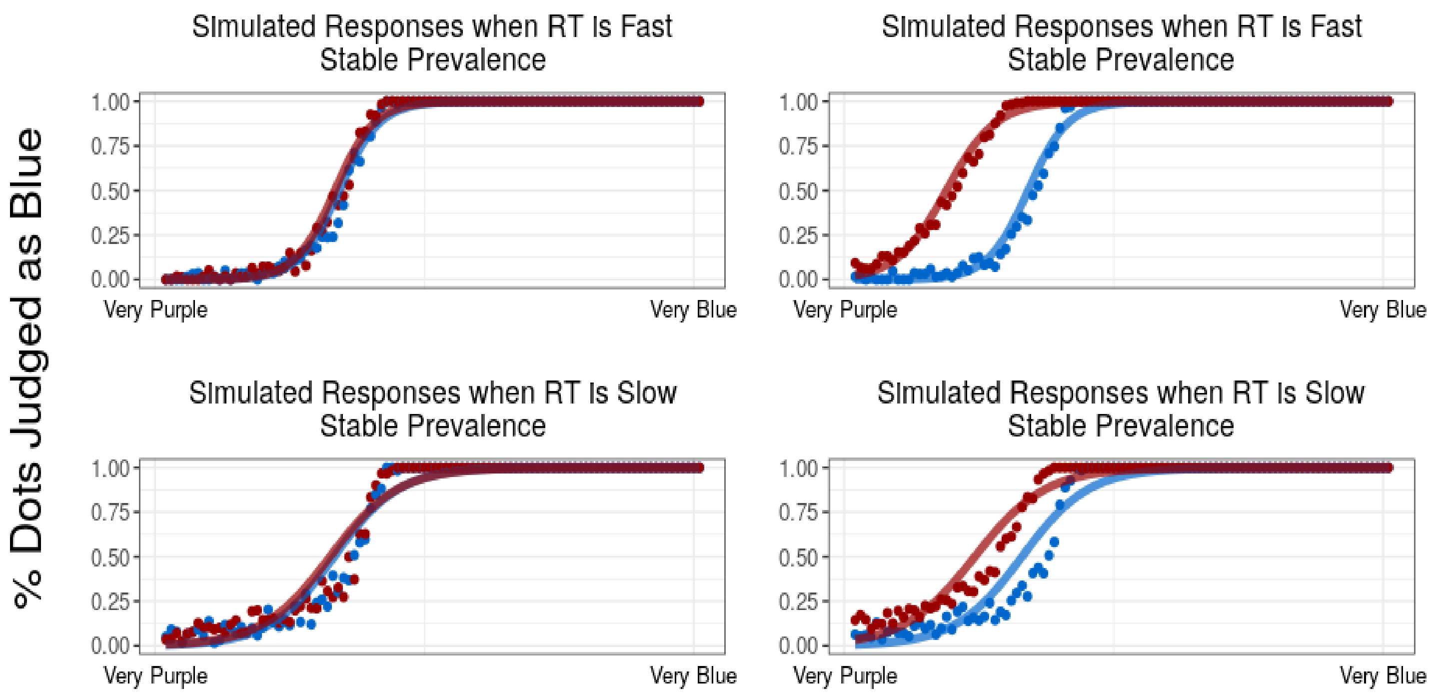

Fig. 4. Exploratory data simulation using a modified model that incorporates response times. In both fast and slow response time groups, the same free parameters are used; only response times vary. Note that distributions are shifted to the left because parameters are not optimised for this modified model.

Figure 5

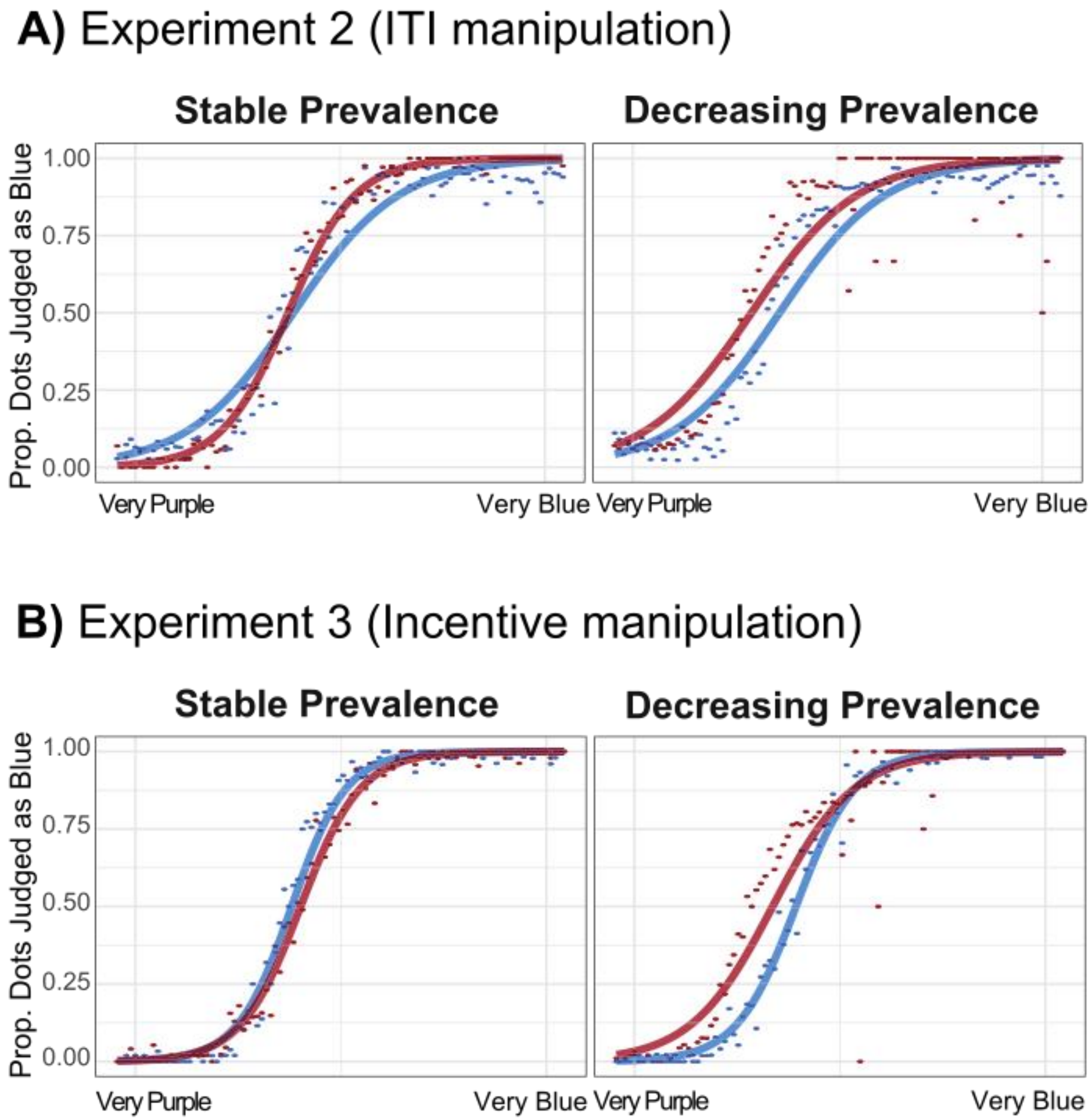

Fig. 5. Concept judgements in (A) ITI-modified Dots Task and (B) the Rewarded Dots Task. The x-axis represents stimulus strength (i.e., blueness). The y-axis represents the percentage of dots judged as blue. Points represent the percent of choices for the corresponding stimulus strength, averaged across subjects within that cell. Curves represent fitted binomial regression curves. Blue points and lines represent the first 200 trials in the Dots Task and red ones represent the final 200 trials in the Dots Task.

Supplementary Materials

HRR Colour Vision Test

The HRR colour vision test is a short screening test to ensure that participants' colour vision is adequate for the Dots Task. Specifically, this test was included to ensure that participants did not differ in how they experienced the stimuli in the Dots Task. Furthermore, given that many older adults experience sensory deficits that compromise their ability to discriminate colours, we also wanted to control for age-related impairments in colour discrimination.

The test contains 24 plates (pages), each displaying either one or two symbols, which can be a circle, a cross, or a triangle, four of which are demonstration plates to explain the task and six of which are screening plates used to classify participants based on their colour vision. The remaining plates are used to grade the severity of certain deficiencies. Only the first 10 plates were used in this study, as is standard in assessing basic colour discrimination (Cole, Lian, & Lakkis, 2006). The symbols on each plate are constructed of coloured dots that would be difficult or impossible to discern if someone were colourblind (Cole, Lian, & Lakkis, 2006).

An experimenter presented the plates to participants one at a time and asked them to identify how many symbols they saw, what the symbols were, and to outline those symbols with a brush. Participants in this study were graded as pass/fail, receiving a failing grade as soon as they either failed to identify one of the symbols or misidentified a symbol. A passing grade was only given if all plates were correctly identified.

Unsurprisingly, more older adults failed the colour vision test than younger adults (14 older adults vs. 6 younger adults). To test whether these participants affected the overall pattern of results we observed in Experiment 1, we ran the same binomial mixed-effects models described in the main text with them excluded (see Experiment 1 *Statistical Analysis*). However, excluding these participants alone would result in unbalanced data in terms of age group (74 young adults vs. 64 older adults), which is known to contribute to decreased power and imprecise estimates in mixed-effects models (Harrison et al., 2018). As such, we randomly sampled 64 young adults from the 74 who did not fail the colour vision test and ran the models.

The patterns of results from the overall binomial regression match those presented in the main text (Experiment 1 Results). Namely, we found a significant four-way interaction between age group, condition, trial, and colour strength (β = 3.72, SE = 1.37, p = .0064), such that older adults were less sensitive to prevalence-induced concept change than younger adults (see Supplemental Figure 1). Therefore, to avoid the issue of unbalanced data and to increase power, all participants were included in the regressions presented in the main text.

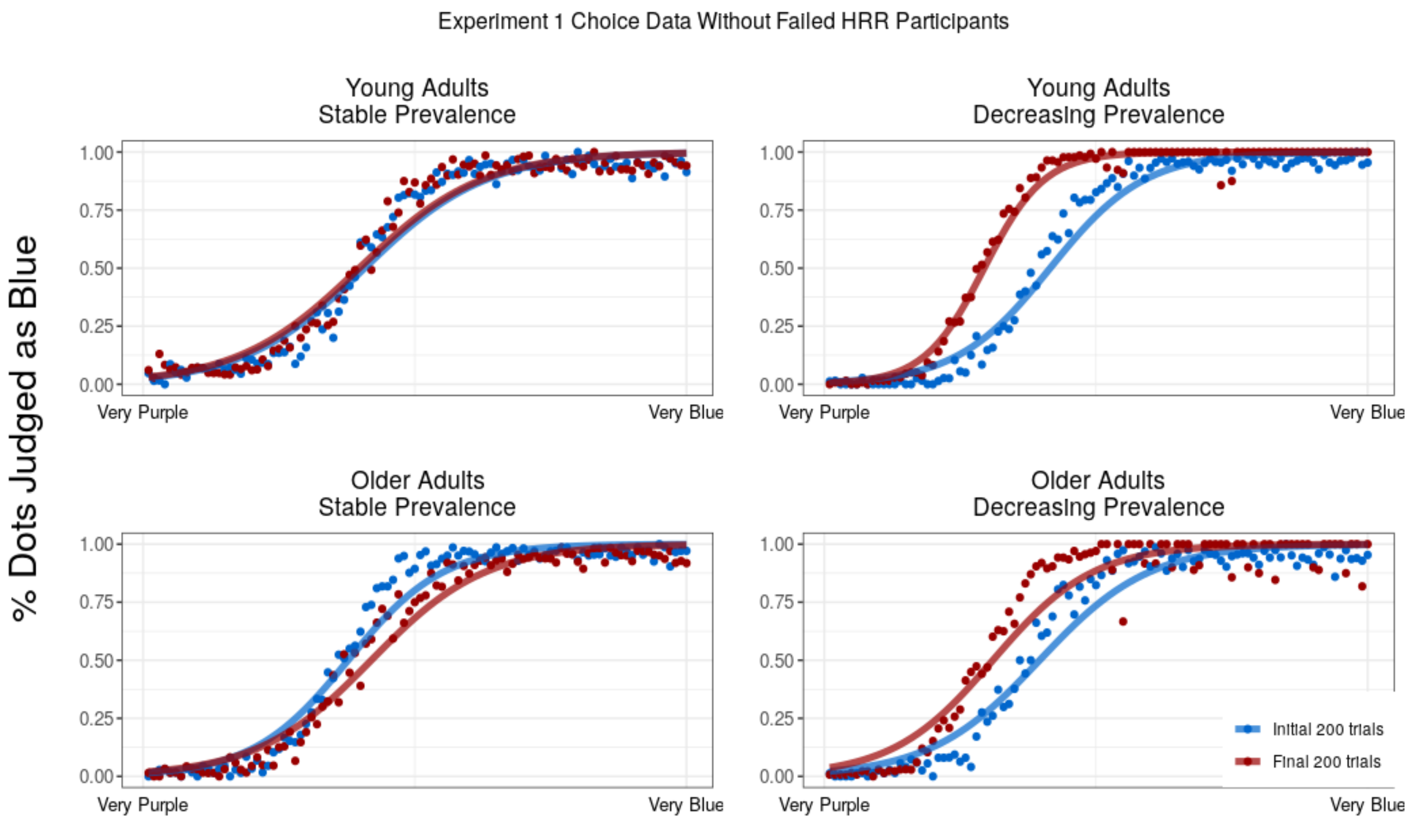

Fig. S1. Concept judgements in the Dots Task after participants who failed the HRR colour vision test are excluded and groups are balanced. The x-axis represents stimulus strength: blueness in the Dots Task and ethicality in the Ethics task. The y-axis represents the percentage of dots/proposals judged as blue/ethical. Points represent the percent of choices for the corresponding stimulus strength, averaged across subjects within that cell. Curves represent fitted binomial regression curves.

Response Time Modified Computational Model

As a first exploration into the role of response times in prevalence-induced concept change, we modified the original computational model (Wilson, 2018; see Method section main text). Specifically, we kept the majority of the model the same, but weighted the decay parameters such that longer responses would incur slower decay on a trial-by-trial basis:

$$F_t = \lambda_F^{RT_t} F_t + f_t \quad \text{(Eq. 4)}$$

$$C_t = \lambda_c^{RT_t} C_t + c_t \quad \text{(Eq. 5)}$$

We used the same parameters in all cases, but manipulated the response time distributions. For the "Low RT" group, we produced a random normal distribution using the same mean and standard deviation we observed in the young adults data above. For the "High RT" group we did the same, using the mean and standard deviation from the older adults. Qualitatively, we can see these simulated data appear to resemble the empirical distributions shown in Figure 2. Simulated data from this modified model is presented in Figure 4.

We used this simulated data as preliminary support that response times may be driving the differences in sensitivity to prevalence-induced concept change we observed between older and younger adults and not participants' sequential decision-making patterns (see main text).