PLOS ONE

An extended hierarchical ordered probit model robust to heteroskedastic vignette perceptions with an application to functional limitation assessment

--Manuscript Draft--

Manuscript Number:	PONE-D-19-23566
Article Type:	Research Article
Full Title:	An extended hierarchical ordered probit model robust to heteroskedastic vignette perceptions with an application to functional limitation assessment
Short Title:	An extended hierarchical ordered probit model
Corresponding Author:	Zhiyong Huang Southwestern University of Finance and Economics Chengdu, Sichuan CHINA
Keywords:	Anchoring vignettes; differential item functioning; hierarchical ordered probit model; psychological distance
Abstract:	To improve interpersonal comparability of self-reported measures, anchoring vignettes are increasingly collected in surveys and modeled as the hierarchical ordered probit (HOPIT) model. This paper – based on the idea of psychological distance - relaxes the assumption of vignette equivalence in the HOPIT by allowing for heteroscedasticity in respondents' perceptions of vignettes. Particularly, we assume that respondents who are more similar to a vignette are more familiar with the condition described and therefore are capable of forming a more precise perception of the vignette. We show evidence in favor of this extended HOPIT through Monte Carlo simulations and an application concerning self-reported vision, pain, cognition and mobility from the WHO Study on Global Aging and Adult Health (SAGE).
Order of Authors:	Zhiyong Huang
Additional Information:	
Question	**Response**
Financial Disclosure Enter a financial disclosure statement that describes the sources of funding for the work included in this submission. Review the submission guidelines for detailed requirements. View published research articles from *PLOS ONE* for specific examples. This statement is required for submission and **will appear in the published article** if the submission is accepted. Please make sure it is accurate.	This research was supported in part by National Natural Science Foundation of China 71804151 (Zhiyong Huang) (http://ms.nsfc.gov.cn/index.php?r=search/index&Projects_page=80&Projects_sort=fund). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Unfunded studies Enter: *The author(s) received no specific funding for this work.* **Funded studies** Enter a statement with the following details: • Initials of the authors who received each award • Grant numbers awarded to each author • The full name of each funder • URL of each funder website • Did the sponsors or funders play any role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript? • **NO** - Include this sentence at the end of your statement: *The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.* • **YES** - Specify the role(s) played. * typeset	
Competing Interests Use the instructions below to enter a competing interest statement for this submission. On behalf of all authors, disclose any competing interests that could be perceived to bias this work—acknowledging all financial support and any other relevant financial or non-financial competing interests. This statement **will appear in the published article** if the submission is accepted. Please make sure it is accurate. View published research articles from *PLOS ONE* for specific examples.	The authors have declared that no competing interests exist

NO authors have competing interests Enter: *The authors have declared that no competing interests exist.* **Authors with competing interests** Enter competing interest details beginning with this statement: *I have read the journal's policy and the authors of this manuscript have the following competing interests: [insert competing interests here]* * typeset	
Ethics Statement Enter an ethics statement for this submission. This statement is required if the study involved: • Human participants • Human specimens or tissue • Vertebrate animals or cephalopods • Vertebrate embryos or tissues • Field research Write "N/A" if the submission does not require an ethics statement. General guidance is provided below. Consult the submission guidelines for detailed instructions. **Make sure that all information entered here is included in the Methods section of the manuscript.**	This paper is a secondary analysis using data from the WHO Study on Global AGEing and Adult Health (SAGE) for which the ethical approval was obtained from the World Health Organization's Ethical Review Board

Format for specific study types **Human Subject Research (involving human participants and/or tissue)** • Give the name of the institutional review board or ethics committee that approved the study • Include the approval number and/or a statement indicating approval of this research • Indicate the form of consent obtained (written/oral) or the reason that consent was not obtained (e.g. the data were analyzed anonymously) **Animal Research (involving vertebrate animals, embryos or tissues)** • Provide the name of the Institutional Animal Care and Use Committee (IACUC) or other relevant ethics board that reviewed the study protocol, and indicate whether they approved this research or granted a formal waiver of ethical approval • Include an approval number if one was obtained • If the study involved *non-human primates*, add *additional details* about animal welfare and steps taken to ameliorate suffering • If anesthesia, euthanasia, or any kind of animal sacrifice is part of the study, include briefly which substances and/or methods were applied **Field Research** Include the following details if this study involves the collection of plant, animal, or other materials from a natural setting: • Field permit number • Name of the institution or relevant body that granted permission	
Data Availability Authors are required to make all data underlying the findings described fully available, without restriction, and from the time of publication. PLOS allows rare exceptions to address legal and ethical concerns. See the PLOS Data Policy and FAQ for detailed information.	Yes - all data are fully available without restriction

A Data Availability Statement describing where the data can be found is required at submission. Your answers to this question constitute the Data Availability Statement and **will be published in the article**, if accepted. **Important:** Stating 'data available on request from the author' is not sufficient. If your data are only available upon request, select 'No' for the first question and explain your exceptional situation in the text box. Do the authors confirm that all data underlying the findings described in their manuscript are fully available without restriction?	
Describe where the data may be found in full sentences. If you are copying our sample text, replace any instances of XXX with the appropriate details. • If the data are **held or will be held in a public repository**, include URLs, accession numbers or DOIs. If this information will only be available after acceptance, indicate this by ticking the box below. For example: *All XXX files are available from the XXX database (accession number(s) XXX, XXX.*). • If the data are all contained **within the manuscript and/or Supporting Information files**, enter the following: *All relevant data are within the manuscript and its Supporting Information files.* • If neither of these applies but you are able to provide **details of access elsewhere**, with or without limitations, please do so. For example: *Data cannot be shared publicly because of [XXX]. Data are available from the XXX Institutional Data Access / Ethics Committee (contact via XXX) for researchers who meet the criteria for access to confidential data.* *The data underlying the results presented in the study are available from (include the name of the third party*	All data files are available from the WHO Study on global AGEing and adult health (SAGE).

and contact information or URL). • This text is appropriate if the data are owned by a third party and authors do not have permission to share the data. * typeset	
Additional data availability information:	

An extended hierarchical ordered probit model robust to heteroskedastic vignette perceptions with an application to functional limitation assessment

Zhiyong Huang[1*]

1 Southwestern University of Finance and Economics, Chengdu, China

* zhiyonghuang@swufe.edu.cn

Abstract

To improve interpersonal comparability of self-reported measures, anchoring vignettes are increasingly collected in surveys and modeled as the hierarchical ordered probit (HOPIT) model. This paper – based on the idea of psychological distance - relaxes the assumption of vignette equivalence in the HOPIT by allowing for heteroscedasticity in respondents' perceptions of vignettes. Particularly, we assume that respondents who are more similar to a vignette are more familiar with the condition described and therefore are capable of forming a more precise perception of the vignette. We show evidence in favor of this extended HOPIT through Monte Carlo simulations and an application concerning self-reported vision, pain, cognition and mobility from the WHO Study on Global Aging and Adult Health (SAGE).

Introduction

Many studies from different domains such as population health or political science are commonly using self-assessments as an alternative to objective measures, which might be infeasible or too costly to collect in surveys. Despite the widespread use of self-assessments, there is some concern as to their comparability among individuals with different traits, such as age, gender, socioeconomic status, culture, and nationality [1–3]. While a respondent's underlying objective condition - which is often the variable of interest - potentially depends on these traits, the response scales underlying her self-assessment may depend on the same traits as well. The use of different response scales in self-assessments by different individuals can compromise the interpersonal comparability of such self-assessed measures. The lack of interpersonal comparability of self-assessments in social surveys has been referred to as differential item functioning (DIF) [1, 4], reporting heterogeneity [5, 6], or cut-point shifts [7]. Figure 1 portrays an example of DIF regarding visual acuity, which is commonly measured by self-reports in general surveys. As different reporting scales are adopted, the difference of self-assessments fails to reflect the true difference of vision conditions between the two respondents.

The method of anchoring vignettes proposed by [1] is widely used to adjust for potential heterogeneity in response scales. An anchoring vignette is a brief description of a hypothetical person or situation for a concept relevant to the research question. For instance, one of the anchoring vignettes included in the WHO Study on Global Aging and Adult Health (SAGE) in the domain of visual acuity reads as: "[Eddy] needs a magnifying glass to read small print and look at details on pictures. He also takes a

Fig 1. DIF of the item: self-reported vision. The graph illustrates possible mappings from actual visual acuity to respondents' self-assessments. Despite better visual acuity of the Respondent 2 than the Respondent 1, their self-assessments can be the same when different response scales are used by different respondents. Estimates based on such self-assessments are biased.

while to recognize objects if they are too far from him. Overall in the last 30 days, how much difficulty do you think [Eddy] had in seeing and recognizing a person he knows across the road (from a distance of about 20 meters)?"

Respondents are asked to evaluate anchoring vignettes in addition to their self-assessments. Since the objective situation described in a vignette is the same across respondents, responses to vignette questions can help to reveal response heterogeneity and be used to adjust for individual response scales. Once response heterogeneity has been accounted for, true differences between individuals can be identified from their self-assessments. The general idea can be illustrated in Figure 2, where each respondent rates his/her own condition as in Figure 1, but also the condition of a vignette. Using information revealed by anchoring vignettes, individual responses scales can be aligned. Self-assessments adjusted by this aligned response scale become comparable.

Fig 2. Correction for self-reports using anchoring vignettes

Anchoring vignettes have been included in a number of social surveys including the Health and Retirement Study (HRS), the Survey of Health, Aging and Retirement in Europe (SHARE), the WHO Study on Global Aging and Adult Health (SAGE) and the China Health and Retirement Longitudinal Study (CHARLS) to name but a new. The methodology of anchoring vignettes, has been used for a broad range of interpersonal comparisons with regard to health [3, 8–14], healthcare [15, 16], political efficacy [1, 17, 18], life satisfaction [19–22], job satisfaction [23], working disability [2], social status [24], poverty [22] and quality of life [25].

A commonly used parametric model for incorporating anchoring vignettes is the so-called hierarchical ordered probit (HOPIT) model [1, 10, 21, 26]. The validity of the HOPIT model hinges on two assumptions: vignette equivalence and response consistency. Response consistency assumes that the same reporting scale is used by a respondent when evaluating one's own conditions and anchoring vignettes, while vignette equivalence posits that the perceptions of vignettes are systematically invariant across all respondents.

Vignette equivalence assumes that the perception errors of vignettes are homoskedastic across individuals. This homoskedasticity assumption, however, may be too restrictive. Given several vignettes, a respondent may form a more accurate perception of vignettes which resemble his own condition than those which seem more exotic. Therefore, a vignette can be better perceived by respondents who are similar to the description of the vignette than those who are not familiar with the described condition. For instance, the situation of "Eddy"- the vignette above - might be easily understood and evaluated by an older person with similar vision loss, while a 20-year-old with 20/20 vision may find it much harder to evaluate Eddy's condition. Even though the 20-year-old can assert that Eddy's vision is worse than his own, he might still have some difficulty in assessing the degree of severity of Eddy's visual impairment. As a result, his perception of Eddy's vision limitation may have a relatively large variance, compared to a person in a more similar situation to Eddy. The post-survey interview in a recent validation study of anchoring vignettes by [27], revealed that some young respondents indeed reported difficulties imaging some vignette scenarios related to limitations such as walking difficulties or chronic pain. The potentially large variance due to the perception distance between one's own condition and vignette scenario may as well contribute to commonly observed ties and inconsistencies in assessments of multiple vignettes, as highlighted by [17] on their discussion on tied or inconsistent vignette rankings:

> ". . . [W]e might reasonably expect respondents to be more likely to give some tied or inconsistent answers among vignettes that are far from their own self-assessment even when they correctly rank the vignettes that matter near their own value. For example, if we are measuring height and a respondent knew his or her height to within an inch, he or she still might have difficulty correctly ranking the heights of two trees 200 and 206 feet tall, swaying in the breeze. Yet, the same respondent would presumably have no difficulty understanding that both trees are taller than himself or herself" [17, p. 51].

In this paper, we relax the assumption of vignette equivalence by allowing for heteroskedasticity in vignette perceptions. Particularly, we consider situations in which the information revealed by a vignette depends on the similarity/dissimilarity between a respondent and condition described in that vignette. We assume that the variance of a vignette perception is positively related to the distance between the respondent's own condition and the location of the vignette. Our extension of the HOPIT model has the advantage that vignettes are locally weighted by their distances to the respondent in each respondent-vignette pair. The perception of an anchoring vignette, which is further from the condition of a respondent, is allowed to have a larger variance, accounting for higher propensities of ties and inconsistency observed in that respondent-vignette pair. In addition, our extended HOPIT model nests the standard HOPIT as a special case, and can thus be tested against the standard HOPIT using a likelihood ratio test.

Our idea behind the introduction of heteroskedastic variance is closely related to the measurement of psychological distance, defined as the similarity between one's direct experience of the "here and now" and hypothetical objects which are not yet directly experienced [28]. The construal level theory in social psychology contends that people perceive hypothetical objects in different ways [29]. People comprehend distant objects in a more abstract way while interpreting nearby objects in a more concrete way. In our extended HOPIT, we relate the distance between a respondent and a hypothetical situation - in our case a vignette - to the variance of perception, which quantifies the degree of "concreteness" or "abstractness" of the vignette to the respondent.

To the best of our knowledge, our proposed model is the first model that incorporates heteroskedastic vignette perceptions into the HOPIT. [30] adopted a HOPIT model where different vignettes can have different variances, but the variance of any particular vignette is still assumed to be the same for all respondents. Our extension, however, allows the information content of any particular vignette to be different across respondents. A vignette can be understood quite accurately by some respondents while being vaguely understood by others. Compared with the model taken by [30], our extended HOPIT model admits different variance of vignette perceptions across both respondents and vignettes.

The rest of the paper is organized as follows. Section 2 introduces the extended HOPIT model. Section 3 presents a Monte Carlo study comparing models and an empirical application concerning functional limitation assessment. Section 4 concludes the paper.

Methods

We model self-assessments and anchoring vignettes jointly. A self-assessment question on vision difficulty, for example, may be formulated as the following question from SAGE: "In the last 30 days, how much difficulty did you have in seeing and recognizing an object or a person you know across the road (from a distance of about 20 meters)" , with responses as one of, "none", "mild", "moderate", "severe", or "extreme". We define anchoring vignettes as descriptions of hypothetical persons in the same domain as self-assessment. A corresponding vignette for the above vision self-assessment reads as, "Eddy needs a magnifying glass to read small print and look at details on pictures. He also takes a while to recognize objects if they are too far from him. Overall in the last 30 days, how much difficulty do you think Eddy had in seeing and recognizing a person he knows across the road (from a distance of about 20 meters) ", with responses as one of, "none", "mild", "moderate", "severe", or "extreme".

In this section, we first introduce the model specification of the standard HOPIT and then move to our extended model with heteroskedasticity after a discussion of potential weakness of the standard HOPIT.

The standard HOPIT model

Let i, $i = 1, \ldots, N$ be respondents, k, $k = 1, \ldots, K$ be response categories, and j, $j = 1, \ldots, J$ be vignette questions. We use $Y = \{y_i \mid i = 1, \ldots, N\}$ for the response to the question of self-assessment, $V = \{V_{ij} \mid i = 1, \ldots, N; j = 1, \ldots, J\}$ for responses to the vignette questions, and $X = \{x_i \mid i = 1, \ldots, N\}$ and $Z = \{z_i \mid i = 1, \ldots, N\}$ for regressors determining the true status underlying self-assessments and the cut-off points, respectively. Note that X and Z are not necessarily of the same set, although most empirical studies assume that the two contain the same variables.

We assume that the latent variable of interest takes a linear form

$$y_i^* = \beta' x_i + \sigma_\varepsilon \varepsilon_i \tag{1}$$

where the error term ε_i follows a standard normal distribution , i.e., $\varepsilon_i \sim N(0, 1)$. Further, we impose as identification restrictions $\beta_0 = 0$ and $\sigma_\varepsilon = 1$ without loss of generality as in the standard ordered response models.

The observed categorical responses follows a standard mapping rule given by

$$y_i = k \Longleftrightarrow \tau_i^{k-1} \leq y_i^* < \tau_i^k \tag{2}$$

where τ_i denotes individual-specific cut-off points modeled as

$$\begin{aligned} \tau_i^1 &= \gamma^{1\prime} z_i + \sigma_u u_i \\ \tau_i^k &= \tau_i^{k-1} + \exp\left(\gamma^{k\prime} z_i\right), \, k = 2., \ldots, K-1 \\ \tau_i^0 &= -\infty, \tau_i^K = \infty \end{aligned} \tag{3}$$

The Figure 3 illustrates the model specification of the standard HOPIT model. This model setup is used, for example, by [2, 13, 15, 19, 21].

Fig 3. The standard HOPIT model

Given the above setup, the probability that a respondent i has a response k, conditional on both u_i and ε_i, is

$$\Pr\left(y_i = k \mid x_i, z_i, u_i, \varepsilon_i\right) \quad = \quad 1\left\{\tau_i^{k-1} \leq \beta' x_i + \varepsilon_i < \tau_i^k\right\} \tag{4}$$

where $1\{\}$is the indicator function. Note that in the standard HOPIT model, we can further derive the conditional probability as $\Pr(y_i = k \mid x_i, z_i, u_i) = \Phi\left(\tau_i^k - \beta' x_i\right) - \Phi\left(\tau_i^{k-1} - \beta' x_i\right)$ by integrating ε_i out. In our extended HOPIT model, however, we need the probability to be conditional on ε_i as well, since ε_i is also incorporated in functions of vignette assessments. In the likelihood function, ε_i will be integrated out as an unobserved individual effect.

Anchoring vignettes are used to identify the effect of the regressor on the outcome from reporting styles. The assumptions needed to ensure the validity of anchoring vignettes are vignette equivalence and response consistency. The assumption of response consistency requires that respondents use the same response scales for both their self-assessments and their vignette ratings.

$$V_{ij} = k \Longleftrightarrow \tau_i^{k-1} \le V_{ij}^* < \tau_i^k \tag{5}$$

The assumption of vignette equivalence requires that respondents interpret vignettes in the same way subject to an idiosyncratic error term. Under this assumption, a respondent i's perception of vignette j is given by

$$V_{ij}^* = \theta_j + \sigma_v v_{ij},\, v_{ij} \sim N(0,1) \tag{6}$$

The extended HOPIT model with heteroskedastic vignette perceptions

In the standard HOPIT model, vignette equivalence (Eq. 6) implies that the perception errors of vignettes are homoskedastic across individuals. The homoskedasticity assumption, nonetheless, can be too restrictive. Given several vignettes, a respondent may form a more accurate perception of vignettes which resemble his own condition than those which seem more exotic. Therefore, a vignette can be better perceived by respondents who are similar to the description of the vignette than those who are not familiar with the described condition.

Fig 4. The extended HOPIT with heteroskedastic vignette perceptions

In this paper, we relax the assumption of vignette equivalence by considering a particular source of heteroskedasticity regarding vignette perception. Specially, we assume that the information that a vignette reveals depends on the similarity/dissimilarity between a respondent and the vignette, i.e., we assume that the precision of vignette perception is negatively related to the distance between the respondent's own condition and the vignette,

$$V_{ij}^* = \theta_j + \sigma_v \exp\left(\alpha\left(\beta' x_i + \varepsilon_i - \theta_j\right)^2\right) v_{ij},\, v_{ij} \sim N(0,1) \tag{7}$$

where the parameter α measures the impact of the similarity/dissimilarity between the respondent and the vignette on the level of noise in a specific respondent-vignette-pair. Note that the respondent's own condition consists of both an observed part $\beta' x_i$ and an unobserved part ε_i.

Our specification of heteroskedasticity is based on the theory of psychological distance, which may describe temporal distance between the present and the future, spatial distance between different physical locations, social distance difference between yourself and others, or hypothetical distance between imaging and experienced events [29]. The theory of psychological distance assumes that people think more concretely when faced with an object or an event of a shorter psychological distance while thinking more abstractly on a distant object or event. Whether the hypothetical event transcends into our mindset in a more concrete or a more abstract way would in turn affect our perception precision to the event. For example, [31] have found that a concrete mindset often achieves higher accuracy on the estimate of risk events.

When respondents assess anchoring vignettes which are a set of hypothetical events which are not directly experienced by respondents, their perceptions are apt to be affected by the hypothetical distance. As the respondent shows a higher degree of psychological proximity to the vignette, the perception can become more concrete and thus more precise. In our specification, the degree of concreteness (or abstractness) in vignette perception is modeled through the variance in the vignette perception function.

Under the assumption of heteroskedastic perception, each respondent-vignette-pair is locally weighted by the distance between the respondent's situation and the vignette. The more similar a vignette is to a respondent, the more precise is the respondent's perception to the vignette, and higher the weight the vignette gets in the correction of reporting heterogeneity across respondents. Figure 4 shows one example of heteroskedastic perceptions. A respondent assesses his own situation as well as situations of two vignettes, "Eddy" and "Eric". Eddy's situation is thereby more similar to the respondent's own condition than the situation of Eric. Following our idea of heteroskedastic vignette perceptions, the respondent would have a more accurate perception of Eddy's situation than the situation of Eric, as the former is more similar to his own condition.

In addition, when $\alpha = 0$, the extended model collapses into the standard HOPIT model. As a result, we can use a standard likelihood ratio test to directly test the assumption of heteroskedasticity.

Under the assumption of response consistency (Eq. 5) and our relaxed assumption of vignette equivalence (Eq. 7), the probability that a respondent i rates a vignette j as k, conditional on both u_i and ε_i, is given by

$$\begin{aligned} &\Pr\left(V_{ij} = k \mid x_i, z_i, u_i, u_i, \varepsilon_i\right) \\ &\quad = \Phi\left(\frac{\tau_i^k - \theta_j}{\sigma_v \exp\left(\alpha\left(\beta' x_i + \varepsilon_i - \theta_j\right)^2\right)}\right) - \Phi\left(\frac{\tau_i^{k-1} - \theta_j}{\sigma_v \exp\left(\alpha\left(\beta' x_i + \varepsilon_i - \theta_j\right)^2\right)}\right) \end{aligned} \tag{8}$$

Estimates for the model parameters can be obtained by maximizing the log-likelihood $L = \sum_{i=1}^{N} \log \ell_i$, with ℓ_i given by

$$\begin{aligned} &\ell_i\left(\beta, \gamma^1, \ldots, \gamma^{K-1}, \theta_1, \ldots, \theta_J, \sigma_v, \sigma_u, \alpha\right) = \int\int \prod_{k_0=1}^{K} \left[1\left\{\tau_i^{k_0-1} \leq \beta' x_i + \varepsilon_i < \tau_i^{k_0}\right\}\right]^{1\{y_i = k_0\}} \\ &\prod_{j=1}^{J}\prod_{k_j=1}^{K}\left[\Phi\left(\frac{\tau_i^{k_j} - \theta_j}{\sigma_v \exp\left(\alpha\left(\beta' x_i + \varepsilon_i - \theta_j\right)^2\right)}\right) - \Phi\left(\frac{\tau_i^{k_j-1} - \theta_j}{\sigma_v \exp\left(\alpha\left(\beta' x_i + \varepsilon_i - \theta_j\right)^2\right)}\right)\right]^{1\{V_{ij} = k_j\}} \\ &\qquad d\Phi\left(\varepsilon_i\right) d\Phi\left(u_i\right) \end{aligned} \tag{9}$$

where $\Phi\left(\right)$ is the standard normal cumulative distribution function.

Comparing the likelihood of the extended HOPIT model with that of the standard HOPIT model, one can see that in the extended model the contribution of the perception of a vignette j by an individual i to the likelihood is weighted by its exponential distance from i to j.

We use a quasi-Monte Carlo method to evaluate the integral in the likelihood function. To reduce the variance of the Monte Carlo integration and thereby obtain more precise estimates of model parameters for a given number of iterations, we use the technique of importance sampling, which is detailed in S3 Appendix.

Results

Simulation studies

Data generating process and model fit measures In this section, we explore the finite sample performance of our extended HOPIT model through a Monte Carlo study. First, we investigate the potential bias of standard HOPIT models, which are misspecified in the presence of perception heteroskedasticity. Second, we examine whether our extended HOPIT models produces comparable results with the standard HOPIT model in the absence of perception heteroskedasticity.

The simulated datasets are generated from the extended HOPIT model as follows: (1) We set the number of response categories K to 3. (2) We generate two exogenous variables: x_{i1} is drawn from a $U(0, 1)$ distribution, which corresponds to continuous regressors such as age or years of schooling, and x_{i2} is drawn from a Bernoulli distribution with equal probabilities, which represents binary regressors such as sex. (3) We set model parameters to the values described in Table 1. (4) We set the number of vignettes J to 1, 3 and 5. (5) We set the parameter of heteroskedasticity α to 0, 0.1 and 0.5. (6) We set the number of observations N to 500, $1,000$ and $2,000$. (7) Each Monte Carlo experiment is replicated 100 times.

The value of α is chosen so that it represents three usual DGPs: no heteroskedasticity ($\alpha = 0$), weak heteroskedasticity ($\alpha = 0.1$) and strong heteroskedasticity ($\alpha = 0.5$). When heteroskedasticity (weak or strong) is present, the standard HOPIT model – which does not take the heteroskedasticity into account – is under-specified. In contrast, our extended HOPIT model is over-specified when there is no heteroskedasticity.

For each generated dataset, we estimate both the standard HOPIT model and our extended HOPIT model. For each Monte Carlo experiment, multiple measures are used to evaluate model fit. Firstly, we calculate the mean squared error (MSE) of the model parameters across all replications of the experiment. Secondly, the mean values of the log-likelihoods and AICs are reported to compare the overall model fit. Finally, since the standard HOPIT model is nested in our extended HOPIT model, likelihood ratio tests are also used to compare the two models. The test results are summarized by the rejection rate of the standard HOPIT model across all replications in each experiment. Note that the size of the likelihood ratio test is 0.05 and the test statistic is compared with the critical value of a standard χ^2-distribution.

Results The results from our simulations depend on the specifications for the value of α, the number of vignettes J and the number of observations N.

Tables 1-3 summarize the results of the experiments with weak heteroskedasticity ($\alpha = 0.1$). In the experiments with only one vignette, the standard HOPIT model is not always rejected by the likelihood ratio test, with rejection rates of 13%, 27% and 48% for $N = 500$, $1,000$ and $2,000$, respectively. As the number of vignettes or the number of observations increases, rejection rates for the standard HOPIT model increase. When we have three vignettes, the rejection rates rise to 57%, 86% and 100% for $N = 500$, $1,000$ and $2,000$, respectively. For experiments with five vignettes, we have rejection rates of 90% for $N = 500$ and of 100% for $N = 1,000$ and $N = 2,000$. In terms of information criteria, such as the AIC, the extended HOPIT model shows a better model fit in all but one experiment when $J = 1$ and $N = 500$, in which the extended HOPIT model yields a slightly larger AIC than the standard HOPIT (1954.393 vs. 1954.361). The extended HOPIT model also almost always has smaller or equal MSEs than the standard HOPIT model, again with a few exceptions for cases with only one vignette and a small sample size.

Tables 4-6 summarize the results of experiments with strong heteroskedasticity ($\alpha = 0.5$). Even for cases with small sample sizes $N = 500$ and $N = 1000$ and only one vignette, we still obtain very high rejection rates of 81% and 91% for the standard HOPIT model. In other experiments with larger sample sizes and more available vignettes, the standard HOPIT model is always rejected. In all experiments, including the one with $N = 500$ and $J = 1$, the extended HOPIT model has a smaller AIC than the standard HOPIT model.

In addition, the extended HOPIT model gives more accurate estimates than the standard HOPIT model. With few exceptions, the extended HOPIT model has smaller MSEs of the parameters than the standard HOPIT model. Both the standard and the extended HOPIT have increasingly smaller MSE as the sample sizes increase or as the number of vignettes increases, but the difference between the two models remains and the extended HOPIT always dominates the standard HOPIT. Take the estimate of β_2 for example. For the experiment with a moderate sample size and a moderate number of vignettes ($N = 1,000$ and $J = 3$), the MSEs of β_2 are 8.8% for the standard HOPIT model and 2.9% for the extended HOPIT model. For the experiment with a large sample size and more vignettes ($N = 2,000$ and $J = 5$), the MSEs of β_2 are 5.3% for the standard HOPIT model and 1.2% for the extended HOPIT model.

Tables 7-9 summarize the results of our experiments with no heteroskedasticity ($\alpha = 0$). Although the extended HOPIT model is over-specified in this case, we do not observe much information loss using the extended HOPIT model against the standard HOPIT model. Firstly, we notice that the estimate of α is always near its true value of zero. The bias of the estimate shrinks as the number of observations or the number of vignettes increases. For the experiment with a moderate sample size and a moderate number of vignettes ($N = 1,000$ and $J = 3$), the bias is already near zero. Secondly, the standard HOPIT model and the extended HOPIT model have almost equal log-likelihoods and AICs. Lastly, the two models exhibit the same level of unbiasedness implied by the similar MSEs of parameters.

Overall, our Monte Carlo experiments show that, in the presence of heteroskedasticity of vignette perceptions, the extended HOPIT model has a better model fit than the standard HOPIT model in terms of the MSEs of the parameters, the AICs and LR tests, for datasets of moderate sample size and more than one vignette. Moreover, in the absence of heteroskedasticity, no information loss has been found using the extended HOPIT model in general.

Table 1. MSE of parameters and model fit for experiments where $\alpha = 0.1$ and $J = 1$

Parameter	True value	$N = 500$		$N = 1000$		$N = 2000$	
		Standard HOPIT	Extended HOPIT	Standard HOPIT	Extended HOPIT	Standard HOPIT	Extended HOPIT
β_1: x_{i1}	1	0.140	0.139	0.086	0.088	0.032	0.029
β_2: x_{i2}	1	0.095	0.077	0.037	0.034	0.017	0.014
γ_{10}: *intercept*	0	0.040	0.036	0.020	0.019	0.009	0.009
γ_{11}: x_{i1}	1	0.142	0.134	0.084	0.084	0.033	0.031
γ_{12}: x_{i2}	1	0.086	0.070	0.033	0.029	0.018	0.015
γ_{20}: *intercept*	0	0.044	0.035	0.019	0.017	0.012	0.010
γ_{21}: x_{i1}	0.1	0.040	0.040	0.019	0.019	0.009	0.010
γ_{22}: x_{i2}	0.1	0.010	0.011	0.006	0.006	0.002	0.003
σ_u	1	0.064	0.041	0.023	0.016	0.020	0.010
σ_v	1	0.151	0.074	0.066	0.034	0.060	0.020
α	0.1	-	0.009	-	0.005	-	0.003
Goodness of fit							
log-likelihood	-	966.181	965.197	1941.455	1940.195	3884.547	3882.342
AIC	-	1954.361	1954.393	3904.910	3904.390	7791.093	7788.684
LR test (% rej. $\alpha = 0$)	-	13		27		48	

Table 2. MSE of parameters and model fit for experiments where $\alpha = 0.1$ and $J = 3$

Parameter	True value	$N = 500$		$N = 1000$		$N = 2000$	
		Standard HOPIT	Extended HOPIT	Standard HOPIT	Extended HOPIT	Standard HOPIT	Extended HOPIT
β_1: x_{i1}	1	0.095	0.075	0.020	0.019	0.020	0.019
β_2: x_{i2}	1	0.058	0.041	0.014	0.007	0.014	0.007
γ_{10}: *intercept*	0	0.041	0.036	0.012	0.010	0.012	0.010
γ_{11}: x_{i1}	1	0.092	0.079	0.020	0.016	0.020	0.016
γ_{12}: x_{i2}	1	0.062	0.042	0.013	0.008	0.013	0.008
γ_{20}: *intercept*	0	0.034	0.024	0.010	0.005	0.010	0.005
γ_{21}: x_{i1}	0.1	0.020	0.020	0.004	0.004	0.004	0.004
γ_{22}: x_{i2}	0.1	0.007	0.008	0.002	0.002	0.002	0.002
σ_u	1	0.049	0.024	0.020	0.005	0.020	0.005
σ_v	1	0.147	0.042	0.082	0.007	0.082	0.007
α	0.1	-	0.002	-	0.001	-	0.001
Goodness of fit							
log-likelihood	-	1822.894	1819.911	7319.352	7308.166	7319.352	7308.166
AIC	-	3671.787	3667.823	14664.704	14644.332	14664.704	14644.332
LR test (% rej. $\alpha = 0$)	-	57		86		100	

Table 3. MSE of parameters and model fit for experiments where $\alpha = 0.1$ and $J = 5$

Parameter	True value	$N = 500$		$N = 1000$		$N = 2000$	
		Standard HOPIT	Extended HOPIT	Standard HOPIT	Extended HOPIT	Standard HOPIT	Extended HOPIT
β_1: x_{i1}	1	0.071	0.059	0.038	0.031	0.017	0.012
β_2: x_{i2}	1	0.036	0.020	0.024	0.014	0.011	0.005
γ_{10}: *intercept*	0	0.031	0.027	0.016	0.013	0.010	0.008
γ_{11}: x_{i1}	1	0.059	0.049	0.035	0.026	0.017	0.012
γ_{12}: x_{i2}	1	0.043	0.024	0.022	0.011	0.014	0.006
γ_{20}: *intercept*	0	0.030	0.016	0.017	0.008	0.013	0.005
γ_{21}: x_{i1}	0.1	0.013	0.013	0.005	0.006	0.004	0.005
γ_{22}: x_{i2}	0.1	0.005	0.006	0.002	0.003	0.001	0.002
σ_u	1	0.045	0.014	0.028	0.007	0.024	0.004
σ_v	1	0.136	0.017	0.108	0.008	0.096	0.005
α	0.1	-	0.001	-	0.000	-	0.000
Goodness of fit							
log-likelihood	-	2660.522	2653.699	5323.627	5309.145	10655.995	10629.549
AIC	-	5351.043	5339.397	10677.253	10650.291	21341.991	21291.098
LR test (% rej. $\alpha = 0$)	-	90		100		100	

Table 4. MSE of parameters and model fit for experiments where $\alpha = 0.5$ and $J = 1$

Parameter	True value	$N = 500$		$N = 1000$		$N = 2000$	
		Standard HOPIT	Extended HOPIT	Standard HOPIT	Extended HOPIT	Standard HOPIT	Extended HOPIT
β_1: x_{i1}	1	1.006	0.278	0.187	0.137	0.095	0.064
β_2: x_{i2}	1	1.369	0.133	0.135	0.058	0.084	0.026
γ_{10}: *intercept*	0	0.069	0.038	0.029	0.021	0.013	0.009
γ_{11}: x_{i1}	1	1.038	0.293	0.197	0.144	0.111	0.070
γ_{12}: x_{i2}	1	1.506	0.135	0.128	0.051	0.095	0.030
γ_{20}: *intercept*	0	0.217	0.053	0.081	0.030	0.078	0.022
γ_{21}: x_{i1}	0.1	0.043	0.044	0.020	0.023	0.011	0.012
γ_{22}: x_{i2}	0.1	0.012	0.014	0.006	0.007	0.003	0.004
σ_u	1	2.014	0.102	0.267	0.051	0.238	0.034
σ_v	1	6.459	0.185	1.627	0.122	1.457	0.067
α	0.5	-	0.054	-	0.037	-	0.025
Goodness of fit							
log-likelihood	-	982.303	977.999	1974.477	1967.152	3946.218	3933.152
AIC	-	1986.607	1979.999	3970.955	3958.305	7914.436	7890.303
LR test (% rej. $\alpha = 0$)	-	81		91		100	

Table 5. MSE of parameters and model fit for experiments where $\alpha = 0.5$ and $J = 3$

Parameter	True value	$N = 500$		$N = 1000$		$N = 2000$	
		Standard HOPIT	Extended HOPIT	Standard HOPIT	Extended HOPIT	Standard HOPIT	Extended HOPIT
β_1: x_{i1}	1	0.335	0.106	0.081	0.057	0.053	0.027
β_2: x_{i2}	1	0.365	0.048	0.088	0.029	0.061	0.016
γ_{10}: *intercept*	0	0.081	0.040	0.030	0.019	0.019	0.012
γ_{11}: x_{i1}	1	0.423	0.120	0.106	0.063	0.072	0.033
γ_{12}: x_{i2}	1	0.466	0.068	0.111	0.034	0.078	0.019
γ_{20}: *intercept*	0	0.235	0.042	0.125	0.024	0.104	0.015
γ_{21}: x_{i1}	0.1	0.025	0.027	0.018	0.017	0.008	0.007
γ_{22}: x_{i2}	0.1	0.013	0.011	0.006	0.006	0.003	0.002
σ_u	1	0.758	0.040	0.324	0.024	0.298	0.016
σ_v	1	3.975	0.086	2.267	0.052	2.143	0.027
α	0.5	-	0.019	-	0.015	-	0.008
Goodness of fit							
log-likelihood	-	1906.972	1885.884	3826.655	3785.055	7658.822	7575.962
AIC	-	3839.943	3799.768	7679.310	7598.110	15343.643	15179.924
LR test (% rej. $\alpha = 0$)	-	100		100		100	

Table 6. MSE of parameters and model fit for experiments where $\alpha = 0.5$ and $J = 5$

Parameter	True value	$N = 500$		$N = 1000$		$N = 2000$	
		Standard HOPIT	Extended HOPIT	Standard HOPIT	Extended HOPIT	Standard HOPIT	Extended HOPIT
β_1: x_{i1}	1	0.172	0.071	0.092	0.035	0.035	0.020
β_2: x_{i2}	1	0.158	0.030	0.107	0.024	0.053	0.012
γ_{10}: *intercept*	0	0.049	0.026	0.031	0.015	0.019	0.011
γ_{11}: x_{i1}	1	0.168	0.068	0.130	0.044	0.059	0.023
γ_{12}: x_{i2}	1	0.206	0.035	0.132	0.023	0.088	0.015
γ_{20}: *intercept*	0	0.192	0.030	0.157	0.018	0.133	0.012
γ_{21}: x_{i1}	0.1	0.023	0.022	0.011	0.010	0.007	0.005
γ_{22}: x_{i2}	0.1	0.008	0.007	0.004	0.003	0.003	0.002
σ_u	1	0.539	0.027	0.428	0.016	0.357	0.012
σ_v	1	3.263	0.038	2.891	0.026	2.536	0.019
α	0.5	-	0.011	-	0.008	-	0.006
Goodness of fit							
log-likelihood	-	2831.483	2784.466	5662.503	5567.058	11342.691	11153.864
AIC	-	5692.966	5600.933	11355.005	11166.117	22715.382	22339.729
LR test (% rej. $\alpha = 0$)	-	100		100		100	

Table 7. MSE of parameters and model fit for experiments where $\alpha = 0$ and $J = 1$

Parameter	True value	$N = 500$		$N = 1000$		$N = 2000$	
		Standard HOPIT	Extended HOPIT	Standard HOPIT	Extended HOPIT	Standard HOPIT	Extended HOPIT
β_1: x_{i1}	1	0.106	0.108	0.071	0.071	0.028	0.027
β_2: x_{i2}	1	0.067	0.064	0.030	0.030	0.013	0.012
γ_{10}: *intercept*	0	0.036	0.035	0.018	0.018	0.008	0.008
γ_{11}: x_{i1}	1	0.119	0.124	0.071	0.070	0.028	0.027
γ_{12}: x_{i2}	1	0.061	0.059	0.027	0.026	0.013	0.013
γ_{20}: *intercept*	0	0.036	0.034	0.016	0.016	0.008	0.007
γ_{21}: x_{i1}	0.1	0.037	0.038	0.017	0.017	0.009	0.009
γ_{22}: x_{i2}	0.1	0.011	0.011	0.006	0.006	0.002	0.002
σ_u	1	0.034	0.030	0.012	0.011	0.007	0.007
σ_v	1	0.061	0.056	0.024	0.026	0.012	0.015
α	0	-	0.010	-	0.003	-	0.002
Goodness of fit							
log-likelihood	-	959.445	958.991	1928.574	1928.285	3859.206	3858.782
AIC	-	1940.890	1941.982	3879.147	3880.570	7740.413	7741.564
LR test (% rej. $\alpha = 0$)	-	8		1		6	

Table 8. MSE of parameters and model fit for experiments where $\alpha = 0$ and $J = 3$

Parameter	True value	$N = 500$		$N = 1000$		$N = 2000$	
		Standard HOPIT	Extended HOPIT	Standard HOPIT	Extended HOPIT	Standard HOPIT	Extended HOPIT
β_1: x_{i1}	1	0.069	0.068	0.032	0.031	0.017	0.017
β_2: x_{i2}	1	0.033	0.032	0.014	0.014	0.006	0.006
γ_{10}: *intercept*	0	0.035	0.035	0.014	0.014	0.010	0.010
γ_{11}: x_{i1}	1	0.061	0.060	0.020	0.020	0.013	0.013
γ_{12}: x_{i2}	1	0.030	0.030	0.010	0.010	0.006	0.006
γ_{20}: *intercept*	0	0.017	0.017	0.005	0.005	0.003	0.003
γ_{21}: x_{i1}	0.1	0.016	0.015	0.009	0.009	0.004	0.004
γ_{22}: x_{i2}	0.1	0.006	0.006	0.004	0.004	0.002	0.002
σ_u	1	0.014	0.013	0.005	0.005	0.002	0.002
σ_v	1	0.018	0.018	0.004	0.006	0.003	0.003
α	0	-	0.001	-	0.000	-	0.000
Goodness of fit							
log-likelihood	-	1777.444	1777.190	3562.782	3562.603	7137.357	7137.150
AIC	-	3580.889	3582.380	7151.563	7153.207	14300.714	14302.299
LR test (% rej. $\alpha = 0$)	-	2		3		3	

Table 9. MSE of parameters and model fit for experiments where $\alpha = 0$ and $J = 5$

Parameter	True value	$N = 500$		$N = 1000$		$N = 2000$	
		Standard HOPIT	Extended HOPIT	Standard HOPIT	Extended HOPIT	Standard HOPIT	Extended HOPIT
β_1: x_{i1}	1	0.056	0.055	0.027	0.026	0.011	0.011
β_2: x_{i2}	1	0.021	0.020	0.011	0.011	0.005	0.005
γ_{10}: *intercept*	0	0.026	0.026	0.013	0.013	0.009	0.009
γ_{11}: x_{i1}	1	0.042	0.041	0.020	0.020	0.011	0.011
γ_{12}: x_{i2}	1	0.023	0.021	0.009	0.009	0.005	0.005
γ_{20}: *intercept*	0	0.013	0.013	0.006	0.006	0.002	0.002
γ_{21}: x_{i1}	0.1	0.013	0.013	0.005	0.005	0.004	0.004
γ_{22}: x_{i2}	0.1	0.004	0.004	0.002	0.002	0.001	0.001
σ_u	1	0.012	0.011	0.004	0.004	0.002	0.002
σ_v	1	0.012	0.012	0.004	0.004	0.002	0.002
α	0	-	0.000	-	0.000	-	0.000
Goodness of fit							
log-likelihood	-	2563.538	2563.329	5129.328	5129.104	10271.605	10271.477
AIC	-	5157.076	5158.658	10288.656	10290.208	20573.211	20574.954
LR test (% rej. $\alpha = 0$)	-	3		1		0	

Empirical application

Data As shown in the simulation study, the advantage of our extended HOPIT model against the standard HOPIT depends on the size of potential heteroskedasticity, the number of observations and the number of vignettes available in the dataset. In our empirical application, we use data on Chinese adults from the WHO Study on Global Aging and Adult Health (SAGE). SAGE asked respondents to complete self-assessment questions and questions regarding vignettes. Each self-assessment question was supplemented with five vignette questions, which described varying levels of functional limitations in a health domain. As an example, we include survey questions on self-assessment and anchoring vignettes on visual acuity from SAGE in Table 10 and survey questions on domains of pain, cognition and mobility are included in S1 Appendix. We speculate that perception on vignettes of extreme scenarios such as Vignette 5 might be quite imprecise and the degree of precision could vary largely among respondents.

Table 11 shows the distributions of the self-assessments and vignette responses in each domain. The majority of the respondents perceive no physical limitations in each domain, and only a few have “severe” or “extreme” self-assessments. The distributions of the vignettes differ greatly, indicating that the information content of a vignette, which is the degree of limitation in our example, varies significantly.

In this application, we would examine how the actual health status as well as the health perception of the respondents are affected by socio-demographic factors such as age, gender, education level, partnered and chronic disease. Self-assessed vision, body pain, cognition, and mobility, are studied using both the standard HOPIT model and the extended HOPIT model. We keep individuals over 50 years and with no missing values for any self-assessment, vignette and covariate of interest in our samples. To maintain as much information as possible, we use different samples for the analyses of different domains.

Table 12 describes the covariates in our analysis, including dummy variables for age-groups of five years, education levels categorized as primary, secondary and higher

Table 10. Questions from SAGE on self-assessments and anchoring vignettes on visual acuity

	Questions
Self-assessment	In the last 30 days, how much difficulty did you have in seeing and recognizing an object or a person you know across the road (from a distance of about 20 meters)?
Vignette 1	[X] can read words in newspaper articles (and can recognize faces on a postcard size photograph). He can recognize shapes and colors from across 20 meters but misses out the fine details. Overall in the last 30 days, how much difficulty do you think [X] had in seeing and recognizing a person he knows across the road (from a distance of about 20 meters)?
Vignette 2	[X] only reads if the text is in very large print, such as 10 lines per page. Otherwise she does not read anything. Even when people are close to her, she sees them blurred. Overall in the last 30 days, how much difficulty do you think [X] had in seeing and recognizing a person he knows across the road (from a distance of about 20 meters)?
Vignette 3	[X] needs a magnifying glass to read small print and look at details on pictures. He also takes a while to recognize objects if they are too far from him. Overall in the last 30 days, how much difficulty do you think [X] had in seeing and recognizing a person he knows across the road (from a distance of about 20 meters)?
Vignette 4	[X] can read words in newspaper articles (and can recognize faces on a postcard size photograph). He can recognize familiar people's faces all the time and picks out most details in pictures from across 20 meters. Overall in the last 30 days, how much difficulty do you think [X] had in seeing and recognizing a person he knows across the road (from a distance of about 20 meters)?
Vignette 5	[X] cannot detect any movement close to the eyes or even the presence of a light. Overall in the last 30 days, how much difficulty do you think [X] had in seeing and recognizing a person he knows across the road (from a distance of about 20 meters)?

levels of education, gender, being partnered and a dummy for chronic disease. The summary statistics of these variables are shown in Table 13, where we observe that about half of the respondents are female, most of the respondents only have completed primary or secondary education, over 85% of the respondents live with their spouses, and nearly half have at least one chronic disease.

Results Tables 14 - 17 present parameter estimates (of outcome equations) using both the standard HOPIT model and our extended HOPIT model. The complete estimates of all parameters are provided in S2 Appendix. We observe that the extended HOPIT has a smaller AIC than the standard HOPIT model, and we can always reject the standard HOPIT model by the likelihood ratio test in any of the four health domains studied. This implies that the extended HOPIT has a better model fit than the standard HOPIT model.

We observe that estimates from the standard and extended HOPIT differ. Regarding visual acuity, our extended HOPIT model shows smaller effects of age and gender but larger effects of education and chronic disease by comparison with estimates in the standard HOPIT model. Concerning body pain, the extended HOPIT shows a smaller effect of education but larger effects of age, gender, and chronic disease than that of the standard HOPIT. Concerning cognition, estimates of age effects are very similar between

Table 11. Self-assessment and vignette evaluations (%)

	Self-assessment	Vignette 1	Vignette 2	Vignette 3	Vignette 4	Vignette 5
Vision						
None	66.09	42.58	6.00	3.64	86.18	1.73
Mild	24.76	42.04	13.16	17.60	7.96	2.18
Moderate	7.16	12.8	38.44	34.58	3.78	4.04
Severe	1.73	2.49	35.87	36.27	1.91	15.07
Extreme	0.27	0.09	6.53	7.91	0.18	76.98
Pain						
None	56.59	2.14	8.94	8.31	77.52	2.14
Mild	30.04	34.19	60.86	54.57	13.38	2.37
Moderate	11.20	44.91	25.17	28.89	4.83	6.53
Severe	2.02	17.41	4.71	7.99	2.57	29.13
Extreme	0.16	1.35	0.32	0.24	1.70	59.83
Cognition						
None	58.85	6.92	3.24	51.47	92.10	1.88
Mild	30.30	50.82	27.27	38.08	4.55	3.64
Moderate	8.89	35.50	43.82	8.97	1.97	12.04
Severe	1.97	6.55	23.26	1.39	0.66	35.83
Extreme	0.00	0.20	2.42	0.08	0.74	46.60
Mobility						
None	83.82	44.40	3.06	9.85	94.05	2.43
Mild	11.66	40.92	10.19	33.29	2.68	1.84
Moderate	3.31	12.24	34.88	38.16	2.10	4.03
Severe	1.09	2.43	43.31	17.32	0.96	26.42
Extreme	0.13	1.00	8.55	1.38	0.21	65.28

Note: Different samples are used for each domain. Samples are selected on the availability (i.e., no missing value) of self-assessment and vignettes of each domain and covariates included in our analysis. We have 2250 obs. in vision, 2527 obs. in body pain, 2442 obs. in cognition and 2385 obs. in mobility.

Table 12. Description of covariates

Covariate	Description
Age	We include dummy variables for age groups of five years, namely, 50-54 (baseline group), 55-59, 60-64, 65-69, 70-74, 75-79 and 80 or above.
Education	In SAGE, education is measured as: (1) less than primary education, (2) primary school completed, (3) second school completed, (4) high school completed, (5) college, pre-university or university completed, and (6) post graduate degree completed. We define level (1) and (2) as primary education (baseline category), (3) and (4) as secondary education, and (5) and (6) as higher education.
Female	A dummy variable for gender.
Partnered	In SAGE, marital status belongs to one of the following categories: (1) never married, (2) currently married, (3) cohabiting, (4) separated/divorced and (5) widowed. We include a dummy variable for whether a respondent is currently living alone or not. The dummy has a value of 1 for individuals who are either (2) currently married or (3) cohabiting, and 0 otherwise.
Chronic disease	A dummy variable for diagnosed chronic diseases, such as arthritis, stroke, angina, diabetes, chronic lung disease, asthma, depression, and hypertension.

the two models; the gradient of education is smaller in the extended HOPIT; the gradients of partnered and chronic disease are larger in the extended HOPIT. Finally, with respect to mobility, the extended HOPIT demonstrates smaller effects for all

Table 13. Summary statistics

Variable	Vision		Pain		Cognition		Mobility	
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
Age 50-54 years	0.255	0.436	0.255	0.436	0.249	0.432	0.258	0.438
Age 55-59 years	0.253	0.435	0.240	0.427	0.246	0.431	0.237	0.426
Age 60-64 years	0.164	0.370	0.178	0.383	0.182	0.386	0.178	0.382
Age 65-69 years	0.143	0.350	0.132	0.338	0.129	0.335	0.131	0.337
Age 70-74 years	0.100	0.299	0.095	0.293	0.103	0.304	0.106	0.308
Age 75-79 years	0.056	0.229	0.061	0.239	0.062	0.241	0.058	0.234
Age >= 80 years	0.030	0.171	0.039	0.194	0.030	0.170	0.032	0.177
Female	0.459	0.498	0.469	0.499	0.476	0.500	0.449	0.498
Primary education	0.520	0.500	0.501	0.500	0.496	0.500	0.497	0.500
Secondary education	0.434	0.496	0.440	0.497	0.432	0.495	0.443	0.497
Higher education	0.047	0.211	0.058	0.234	0.072	0.258	0.060	0.238
Partnered	0.872	0.334	0.898	0.302	0.881	0.324	0.879	0.326
Chronic disease	0.482	0.500	0.508	0.500	0.509	0.500	0.500	0.500
N	2,250		2,527		2,442		2,385	

Table 14. Estimates of self-assessed vision

	Standard HOPIT		Extended HOPIT	
	Estimate	t-statistic	Estimate	t-statistic
Age 55-59 years	0.160*	1.789	0.134*	1.841
Age 60-64 years	0.403***	4.287	0.379***	5.041
Age 65-69 years	0.575***	5.952	0.547***	7.122
Age 70-74 years	0.670***	6.606	0.591***	6.950
Age 75-79 years	0.806***	7.000	0.723***	7.049
Age >= 80 years	1.063***	7.082	0.990***	6.460
Female	0.239***	3.537	0.208***	4.042
Secondary education	-0.180**	-2.655	-0.234***	-4.202
Higher education	-0.240*	-1.789	-0.338***	-2.886
Partnered	-0.051	-0.569	-0.014	-0.220
Chronic disease	0.183***	2.660	0.241***	4.649
σ_u	0.374***	18.269	0.382***	18.352
σ_v	0.952***	25.721	0.870***	29.110
α	-	-	0.024***	16.398
AIC	27,491		27,150	
LR test	$\chi^2 = 343.085$, $p-value = 0.000$			

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. LR test has the extended HOPIT as the full model and standard HOPIT as the reduced one. $N = 2,250$.

variables except higher education, female and one age group dummy (*Age 55-59 years*).

The estimates of parameter α, which measure the degree of heteroskedasticity, are statistically significant and take on similar values of 0.024, 0.025, 0.024 and 0.029 in the analyses of vision, pain, cognition, and mobility, respectively. To assess the degree of heteroskedasticity in the error term of the vignette equation (Eq. 7), we evaluate the heteroskedastic error (after integrating out the unobserved ε_i) which is given by

$$\mathbb{E}_{\varepsilon_i}\left\{\sigma_v \exp\left(\alpha\left(\beta' x_i + \varepsilon_i - \theta_j\right)^2\right)\right\}$$

The derivation of the expression is included in S4 Appendix. We compare this heteroskedastic error with the standard error in the standard HOPIT.

Table 15. Estimates of self-assessed pain

	Standard HOPIT		Extended HOPIT	
	Estimate	t-statistic	Estimate	t-statistic
Age 55-59 years	-0.085	-1.108	-0.043	-0.539
Age 60-64 years	0.136	1.551	0.160*	1.851
Age 65-69 years	0.123	1.246	0.192**	2.164
Age 70-74 years	0.123	1.160	0.198**	2.176
Age 75-79 years	0.259**	2.342	0.283***	2.938
Age >= 80 years	0.244**	1.993	0.290**	2.560
Female	0.174***	3.346	0.207***	2.877
Secondary education	-0.187***	-3.463	-0.140**	-2.357
Higher education	-0.528***	-4.922	-0.440***	-5.163
Partnered	0.051	0.607	0.036	0.446
Chronic disease	0.432***	7.660	0.458***	7.353
σ_u	0.276***	20.351	0.303***	21.323
σ_v	0.736***	34.386	0.692***	38.484
α	-	-	0.025***	12.290
AIC	32,451		32,332	
LR test	$\chi^2 = 121.559$, $p-value = 0.000$			

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. LR test has the extended HOPIT as the full model and standard HOPIT as the reduced one. $N = 2,527$.

Table 16. Estimates of self-assessed cognition

	Standard HOPIT		Extended HOPIT	
	Estimate	t-statistic	Estimate	t-statistic
Age 55-59 years	-0.069	-0.839	-0.070	-0.964
Age 60-64 years	0.216**	2.409	0.208***	2.592
Age 65-69 years	0.335***	3.412	0.325***	4.413
Age 70-74 years	0.562***	5.553	0.570***	7.152
Age 75-79 years	0.437***	4.091	0.429***	5.432
Age >= 80 years	0.692***	5.535	0.667***	7.839
Female	0.013	0.231	-0.028	-0.436
Secondary education	-0.314***	-5.252	-0.256***	-4.187
Higher education	-0.756***	-6.983	-0.682***	-7.382
Partnered	-0.136*	-1.733	-0.176***	-4.056
Chronic disease	0.240***	3.908	0.274***	5.346
σ_u	0.285***	16.566	0.311***	17.301
σ_v	0.842***	30.748	0.756***	42.292
α	-	-	0.024***	10.792
AIC	29,223		29,109	
LR test	$\chi^2 = 115.631$, $p-value = 0.000$			

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. LR test has the extended HOPIT as the full model and standard HOPIT as the reduced one. $N = 2442$.

Fig 5. Probability density of heteroskedastic errors with data of vision. Probability densities of heteroskedastic errors of the five vignettes are plotted from top to bottom. The straight red line indicates the standard error from the standard HOPIT model.

Table 17. Estimates of self-assessed mobility

	Standard HOPIT		Extended HOPIT	
	Estimate	t-statistic	Estimate	t-statistic
Age 55-59 years	-0.026	-0.261	-0.048	-0.577
Age 60-64 years	0.113	1.138	0.111	1.269
Age 65-69 years	0.366***	3.558	0.292***	3.531
Age 70-74 years	0.485***	4.611	0.398***	3.935
Age 75-79 years	0.573***	4.769	0.476***	4.565
Age >= 80 years	0.917***	6.719	0.911***	6.165
Female	-0.035	-0.468	-0.091*	-1.881
Secondary education	-0.384***	-5.009	-0.270***	-4.743
Higher education	-0.584***	-4.216	-0.592***	-4.585
Partnered	0.032	0.333	-0.029	-0.492
Chronic disease	0.349***	4.824	0.292***	5.333
σ_u	0.306***	16.205	0.296***	18.740
σ_v	0.673***	18.967	0.561***	22.007
α	-	-	0.029***	14.348
AIC	26,568		26,336	
LR test	$\chi^2 = 233.612$, $p - value = 0.000$			

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. LR test has the extended HOPIT as the full model and standard HOPIT as the reduced one. $N = 2385$.

Figure 5 shows the probability density of the heteroskedastic errors for each vignette for the analysis of vision. Apparently, the standard errors are different for each vignette. For most respondents, the first and the fourth vignettes have standard errors smaller than the standard error in the standard HOPIT, while the other three vignettes, especially the last one, have standard errors larger than that in the standard HOPIT. For a given vignette, standard errors also differ across respondents. Especially for the last vignette, standard errors show a relatively large dispersion. Given the extreme scenario described in the last vignette, it is not surprising to observe that on average respondents are less precise in its perception and the degree of precision varies greatly among respondents.

We can safely reject the standard HOPIT based on our test on the significance of α which measures the degree of perception heteroskedasticity in our extended model. Yet, the standard HOPIT can also be rejected when other maintained assumptions such as response consistency are violated. If so, our extended HOPIT model could be favored by test results simply because it captures other aspects of model deviation from the truth. If the vignette perception is directly observable, the heteroskedasticity assumption can be easily tested by an auxiliary regression which regresses the squared residuals derived from Eq. 7 on the distance between vignettes and respondent health indexes. However, the ordinal nature of vignette assessments mapped from the latent unobservable vignette perceptions precludes such a test. Although an overidentification test like ours could also be affected by model deviation other than vignette perception heteroskedasticity, our extension seems to be able to capture such deviation better than the standard HOPIT indicated by a better overall model fit and hence result in a better estimate of the outcome equation.

Conclusions

The HOPIT model introduced by [1] has been used in a large number of studies to account for reporting heterogeneity of self-assessments when anchoring vignettes are available. The validity of the model, nevertheless, relies on the assumptions of response consistency and vignette equivalence. The assumption of vignette equivalence assumes that perceptions of vignettes are the same across individuals. In this paper, we relax the assumption of vignette equivalence by allowing for heteroskedasticity of perceptions of vignettes across individuals. Particularly, we assume that the perception precision of a vignette by an individual is negatively proportional to the (exponential) distance between the individual and the vignette, which measures the similarity/dissimilarity between the individual and the vignette.

A series of Monte Carlo simulations show that the extended HOPIT model has a better model fit than the standard HOPIT in terms of MSEs of the parameters, AICs and LR tests in the presence of heteroskedasticity of vignette perceptions. In the absence of heteroskedasticity, we find almost no information loss using the extended HOPIT model.

When we adopt the extended HOPIT model to analyze self-assessments of four health domains using Chinese data from SAGE, we find that the extended HOPIT model has a better model fit than the standard HOPIT model, as indicated by both information criteria and likelihood ratio tests. The standard errors of the extended HOPIT could significantly differ from that of the standard HOPIT, especially for vignettes which describe extreme scenarios.

Overall, compared with the standard HOPIT model, our extended HOPIT model facilitates a more flexible way of utilizing information of anchoring vignettes and seems to often have a better model fit than the standard HOPIT model. One potential disadvantage of our model is that we assume the similarity/dissimilarity constituents the only source of heteroskedasticity of vignette perceptions. One of the future research directions on the extension of the standard HOPIT model would be finding statistical models which allow for more generic form of heteroskedasticity and meanwhile impose no identification difficulty.

Our extended model also sheds light on the vignette design using the measure of heteroskedasticity in each respondent-vignette pair. The standard HOPIT model, which assumes variance in vignette perception is the same across different individuals. Vignette which have the smaller variance estimates are deemed to contain more information and therefore selected in subsequent surveys. Yet, application of this criterion may result in dismissing vignettes which are only precisely perceived by a particular group of respondents. Based on the estimates of our extended HOPIT model, we are able to evaluate the information content of a given vignette for every subgroup of respondents and select vignettes accordingly. For example, if we are particularly interested in the health condition of a SES group, we could calculate the standard error of vignette perception of each vignette regarding this group and select vignettes or design new vignettes with the least perception variance.

Supporting information

S1 Appendix. Survey questions on domains of pain, cognition and mobility

S2 Appendix. Estimates of the standard and extended HOPIT models

S3 Appendix. Algorithm used for important sampling of the likelihood function

S4 Appendix. Calculation of the heteroskedastic error

References

1. King G, Murray CJL, Salomon JA, Tandon A. Enhancing the Validity and Cross-cultural Comparability of Measurement in Survey Research. American Political Science Review. 2004;98(1):191–207.

2. Kapteyn A, Smith JP, van Soest A. Vignettes and Self-Reports of Work Disability in the United States and the Netherlands. American Economic Review. 2007;97(1):461–473.

3. Jürges H. True health vs response styles: exploring cross-country differences in self-reported health. Health Economics. 2007;16(2):163–178. doi:10.1002/hec.1134.

4. Hays RD, Morales LS, Reise SP. Item Response Theory and Health Outcomes Measurement in the 21st Century. Medical care. 2000;38(9 Suppl):II28–II42.

5. Shmueli A. Reporting Heterogeneity in the Measurement of Health and Health-Related Quality of Life. PharmacoEconomics. 2002;20(6):405–412.

6. Shmueli A. Socio-economic and demographic variation in health and in its measures: the issue of reporting heterogeneity. Social Science & Medicine. 2003;57(1):125 – 134. doi:http://dx.doi.org/10.1016/S0277-9536(02)00333-7.

7. Lindeboom M, van Doorslaer E. Cut-point shift and index shift in self-reported health. Journal of Health Economics. 2004;23(6):1083 – 1099.

8. Salomon JA, Tandon A, Murray CJL. Comparability of self rated health: cross sectional multi-country survey using anchoring vignettes. BMJ. 2004;.

9. Bago d'Uva T, Van Doorslaer E, Lindeboom M, O'Donnell O. Does reporting heterogeneity bias the measurement of health disparities? Health Economics. 2008;17(3):351–375.

10. Bago d'Uva T, O'Donnell O, van Doorslaer E. Differential health reporting by education level and its impact on the measurement of health inequalities among older Europeans. International Journal of Epidemiology. 2008;37:1375–1383.

11. Mu R. Regional disparities in self-reported health: evidence from chinese older adults. Health Economics. 2014;23(5):529–549.

12. Hanandita W, Tampubolon G. Does reporting behaviour bias the measurement of social inequalities in self-rated health in Indonesia? An anchoring vignette analysis. Quality of Life Research. 2016;25(5):1137–1149.

13. Molina T. Reporting Heterogeneity and Health Disparities Across Gender and Education Levels: Evidence From Four Countries. Demography. 2016;53(2):295–323. doi:10.1007/s13524-016-0456-z.

14. Xu H, Xie Y. Socioeconomic Inequalities in Health in China: A Reassessment with Data from the 2010–2012 China Family Panel Studies. Social Indicators Research. 2016; p. 1–21. doi:10.1007/s11205-016-1244-2.

15. Rice N, Robone S, Smith PC. Vignettes and health systems responsiveness in cross-country comparative analyses. Journal of the Royal Statistical Society: Series A (Statistics in Society). 2012;175(2):337–369.

16. Malhotra C, Do Y. Socio-economic disparities in health system responsiveness in India. Health Policy Plan. 2013;28(2):197–205.

17. King G, Wand J. Comparing Incomparable Survey Responses: Evaluating and Selecting Anchoring Vignettes. Political Analysis. 2007;15:46–66.

18. Hopkins D, King G. Improving Anchoring Vignettes: Designing Surveys to Correct Interpersonal Incomparability. Public Opinion Quarterly. 2010; p. 1–22.

19. Kapteyn A, Smith JP, van Soest A. Are Americans Really Less Happy With Their Incomes? RAND Corporation; 2011.

20. Bonsang E, van Soest A. Satisfaction with Job and Income Among Older Individuals Across European Countries. Social Indicators Research. 2012;105(2):227–254. doi:10.1007/s11205-011-9879-5.

21. Angelini V, Cavapozzi D, Corazzini L, Paccagnella O. Do Danes and Italians rate life satisfaction in the same way? using vignettes to correct for individual-specific scale biases. Oxford Bulletin of Economics and Statistics. 2014;76(5):643–666. doi:10.1111/obes.12039.

22. Ravallion M, Himelein K, Beegle K. Can Subjective Questions on Economic Welfare Be Trusted? Economic Development and Cultural Change. 2016;64(4):697–726. doi:10.1086/686793.

23. Kristensen N, Johansson E. New evidence on cross-country differences in job satisfaction using anchoring vignettes. Labour Economics. 2008;15(1):96 – 117.

24. Wang J. Rural-to-urban Migration and Rising Evaluation Standards for Subjective Social Status in Contemporary China. Social Indicators Research. 2016; p. 1–22. doi:10.1007/s11205-016-1457-4.

25. Crane M, Rissel C, Greaves S, Gebel K. Correcting bias in self-rated quality of life: an application of anchoring vignettes and ordinal regression models to better understand QoL differences across commuting modes. Quality of Life Research. 2016;25(2):257–266. doi:10.1007/s11136-015-1090-8.

26. van Soest A, Delaney L, Harmon C, Kapteyn A, Smith JP. Validating the use of anchoring vignettes for the correction of response scale differences in subjective questions. Journal of the Royal Statistical Society: Series A (Statistics in Society). 2011;174(3):575–595.

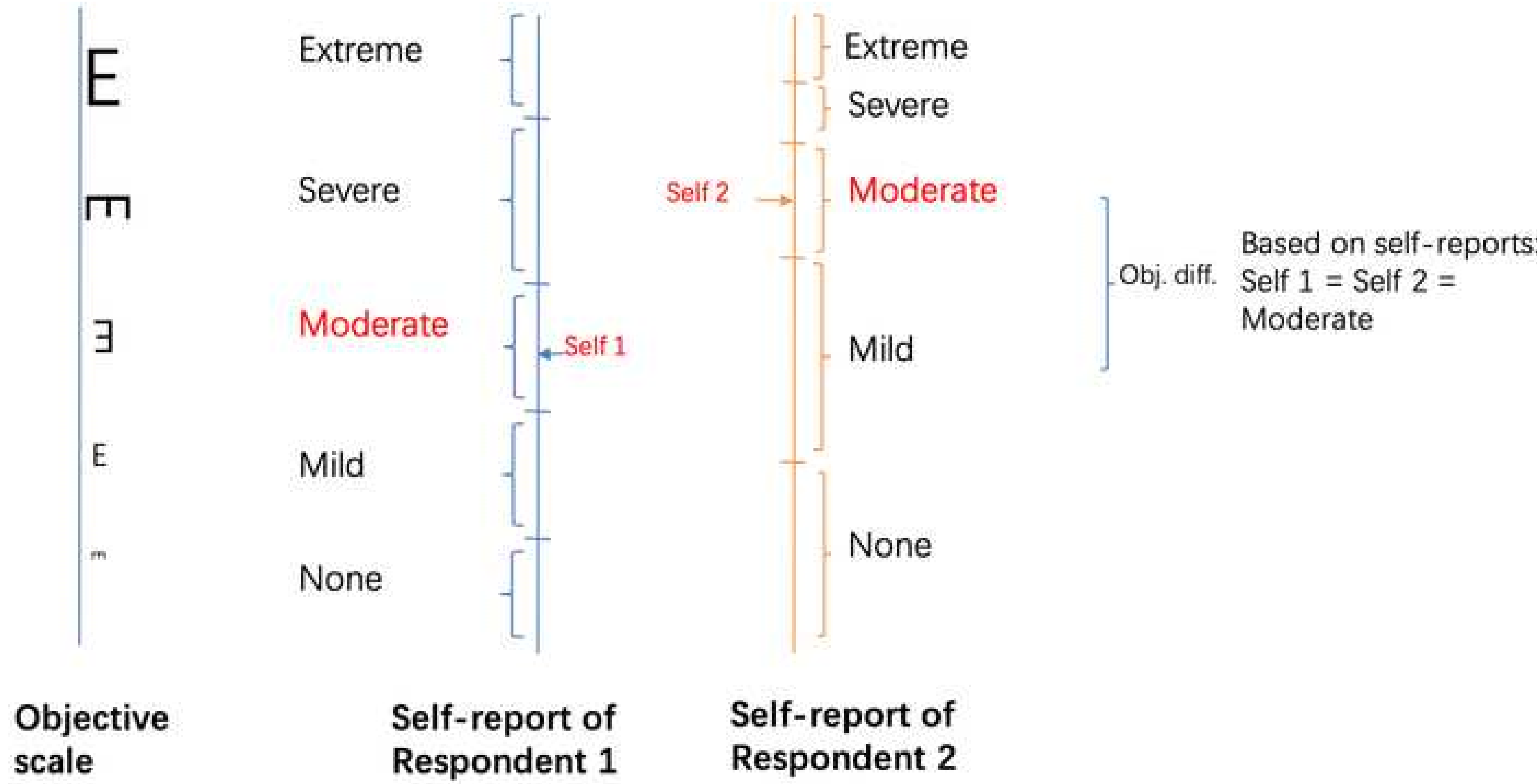
27. Au N, Lorgelly PK. Anchoring vignettes for health comparisons: an analysis of response consistency. Quality of Life Research. 2014;23(6):1721–1731. doi:10.1007/s11136-013-0615-2.

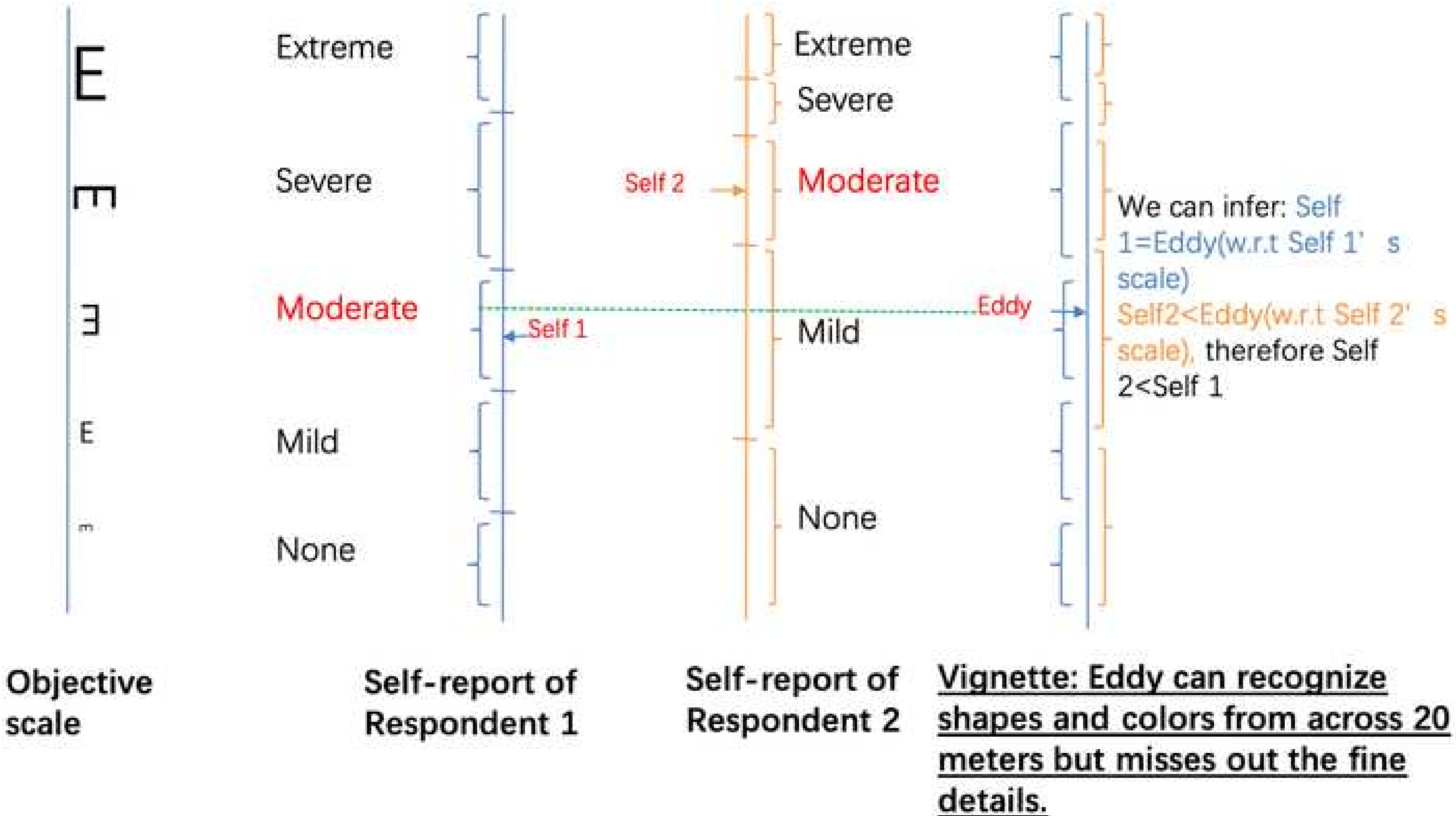
28. Liberman N, Trope Y. The Psychology of Transcending the Here and Now. Science. 2008;322(5905):1201–1205. doi:10.1126/science.1161958.

29. Trope Y, Liberman N. Construal-Level Theory of Psychological Distance. Psychological review. 2010;117(2):440–463.

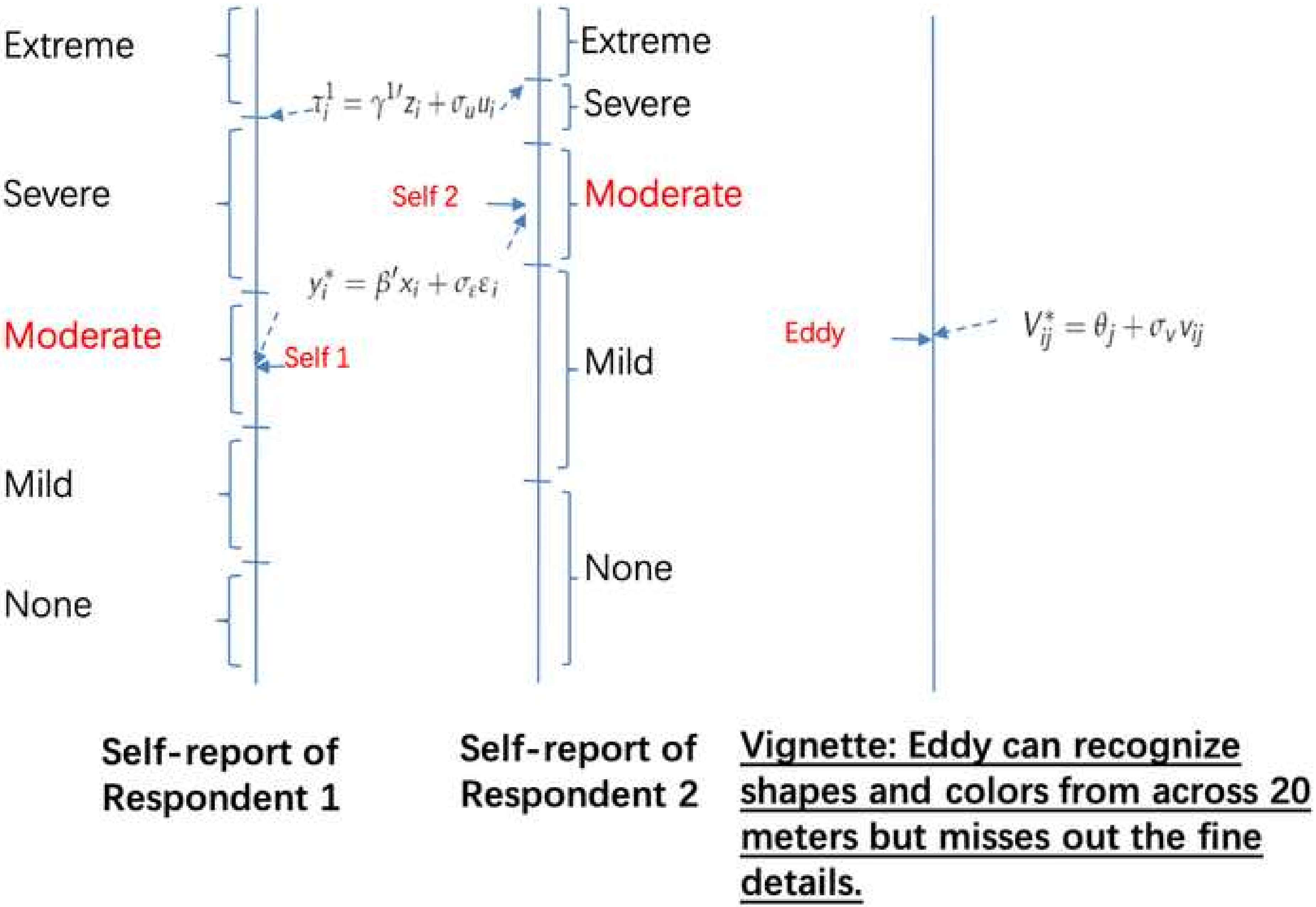
30. Voňková H, Hullegie P. Is the anchoring vignette method sensitive to the domain and choice of the vignette? Journal of the Royal Statistical Society: Series A (Statistics in Society). 2011;174(3):597–620.

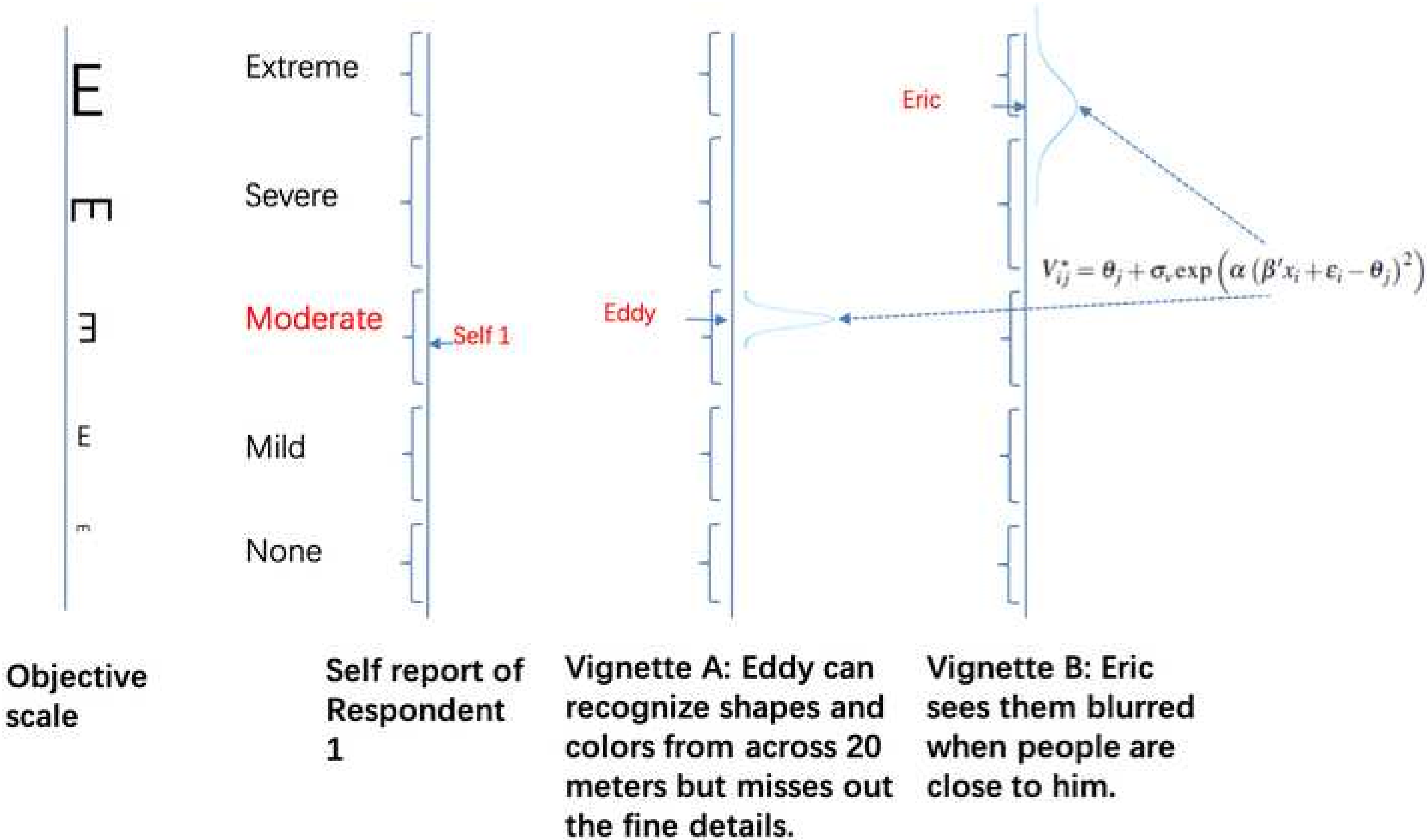
31. Lermer E, Streicher B, Sachs R, Raue M, Frey D. Thinking Concretely Increases the Perceived Likelihood of Risks: The Effect of Construal Level on Risk Estimation. Risk Analysis. 2016;36(3):623–637. doi:10.1111/risa.12445.


Extreme
Severe
Moderate
Mild
None
Self 1
Self 2
Extreme
Severe
Moderate
Mild
None
Obj. diff.
Based on self-reports:
Self 1 = Self 2 =
Moderate
Objective scale
Self-report of Respondent 1
Self-report of Respondent 2

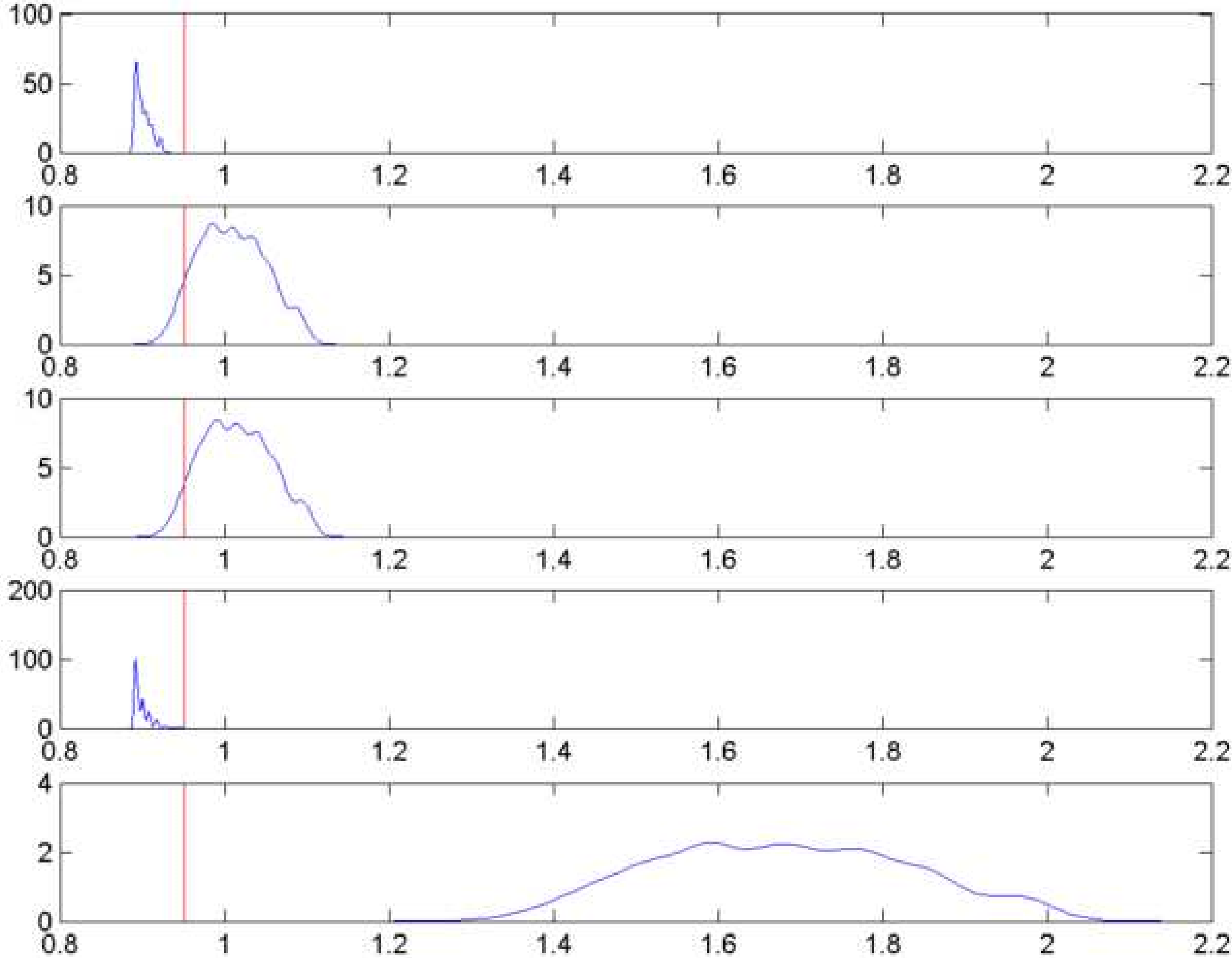


Extreme
Severe
Moderate
Mild
None
Self 1
Self 2
Extreme
Severe
Moderate
Mild
None
Eddy
We can infer: Self 1=Eddy(w.r.t Self 1' s scale)
Self2<Eddy(w.r.t Self 2' s scale), therefore Self 2<Self 1
Objective scale
Self-report of Respondent 1
Self-report of Respondent 2
Vignette: Eddy can recognize shapes and colors from across 20 meters but misses out the fine details.


Extreme
Severe
Moderate
Mild
None
$\tau_i^1 = \gamma^{1\prime} z_i + \sigma_u u_i$
Extreme
Severe
Moderate
Mild
None
Self 2
$y_i^* = \beta' x_i + \sigma_\varepsilon \varepsilon_i$
Self 1
Eddy
$V_{ij}^* = \theta_j + \sigma_v v_{ij}$
Self-report of Respondent 1
Self-report of Respondent 2
Vignette: Eddy can recognize shapes and colors from across 20 meters but misses out the fine details.


Extreme
Severe
Moderate
Mild
None
Self 1
Eddy
Eric
$V_{ij}^{*} = \theta_j + \sigma_v \exp\left(\alpha\left(\beta' x_i + \varepsilon_i - \theta_j\right)^2\right)$
Objective scale
Self report of Respondent 1
Vignette A: Eddy can recognize shapes and colors from across 20 meters but misses out the fine details.
Vignette B: Eric sees them blurred when people are close to him.

Click here to download Figure Fig5.tiff



Click here to access/download
Supporting Information
S1_appendix.pdf

Click here to access/download
Supporting Information
S2_appendix.pdf

Click here to access/download
Supporting Information
S3_appendix.pdf