

DOI: [10.14515/monitoring.2022.6.2290](https://doi.org/10.14515/monitoring.2022.6.2290)



I. F. Deviatko, A. A. Byzov

MEASURING RESPONDENTS' COGNITIVE LOAD RELATED TO MAKING FACTUAL AND NORMATIVE JUDGMENTS

For citation:

Deviatko I. F., Byzov A. A. (2022) Measuring Respondents' Cognitive Load Related to Making Factual and Normative Judgments. *Monitoring of Public Opinion: Economic and Social Changes*. No. 6. P. 291–308. <https://doi.org/10.14515/monitoring.2022.6.2290>.

Правильная ссылка на статью:

Девятко И.Ф., Бызов А.А. Измерение когнитивной нагрузки респондентов, связанной с вынесением фактических и нормативных суждений // Мониторинг общественного мнения: экономические и социальные перемены. 2022. № 6. С. 291—308. <https://doi.org/10.14515/monitoring.2022.6.2290>. (In Eng.)

Получено: 24.07.2022. Принято к публикации: 07.11.2022.

MEASURING RESPONDENTS' COGNITIVE LOAD RELATED TO MAKING FACTUAL AND NORMATIVE JUDGMENTS

Inna F. DEVIATKO^{1,2} — Dr. Sci. (Soc.), Full Professor; Chief Researcher
E-MAIL: deviatko@gmail.com
<https://orcid.org/0000-0002-1955-7592>

Alexandr A. BYZOV³ — Independent Researcher
E-MAIL: debesergopotes12@gmail.com
<https://orcid.org/0000-0002-6253-8581>

¹ HSE University, Moscow, Russia

² Institute of Sociology of the Federal Center of Theoretical and Applied Sociology of the Russian Academy of Sciences, Moscow, Russia

³ Moscow, Russia

Abstract. The use of various methods for measuring cognitive load and mental effort in recent years has become increasingly popular in various fields of social and affective neuroscience, in applied research on the comparative effectiveness of teaching methods and training platforms, in the study of the distribution of attention in solving various problems and using informational tips in decision making, etc. In this broader context, the specific request for a multimodal assessment of the cognitive load of interviewers and respondents and of its impact on the quality of the survey data, including the use of paradata and webcams for this purpose, has been also growing recently. We conducted a within-subject methodological experiment ($N=50$) aiming at comparative measurement of task-evoked

ИЗМЕРЕНИЕ КОГНИТИВНОЙ НАГРУЗКИ РЕСПОНДЕНТОВ, СВЯЗАННОЙ С ВЫНЕСЕНИЕМ ФАКТИЧЕСКИХ И НОРМАТИВНЫХ СУЖДЕНИЙ

ДЕВЯТКО Инна Феликсовна — доктор социологических наук, профессор, Национальный исследовательский университет «Высшая школа экономики», Москва, Россия; главный научный сотрудник, Институт социологии ФИНСЦ РАН, Москва, Россия
E-MAIL: deviatko@gmail.com
<https://orcid.org/0000-0002-1955-7592>

БЫЗОВ Александр Александрович — независимый исследователь, Москва, Россия
E-MAIL: debesergopotes12@gmail.com
<https://orcid.org/0000-0002-6253-8581>

Аннотация. Использование различных методов измерения когнитивной нагрузки и умственных усилий приобретает все большую популярность в различных областях социальной и аффективной нейронауки, в прикладных исследованиях сравнительной эффективности методов обучения и обучающих платформ, в исследованиях распределения внимания при решении различных задач или использования информационных подсказок при принятии решений и т.д. В этом более широком контексте специфический запрос на мультимодальную оценку когнитивной нагрузки интервьюеров и респондентов и оценку ее влияния на качество опросных данных, в том числе с использованием для этой цели параданных и веб-камер, в последнее

cognitive load of respondents related to two tasks of making factual and normative judgments. The first task implied making causal and blame judgments for two institutional domains (medical, work dress-code) using the similar factorial vignettes, while the second task presupposed making lay factual and normative-deontic judgments concerning migrant rights to free health care. We used in parallel two measures of task-evoked cognitive load — pupillometry (Pupil Lab glasses), and the Paas scale of mental effort. The results provide limited evidence in support of the difference that exists between ordinary judgments of cause, blame, and severity of harm in terms of their propensity to evoke psychosensory pupillary response and subjectively perceived mental effort, both reflecting the variability in the cognitive load imposed on survey respondent when performing a pertinent survey task. We also briefly discuss the evidence obtained in support of the task-specific difference in sensitivity and validity of neurophysiological and self-report-based measures of survey-related cognitive load.

время также растет. Мы провели внутрисубъектный методический эксперимент ($N=50$), целью которого было сравнительное измерение вызванной заданием когнитивной нагрузки респондентов, связанной с двумя задачами на вынесение фактических и нормативных суждений. Первая задача подразумевала вынесение суждений о причинности, вине и тяжести вреда для двух институциональных областей (медицина, работа) с использованием аналогичных факторных виньеток, тогда как вторая задача предполагала вынесение обыденных фактических и нормативно-деонтических суждений о правах мигрантов на бесплатное медицинское обслуживание. Мы использовали параллельно две меры когнитивной нагрузки, связанной с заданием, — pupillometriю (с использованием очков-айтрекера Pupil Lab) и шкалу оценки умственного усилия Ф. Пааса. Представленные в статье результаты предоставляют ограниченные доказательства в поддержку различия, которое существует между обыденными суждениями о причине, вине и тяжести вреда, с точки зрения их способности вызывать психосенсорную зрачковую реакцию и субъективно воспринимаемое умственное усилие, что отражает изменчивость когнитивной нагрузки респондента при выполнении соответствующего опросного задания. Мы также кратко обсуждаем доказательства, полученные в поддержку существования различий в чувствительности и достоверности нейрофизиологических и основанных на субъективных самоотчетах показателей когнитивной нагрузки респондента для специфических заданий.

Keywords: mental effort, cognitive load, factual judgements, normative judgements, vignette experiment

Ключевые слова: умственное усилие, когнитивная нагрузка, фактические суждения, нормативные суждения, факторный эксперимент с виньетками

Acknowledgments. The research was supported by RSF (project number 22-28-00968, project title: "Eye-tracking data and pupillometry in multimodal measurement of the respondents' cognitive load").

Благодарность. Данное исследование осуществлено при финансовой поддержке Российского научного фонда (проект № 22-28-00968, тема «Окулография в мультимодальном измерении когнитивной нагрузки респондента»).

Disclosure statement. No potential conflict of interest was reported by the authors.

Заявление о конфликте интересов. Авторы не сообщили о потенциальном конфликте интересов.

Introduction

Recently, the use of various methods for measuring cognitive load and mental effort has become increasingly popular in various fields of social and affective neuroscience, particularly, in applied research on the comparative effectiveness of teaching methods and training platforms, in studies of the distribution of attention and using informational tips in decision making, etc. [Chen et al., 2016; Hoogerheide et al., 2019; Jbara, Feitelson, 2017; Schmeck et al., 2015]. In the related field of sociological methodology, the specific request for a multimodal assessment of respondents' cognitive load and perceived mental effort and their impact on the quality of the survey data has also been growing over the recent years [Deviatko, Lebedev, 2017; Höhne, Schlosser, Krebs, 2017; Höhne, Lenzner, 2018; Kaminska, Foulsham, 2014; Neuert, 2021; Stodel, 2015]. At the same time, the possibilities of relatively new approaches to measuring survey-related cognitive load using unobtrusive and non-invasive neurophysiological methods such as modern portable and wearable devices for eye tracking and pupillometry remain rather underestimated, despite the fact that these devices proved to be instrumental in conducting the accurate comparisons of the oculographic indicators of cognitive effort related to processing the specific question formats and response categories [Höhne, 2019], the different survey modes [Deviatko, Bogdanov, Lebedev, 2021], as well as in identifying problematic survey questions leading to excessive respondents' burden [Neuert, 2020]. The latter strain of research demonstrated, in particular, that the long-debated possible advantage of the item-specific question format over the agree/disagree (A/D) one in susceptibility to response bias is counterbalanced by deeper cognitive processing as measured by markedly longer fixations on response categories for A/D format [Höhne, 2019], while fixation times seemingly turned out to be more sensitive in revealing the problematic, poorly worded questions when compared to pupil data [Neuert, 2020]. However, the possible differences in task-related cognitive load associated with making either normative or factual judgments made by survey respondents, which are the focus of this article, still remains relatively unexplored with both more traditional and relatively newer methods.

Nowadays, the growing number of factorial survey experiments in social sciences deploy the vignettes constructed by systematically varying informational cues used as experimental factors purportedly influencing respondents' attitudes and opinions in order to statistically estimate main and interaction effects of these factors upon respondents' normative and factual judgments, which include the causal judgments, about scenarios described in the vignettes [Deviatko, 2007; Lavrakas et al., 2019; Rossi, Anderson, 1982; Sniderman, Grob, 1996]. Response formats used for these judgments are usually based on relevant ordinal rating scales (e.g., ratings of causal impact, blame, moral worth, etc.). Exploring in parallel the task-related cognitive load for both normative and factual judgments about identical vignettes using unobtrusive wearable devices for eye tracking and pupillometry can be of great benefit for at least two intertwined tasks—the better understanding of response models underlying the cognitive processes which are involved in making these different types of lay judgments and the estimation of validity and reliability of the pertinent types of survey data.

It is worth mentioning that the advent of wearable oculographic devices coincided in time with the formation of a deeper understanding of neurophysiological processes associated with pupil size dynamics and their reciprocal relationship with the processes of distribution of cognitive load, attention, decision-making, etc. [Mathôt, 2018]. Small changes in pupil size (< 1 mm) driven by the noradrenergic system associated with neuronal activity in the locus coeruleus [Costa, Rudebeck, 2016] reflect the dynamics of cognitive load with high temporal resolution and are considered to be inaccessible for voluntary control¹. This dynamics is associated with the use of attention and working memory resources and acts as a kind of “window” in the underpinnings of information processing and decision-making, allowing to grasp the cognitive load associated with specific tasks [Laeng, Sirois, Gredebäck, 2012].

As briefly stated above, the important yet understudied aspect of cognitive load measurement relates to the possibility of elaboration and verification of empirically-based models of making lay factual, explanatory, and normative judgments. To name just a few, these judgments may be based on evaluating the prevalence of specific behaviors (for herself/himself or for others), making predictions of everyday social facts, estimating distributive or procedural justice, attributing responsibility and blame, etc. These judgments are currently considered not so much as based on some form of “simple theory of the survey response” which describes response choices during the survey as mostly reflecting currently accessible ideas retrieved from the memory [Zaller, Feldman, 1992], but rather as non-trivial outcomes predicted by a more complicated dual-system information-processing framework, involving either fast, intuitive decisions based on “associative machine” of system-1 [Morewedge, Kahneman, 2010], or mostly the reflective/conscious consequential decision making based on capacity-limited system-2, or, at last, the dual-process production of task-specific judgments [Evans, Stanovich, 2013; Guglielmo, 2015]. Some empirical findings support dual-processing models, in particular, the revealed pattern of differences in RTs² depending on increase

¹ Despite the existing data on the possible influences from high-level cognitive processes on the psychosensory and other reactions of the pupil which requires further clarification.

² RT — reaction time. Reaction time data are used, in particular, as cognitive load indicator, with increases in RT reflecting increases in cognitive load.

in extraneous cognitive load (due to imposing an additional control-demanding task) for conscious reasoning-based utilitarian judgments in high-conflict, difficult moral dilemmas as compared to more emotion-based non-utilitarian moral judgments which are not sensitive to such increase [Greene et al., 2008].

Another line of research demonstrates that judgment timing can be indicative in adjudicating between various models of moral judgments [Guglielmo, 2015]. Many models presume, particularly, that certain judgments (causality, mental states) usually precede judgments of blame or responsibility [Cushman, 2008; Malle, Guglielmo, Monroe, 2014]. Blame and responsibility judgments are considered in these models as requiring orderly incoming information units (agent causal role, intentionality, etc.) obtained as an output from the previous steps of cognitive processing, e.g., Path Model of Blame by Malle et al. [2014]. However, other models of moral judgment (e.g., [Alicke, 2000; Knobe, 2010; Schein, Gray, 2014]) postulate, contrariwise, that blame attribution is mostly based on fast, intuitive and emotional normative-evaluative judgments preceding the attribution of mental states and causality.

Recent empirical findings give some support to the role of time and the canonical order of information unit processing posited by Path Model of Blame [Guglielmo, Malle, 2017]. To sum up, the deliberative, conscious information processing models of this type describe the stepwise “rationalist” processing of constituent factual information units concerning probable causality chains, agent’s intents, possibilities of control over action consequences, etc., considered as preconditions for resulting normative-evaluative judgment. Such reflective processing models explicitly or implicitly presuppose that total cognitive load and mental effort related to the task of blame judgment should surpass in this case the cognitive load predicted by the competing models of fast and spontaneous intuitive evaluation of blame (sometimes called biased information models). However, it still remains unclear whether more directly obtained data on cognitive load and perceived mental effort, related to tasks of attributing blame or making other normative judgments, could be helpful in adjudicating between these two classes of information models that explain how respondents answer these different types of questions while participating in factorial surveys and opinion polls. The present study might be a first step in elucidating this question.

In sum, the need for this research arises from two reasons. From theoretical perspective, this research could greatly improve our understanding of information processing underlying everyday moral judgments about typical social situations, especially judgments on blame. Are normative judgments, by and large, fast and intuitive or do they rather lean on deliberative, conscious processes, and, correspondingly, might depend in the latter case on the consecutive input of information elements leading, in turn, to differential cognitive load for factorial survey respondents? From methodological perspective, this research shows relative promise and pitfalls of using neurophysiological and self-report-based methods of measurement of respondent burden evoked by a specific task in judgment formation research and sociology of morality with a view to improve the quality of data collected. This research could also inform future scholars in their decision to invest into wearable oculographic devices, alongside with using more traditional methods of mental effort measurement, for cognitive pretesting of factorial survey instruments and, more generally, for using these devices in judgment and decision studies.

The current research

This study aims at the comparative measurement of task-evoked cognitive load of respondents that is related to making factual and normative judgments concerning life-based scenarios described in factorial vignettes: (1) causal and blame judgments for two institutional domains (medical, corporate dress-code); (2) factual and normative-deontic judgments about migrants' right to free health care. Causal factual judgments assume, in a case of first task, evaluating the causal role that a vignette protagonist played in inadvertently provoking an aversive outcome for another person, while normative-evaluative blame judgments relate to rating the protagonist's blame for the same vignette (respondents are also asked to rate the severity of harm inflicted). Factual descriptive judgment for the second task presupposes answering a question concerning an actual right to free medical care for individual migrants, while normative-prescriptive judgments relate to ought-questions, i. e., whether individual migrants should have this specific right (see the next section for more details).

Basically, we want to see if pupillometry as a neurophysiological method of cognitive load measurement, alongside with more traditional measure of subjectively perceived mental effort, could detect theoretically expected differences between these types of moral and factual judgments, thereby demonstrating its construct validity. In order to substantiate our expectations concerning task-specific cognitive load for different types of tasks, we briefly summarize some predictions from the existing models of moral judgments described above.

Two major types of information models described above ground our expectations for causal and blame judgments: "rationalist" and biased information models (see [Guglielmo, 2015]). The first type of models describes the consequential analysis of such features of an agent's behavior as causality, intentionality, and harmful consequences (e. g., [Cushman, 2008; Malle, Guglielmo, Monroe, 2014]). It predicts that:

H1.1: Cognitive load evoked by the causality judgment related to a specific vignette will be equal or even smaller compared to cognitive load evoked by the blame judgment conditional upon the causality judgment, and both will produce larger cognitive loads when compared to one related to severity of harm judgment as a precondition for blame attribution.

Biased information models—e.g., culpable control model by Alicke [2000] or, in a way, the theory of dyadic morality [Schein, Gray, 2018] specifies the principal and direct contribution from harm-based spontaneous affective moral evaluations upon blame attribution and, indirectly, upon causality attribution. It predicts:

H1.2: The absence of the marked differences between cognitive load related to causality, blame and harm judgments for identical scenarios.

Another hypothesis arises from previous research on the effect of institutional domains on causal, blame, and harm judgments [Deviatko, Gavrilov, 2020]. This research demonstrated a significant difference between two institutional action domains: actors in "medical"-related vignettes were generally estimated to be more causally effective and blameworthy than actors in "dress code"—related vignettes. Based on this research, we predict that:

H2: Cognitive load evoked in making causality and blame judgments about negative side effects of intentional actions related to medical institutional domain differs from one evoked while making judgments related to work domain.

As to factual and normative-deontic judgments on the different types of migrant rights, we did not have comparable, albeit preliminary, predictions from information models. We were aware only about processing models [Guglielmo, 2015] describing this type of lay deontic judgments from informational cues influencing the resulting judgments (cultural distance, skill level, etc.). Thus, in this experimental block, we followed an exploratory approach comparing respondents' cognitive load evoked by factual and deontic judgments.

Method

This research focuses on the investigation of cognitive load during the making of ordinary factual and normative (evaluative and deontic) judgments. We used two pairs of tasks. The first pair of tasks is related to cause and blame judgments for two institutional domains (medical, work dress-code). The second pair of tasks relates to judgments on whether an individual immigrant from one of two countries (Belarus, Uzbekistan) has a particular right and should have the right. Both of these tasks were administered to all participants of the experiment.

Materials and procedure

Each experiment was conducted under similar conditions: the same dimly lit room, draped windows, closed door. After a participant had arrived and took a seat at a table opposite to the window, an experimenter explained the procedure, and put on the participant a wearable eye-tracker (glasses).

Tasks

For both tasks (Task 1 and Task 2) respondents were asked to answer two or three related questions for two vignettes used as stimuli (see below for details). Each question was placed on a separate page. After each question, we asked a participant to rate the task on the original version of 9-point Paas scale (ranging from *very, very low mental effort* (1) to *very, very high mental effort* (9)), which is a self-report rating scale on the amount of mental effort spent on the task [Paas, 1992] and then, to count to five to give a participant's pupil time to readjust.

We employed two counterbalanced versions of a questionnaire with direct and reversed order of questions in order to alleviate possible carry-over effect. Each participant received all tasks and all vignettes (within-subject factorial design). Two vignettes in each task had an equal number of words (in Russian) to account for potential variability in cognitive load due to reading-related cognitive load. Among our participants, 19 were randomly assigned to a self-completed paper-based mode of questionnaire, while 20 were assigned to its computerized version. We found no statistical differences between the two modes of administration, so the data were analyzed together.

Vignettes for Task 1 were previously used as a part of a bigger set for substantive cause and blame attribution tasks in another study [Deviatko, Gavrilov, 2020] and had the identical levels of all experimental factors used in this previous study (principal action originator — individual, group, or institution; type of damage — monetary damage, or damage to health; the “remoteness” of a victim), except the factor “the institutional domain of action” which also had two levels in the current study (see table 1). One

institutional domain described adverse situations that emerged during the purchase of medication (the “medical” domain), while the other described situations connected to the negative side effects of implementing a dress code in an organizational setting (the “dress code” domain). The respondents’ answers evaluating causality, blame and severity of harm were given on corresponding 11-point rating scales with increment 10, ranging from 0 (e.g., *not at all the cause*) to 100 (e.g., *completely the cause*).

Task 1. There were two vignettes for this task:

Scenario 1: The Minister of Health had issued a decree to expand the list of prescription drugs. Drug X turned out to be on the list. Mikhail Borisovich, a senior citizen, needs to take this drug regularly. When the medicine ran out, his wife, Anna Nikolaevna, could not purchase this drug at the nearest pharmacy without a prescription. She bought a substitute drug, which cost ten times more.

Scenario 2: The CEO of the corporation has established strict dress code rules for all employees. Elena went to the office wearing a tight skirt and high heels following these rules. She tripped over a small metal threshold in the corporate dining room and dropped her food-laden tray. As a result, Elena’s costly costume was hopelessly flawed.

Table 1. Factors and levels for Task 1

Factor	Levels
Institutional domain	1. Medical 2. Work dress-code
Type of judgment	1. Cause 2. Blame 3. Harm severity

For this pair of scenarios respondents made three judgments: cause, blame, and harm.

Judgments on cause: (1) Is the Minister of Health’s decree the cause that Anna Nikolaevna had to buy a substitute drug that cost ten times more than the previous drug X? or (2) Is the decision of CEO of the corporation the cause why Anna’s expensive costume was hopelessly damaged?

Judgments on blame: (1) Is the Minister of Health to blame for Anna Nikolaevna having to buy a substitute drug that cost almost ten times more than the previous drug X? or (2) Is the CEO of the corporation to blame for the fact that Anna’s expensive costume was hopelessly damaged?

Judgment on harm severity: (1) How severe are the consequences for Anna Nikolaevna? or (2) How severe are the consequences for Elena?

Task 2. Vignettes for this task were previously used in another study³ as a part of a bigger set for exploring determinants of factual (descriptive) and deontic (normative-prescriptive) judgments concerning migrant rights (see table 2).

There were two almost identical vignettes for this task varying only the migrant’s country of origin (and, correspondingly, name): *[Artyom/Azizbek] is a middle-aged immigrant from [Belarus/Uzbekistan]. He moved to Russia two years ago. He regis-*

³ Byzov, Devyatko, in preparation.

tered in the Migration Agency and has lived here legally. He's been working in Russia as a senior software developer since his arrival. [Artyom/Azizbek] is fluent in Russian.

Table 2. **Factors and levels for Task 2**

Factor	Levels
Country of origin	1. Uzbekistan (Azizbek) Belarus (Artyom)
Type of judgment on the individual immigrant's right	1. Has a right to free medical care (descriptive) 2. Should have a right to free medical care (prescriptive)

The participants were tasked to evaluate two statements: (1) Does [Artyom/Azizbek] have the same rights as citizens of the Russian Federation to free medical care? (2) Should [Artyom/Azizbek] have the same rights as the Russian Federation's citizens to free medical care? These statements were evaluated on a similar scale from 0 (absolutely disagree) to 100 (absolutely agree) with increment 10.

The measurement of cognitive load

We measured the diameter of a participant's pupil with Pupil Labs Pupil Core eye-tracker. This eye-tracker records both the participant's pupil (eye camera with sampling frequency 200 Hz) and gaze (world camera). We extract a pupil diameter in mm per frame with Pupil Labs' Pupil Player offline pupil detection algorithm. This data was subjected to several preparation procedures. First, we removed data with low or medium confidence values assigned by the pupil detection algorithm (< 0.7). Second, we omitted observations with abnormal pupil diameters values (less than 1 mm or more than 9 mm). Finally, we standardized pupil data by subtracting from each datum a baseline value, i. e., using subtractive baseline correction [Mathôt et al., 2018]. A baseline value was computed by averaging high confidence pupil diameter data from the experiment's first two minutes.

To analyze cognitive load per judgment we separated pupil diameter data on epochs. The epoch is a particular period from the start of the new page to the time when a participant provides a written response. The epochs were manually coded from world camera recordings. Preprocessed data from each of these epochs were averaged to receive mean pupil diameter for an epoch. It is important to note that we included *only those epochs*, in which 50 % or more observations had good quality (high confidence, typical pupil diameters).

Also, we used a Paas scale in its initial format [Paas, 1992] in order to access mental effort as the subjective component of cognitive load. Mental effort may be defined as the total amount of controlled cognitive processing in which a subject is engaged [Paas, Van Merriënboer, 1993].

Participants

The sample consisted of 50 students from one of the Russia's universities. Participation was voluntary and did not presuppose any remuneration. However due to several technical problems with hardware and software only 39 individual-level observations were analyzed (see table 3). The problems were as follows:

- some respondents had eye structure that does not allow for a reliable measurement of their pupil size in an “ecologically valid” situation,
- some respondents wore mascara or artificial eyelashes that partly covered a camera of an eye-tracker,
- sometimes, a current version of proprietary software used had a bug that stopped the process of recording observations,
- in few cases, a hard disk became full by the end of the recording and did not allow to save a video file.

At last, we used a mobile eye-tracker for one eye only. These types of eye-trackers are notorious for their sensitivity to some head movements, which happen when a person sits for a prolonged time without chin fixation, so some noise and excluded cases in eye pupil size measurement are usually expected.

Table 3. Descriptive statistics of study participants

Characteristic	N=39*
Gender	
Man	3 (7.7 %)
Woman	36 (92 %)
Age	18 (18.21)
Unknown	1

* Statistics presented: n (%);
median (1st quartile, 3rd quartile)

The procedure of data analysis

The main procedure for data analysis is parametric Repeated Measures ANOVA (contrasts set to sum, type III of estimating sum of squares, Greenhouse-Geisser correction). Only results with $p < .05$ are reported and further analyzed with post-hoc tests, which were separately adjusted with Tukey’s method for correcting p value. The data analysis was conducted in R.

Results

Task 1

Pupillometry. We excluded data from 21 participants for this analysis because their pupil diameter contained less than 50 % of quality data. No significant main effects of institutional domain or type of judgments were observed. There was one significant interaction between institutional domain and judgments on the cause, blame, and harm on pupil diameter (see table 4).

Table 4. ANOVA results for Task 1 with pupil's diameter as dependent variable

Effect	Df	MSE	F	pes	p value
Institutional domain	1, 19	0.01	0.02	.001	.886
Type of judgment	1.11, 21.04	0.00	1.63	.079	.217
Institutional domain × Type of judgment	1.10, 20.83	0.00	4.40*	.188	.045

The post-hoc analysis shows that there is a significant difference between causal and severity of harm judgments in a medical institution domain (see table 5).

Table 5. Post-hoc tests for Task 1 with pupil's diameter as dependent variable

Contrast	EMM diff.	SE	DF	Statistic	Adjusted p value
Medical Cause—Medical Harm	-0.03	0.01	73.49	-3.27	0.02

Paas scale. We excluded data from the same 21 participants for this analysis to achieve comparability between these two methods of cognitive load estimation. There was one significant main effect of institutional domain on Paas scale (see table 6).

Table 6. ANOVA results for Task 1 with Paas scale as dependent variable

Effect	df	MSE	F	pes	p value
Institutional domain	1, 18	1.59	8.82**	.329	.008
Type of judgment	1.78, 32.03	0.67	0.55	.030	.564
Institutional domain × Type of judgment	1.43, 25.80	1.03	1.94	.097	.172

The post-hoc analysis of the significant main effect of institutional domain demonstrates significant difference between judgments related to medical and work dress-code related domains (see table 7).

Table 7. Post-hoc tests for Task 1 with Paas scale as dependent variable

Contrast	EMM diff.	SE	DF	Statistic	p value
Work dress-code—Medical	-0.70	0.24	18.00	-2.97	0.01

Task 2

We excluded data from 17 participants for this analysis because their pupil diameter contained less than 50 % of quality data, so observations from 22 participants were analyzed for this task. There were no significant main or pairwise interaction effects of the country of origin or type of judgment on the individual immigrant's right to free medical care for both pupil diameter and Paas scale (see tables 8 and 9).

Table 8. ANOVA results for Task 2 with pupil's diameter as dependent variable

Effect	Df	MSE	F	pes	p value
Country of origin	1, 21	0.00	0.00	<.001	.990
Type of judgement on the individual immigrant's right	1, 21	0.00	1.50	.067	.235
Country of origin × Type of judgement on the individual immigrant's right	1, 21	0.00	3.00	.125	.098

Table 9. ANOVA results for Task 2 with Paas scale as dependent variable

Effect	df	MSE	F	pes	p value
Country of origin	1, 21	2.63	2.29	.098	.146
Type of judgement on the individual immigrant's right	1, 21	0.85	1.62	.071	.218
Country of origin × Type of judgement on the individual immigrant's right	1, 21	0.77	0.01	<.001	.905

Discussion

In this study, we focused on assessing cognitive load of ordinary factual and normative judgments. We chose two pairs of tasks: (1) judgments on the cause, blame, and harm in medical and work institutional domains and (2) descriptive and prescriptive (deontic) judgments on the right to free medical care of an individual immigrant from two countries, Uzbekistan and Belarus.

The present study provides some limited evidence in support of the difference that exists between ordinary judgments of cause, blame, and severity of harm in terms of their propensity to evoke psychosensory pupillary response and subjectively perceived mental effort, both reflecting the variability in the cognitive load imposed on survey respondent when performing a pertinent survey task. The pupillometry data we obtained for Task 1 demonstrate that judgments on the severity of harmful consequences requires more cognitive load measured as pupil dilation in response to increased levels of arousal or mental effort [Mathôt, 2018] than judgments on the cause (though only in the medical institutional context). This finding may indicate that even if biased information models of blame judgment (e.g., [Alicke, 2000]) are well-founded and initial spontaneous evaluation of badness of action consequences and agent's blame directly influence the subsequent judgments of causality and intentionality, the graded evaluations of severity of negative consequences for the victim presupposes the involvement of Type 2 reflective and comparison-based processes loading heavily on working memory [Evans, Stanovich, 2013] and, consequently, the increase in cognitive load. Hence, these findings can be interpreted as a limited support of our H1.2 hypothesis (contra H1.1) and, in that way, a modest but promising demonstration of the construct validity of pupillometric data in adjudicating between different information models of normative judgment. However, our findings from the analysis of Paas's ratings of subjectively perceived mental effort for the same task do not offer any support to this hypothesis. These conflicting results may be explained by possible differences in sensitivity (discriminant validity) and construct validity of various objective and subjective measures of cognitive load that are known to be task-specific, i.e., providing different measurement quality for peculiar assignments, e.g., for driving simulation task, solving arithmetic problems, or, as in the current case, making a graded normative decision [Ayres et al., 2021]. An alternative explanation may follow from the relatively small-scale character of our study, and first and foremost—from the limited number of vignettes used that restrained the variability of scenarios reflecting the possible combinations of factor levels used in previous research. The further research is needed to overcome this potentially significant limitation.

As for the main effect of information cueing the medical institutional domain of action vs. the corporate dress code domain in two similar vignettes, the data from

Paas scale clearly shows that “medical”-related vignette evaluations made by our participants significantly differ from “work dress code”-related ones from the point of view of the task-evoked cognitive load, this finding being in favor of the hypothesis H2. The medical domain vignette scenario requires more mental effort than the dress code one. This result is also in a good agreement with similar findings from the previous substantive research on these types of ordinary judgments, demonstrating the pronounced influence of information about institutional domain of actions leading to negative side effects for a third party [Deviatko, Gavrilov, 2020]. The latter study also discovered a visible predominance in respondents’ sensitivity to comparable negative side effects occurring specifically in medical domain. In turn, the correspondence between findings from these two studies give some evidence of sensitivity and construct validity of perceived mental effort measure as employed at least for this type of judgment task.

Two results are especially noteworthy and need further clarification in future studies. First, the difference observed between general pattern of findings for pupillometry and Paas scale for Task 1. It has already been discussed that measures of pupil dynamics as an objective indicator of cognitive effort and Paas scale as a subjective measure of mental effort probably relate to different aspects of multimodal evaluation of cognitive load involved in specific tasks [Chen et al., 2016; Ayres et al., 2021]. Conceivably, pupil size may dilate, other things being equal, in response to increase in arousal and cognitive effort even during fast non-conscious, parallel processing, while Paas scale as a form of self-report on mental effort relies strictly on conscious, controlled processing. Purportedly, the previously noted high sensitivity of respondents to normative-evaluative judgments related to the medical institutional domain could provoke more intensive subjective experience of task-evoked mental effort in this case, alongside with more pronounced load imposed on the respondent’s cognitive system when performing the task of evaluation of severity of harmful consequences. However, it’s not clear why we did not observe the same pattern for blame judgments. Further research is badly needed here to arrive to more definite conclusions.

Second, we observed no visible difference between factual and normative-deontic judgments related to migrant rights vignettes (Task 2). This could potentially be explained in at least three ways: (1) this domain does not evoke sensitivity of both of our cognitive load measures to the difference in processing information related to descriptive vs. prescriptive judgements, (2) the specific scenarios chosen to measure the comparative cognitive load imposed by performing descriptive and prescriptive judgements were too similar (describing highly professional and educated legal migrants fluent in local language) to cause differential sensitivity to these types of judgements, (3) similar wording of questions might disguise the actual difference between these types of judgment, which in turn impede sensitivity to descriptive/prescriptive judgements. Again, more detailed further studies are needed to test the proposed hypothetical explanations for these preliminary negative findings.

At last, this study could be a good reference for an aspiring social scientist who wants to use a mobile eye-tracker for measuring cognitive load and has means to obtain it. While impressive, these instruments are currently also prone to noise, and particular observations could be excluded for numerous reasons ranging from a participant using

mascara before coming to an experiment to a hard drive becomes full unexpectedly, so research should be planned more carefully. This fact, as well as a relatively time-consuming and non-scaling procedure at present make these studies rather costly, but their potential to record a real-time observation on cognitive load makes it a worthwhile pursuit. Besides, new models of eye tracking devices that look and function like a regular pair of glasses are currently becoming available (and affordable) for academic researchers opening up the prospects for more scalable and statistically-powered future research in order to check the robustness of our present findings.

References

- Alicke M. D. (2000) Culpable Control and the Psychology of Blame. *Psychological Bulletin*. Vol. 126. No. 4. P. 556—574. <https://doi.org/10.1037/0033-2909.126.4.556>.
- Ayres P., Lee J. Y., Paas F., van Merriënboer J. J. G. (2021) The Validity of Physiological Measures to Identify Differences in Intrinsic Cognitive Load. *Frontiers in Psychology*. Vol. 12. Art. 702538. <https://doi.org/10.3389/fpsyg.2021.702538>.
- Chen F., Zhou J., Wang Y., Yu K., Arshad S. Z., Khawaji A., Conway D. (2016) Robust Multimodal Cognitive Load Measurement. Cham: Springer.
- Costa V. D., Rudebeck P. H. (2016) More Than Meets the Eye: The Relationship between Pupil Size and Locus Coeruleus Activity. *Neuron*. Vol. 89. No. 1. P. 8—10. <https://doi.org/10.1016/j.neuron.2015.12.031>.
- Cushman F. (2008) Crime and Punishment: Distinguishing the Roles of Causal and Intentional Analyses in Moral Judgment. *Cognition*. Vol. 108. No. 2. P. 353—380. <https://doi.org/10.1016/j.cognition.2008.03.006>.
- Deviatko I. F. (2007) Causality in Everyday Knowledge and in Sociological Research: An Outline of New Exploratory Approach. *Sociology: Methodology, Methods, Mathematical Modeling*. No. 25. P. 5—21. (In Russ.)
- Девятко И. Ф. Причинность в обыденном сознании и в социологическом объяснении: контуры нового исследовательского подхода // Социология 4М. 2007. № 25. С. 5—21.
- Deviatko I. F., Bogdanov M. B., Lebedev D. V. (2021) Pupil Diameter Dynamics as an Indicator of the Respondent's Cognitive Load: Methodological Experiment Comparing CASI and P&PSI. *RUDN Journal of Sociology*. Vol. 21. No. 1. P. 36—49. <https://doi.org/10.22363/2313-2272-2021-21-1-36-49>. (In Russ.)
- Девятко И. Ф., Богданов М. Б., Лебедев Д. В. Динамика диаметра зрачка как индикатор когнитивной нагрузки респондента: методический эксперимент по сравнению CASI и P&PSI вопросников // Вестник Российского университета дружбы народов. Серия: Социология. 2021. Т. 21. № 1. С. 36—49. <https://doi.org/10.22363/2313-2272-2021-21-1-36-49>.
- Deviatko I. F., Gavrilov K. A. (2020) Causality and Blame Judgments of Negative Side Effects of Actions May Differ for Different Institutional Domains. *SAGE Open*. Vol. 10. No. 4. <https://doi.org/10.1177/2158244020970942>.

- Deviatko I. F., Lebedev D. (2017) Through the Eyes of the Interviewer, through the Eyes of the Respondent: Outlining a New Approach towards the Assessment of Cognitive Load during the Interview. *Monitoring of Public Opinion: Economic and Social Changes*, No. 5. P. 1—19. <https://doi.org/10.14515/monitoring.2017.5.01>. (In Russ.)
Девятко И.Ф., Лебедев Д. В. Глазами интервьюера, глазами респондента: контуры нового подхода к оценке когнитивной нагрузки при проведении опроса // Мониторинг общественного мнения: Экономические и социальные перемены. 2017. № 5. С. 1—19. <https://doi.org/10.14515/monitoring.2017.5.01>.
- Evans J. S. B., Stanovich K. E. (2013) Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science*. Vol. 8. No. 3. P. 223—241. <https://doi.org/10.1177/1745691612460685>.
- Greene J. D., Morelli S. A., Lowenberg K., Nystrom L. E., Cohen J. D. (2008) Cognitive Load Selectively Interferes with Utilitarian Moral Judgment. *Cognition*. Vol. 107. No. 3. P. 1144—1154. <https://doi.org/10.1016/j.cognition.2007.11.004>.
- Guglielmo S. (2015) Moral Judgment as Information Processing: An Integrative Review. *Frontiers in Psychology*. Vol. 6. Art. 1637. <https://doi.org/10.3389/fpsyg.2015.01637>.
- Guglielmo S., Malle B. F. (2017) Information-Acquisition Processes in Moral Judgments of Blame. *Personality and Social Psychology Bulletin*. Vol. 43. No. 7. P. 957—971. <https://doi.org/10.1177/0146167217702375>.
- Hoogerheide V., Renkl A., Fiorella L., Paas F., van Gog T. (2019) Enhancing Example-Based Learning: Teaching on Video Increases Arousal and Improves Problem-Solving Performance. *Journal of Educational Psychology*. Vol. 111. No. 1. P. 45—56. <https://doi.org/10.1037/edu0000272>.
- Höhne J. K. (2019) Eye-Tracking Methodology: Exploring the Processing of Question Formats in Web Surveys. *International Journal of Social Research Methodology*. Vol. 22. No. 2. P. 199—206. <https://doi.org/10.1080/13645579.2018.1515533>.
- Höhne J. K., Schlosser S., Krebs D. (2017) Investigating Cognitive Effort and Response Quality of Question Formats in Web Surveys Using Paradata. *Field Methods*. Vol. 29. No. 4. P. 365—382. <https://doi.org/10.1177/1525822X17710640>.
- Höhne J. K., Lenzner T. (2018) New Insights on the Cognitive Processing of Agree/Disagree and Item-Specific Questions. *Journal of Survey Statistics and Methodology*. Vol. 6. No. 3. P. 401—417. <https://doi.org/10.1093/jssam/smw028>.
- Jbara A., Feitelson D. G. (2017) How Programmers Read Regular Code: A Controlled Experiment Using Eye Tracking. *Empirical Software Engineering*. Vol. 22. No. 3. P. 1440—1477. <https://doi.org/10.1007/s10664-016-9477-x>.
- Kaminska O., Foulsham T. (2014) Real-World Eye-Tracking in Face-to-Face and Web Modes. *Journal of Survey Statistics and Methodology*. Vol. 2. No. 3. P. 343—359. <https://doi.org/10.1093/jssam/smw010>.
- Knobe J. (2010) Person as Scientist, Person as Moralist. *Behavioral and Brain Sciences*. Vol. 33. No. 4. P. 315—329. <https://doi.org/10.1017/S0140525X10000907>.

Laeng B., Sirois S., Gredebäck G. (2012). Pupillometry: A Window to the Preconscious? *Perspectives on Psychological Science*. Vol. 7. No. 1. P. 18—27. <https://doi.org/10.1177/1745691611427305>.

Lavrakas P.J., Traugott M. W., Kennedy C., Holbrook A. L., de Leeuw E. D., West B. T. (eds.) (2019) Experimental Methods in Survey Research: Techniques that Combine Random Sampling with Random Assignment. Hoboken, NJ: John Wiley and Sons.

Malle B. F., Guglielmo S., Monroe A. E. (2014) A Theory of Blame. *Psychological Inquiry*. Vol. 25. No. 2. P. 147—186. <https://doi.org/10.1080/1047840X.2014.877340>.

Mathôt S. (2018) Pupillometry: Psychology, Physiology, and Function. *Journal of Cognition*. Vol. 1. No. 1. P. 16. <https://doi.org/10.5334/joc.18>.

Mathôt S., Fabius J., Van Heusden E., Van der Stigchel (2018) Safe and Sensible Pre-processing and Baseline Correction of Pupil-Size Data. *Behavior Research Methods*. Vol. 50. P. 94—106. <https://doi.org/10.3758/s13428-017-1007-2>.

Morewedge C. K., Kahneman D. (2010) Associative Processes in Intuitive Judgment. *Trends in Cognitive Sciences*. Vol. 14. No. 10. P. 435—440. <https://doi.org/10.1016/j.tics.2010.07.004>.

Neuert C. E. (2020) How Effective Are Eye-Tracking Data in Identifying Problematic Questions? *Social Science Computer Review*. Vol. 38. No. 6. P. 793—802. <https://doi.org/10.1177/0894439319834289>.

Neuert C. E. (2021) The Effect of Question Positioning on Data Quality in Web Surveys. *Sociological Methods & Research*. Online First published: February. <https://doi.org/10.1177/0049124120986207>.

Paas F. G. (1992) Training Strategies for Attaining Transfer of Problem-Solving Skill in Statistics: A Cognitive-Load Approach. *Journal of Educational Psychology*. Vol. 84. No. 4. P. 429—434. <https://doi.org/10.1037/0022-0663.84.4.429>.

Paas F. G., Van Merriënboer J. J. (1993) The Efficiency of Instructional Conditions: An Approach to Combine Mental Effort and Performance Measures. *Human Factors*. Vol. 35. No. 4. P. 737—743. <https://doi.org/10.1177/001872089303500412>.

Rossi P., Anderson A. (1982) The Factorial Survey Approach: An Introduction. In: Rossi P. H., Nock S. L. (eds.). *Measuring Social Judgments: The Factorial Survey Approach*. Beverley Hills, CA: Sage. P. 15—67.

Schein C., Gray K. (2014) The Prototype Model of Blame: Freeing Moral Cognition from Linearity and Little Boxes. *Psychological Inquiry*. Vol. 25. No. 2. P. 236—240. <https://doi.org/10.1080/1047840X.2014.901903>.

Schein C., Gray K. (2018) The Theory of Dyadic Morality: Reinventing Moral Judgment by Redefining Harm. *Personality and Social Psychology Review*. Vol. 22. No. 1. P. 32—70. <https://doi.org/10.1177/1088868317698288>.

Schmeck A., Opfermann M., van Gog T., Paas F., Leutner D. (2015) Measuring Cognitive Load with Subjective Rating Scales during Problem Solving: Differences between

Immediate and Delayed Ratings. *Instructional Science*. Vol. 43. No. 1. P. 93—114.
<https://doi.org/10.1007/s11251-014-9328-3>.

Sniderman P. M., Grob D. B. (1996) Innovations in Experimental Design in Attitude Surveys. *Annual Review of Sociology*. Vol. 22. P. 377—399. <https://doi.org/10.1146/annurev.soc.22.1.377>.

Stodel M. (2015) But What will People Think? Getting beyond Social Desirability Bias by Increasing Cognitive Load. *International Journal of Market Research*. Vol. 57. No. 2. P. 313—322. <https://doi.org/10.2501/IJMR-2015-024>.

Zaller J., Feldman S. (1992) A Simple Theory of the Survey Response: Answering Questions versus Revealing Preferences. *American Journal of Political Science*. Vol. 36. No. 3. P. 579—616. <https://doi.org/10.2307/2111583>.