



RESEARCH ARTICLE

Inconsistencies in repeated refugee status decisions

Philip Pärnamets^{1,2,3} | Alexander Tagesson³ | Annika Wallin³

¹Division of Psychology, Department of Clinical Neuroscience, Karolinska Institutet, Solna, Sweden

²Department of Psychology, New York University, New York, NY

³Lund University Cognitive Science, Lund University, Lund, Sweden

Correspondence

Philip Pärnamets, Division of Psychology, Department of Clinical Neuroscience, Karolinska Institutet, Solna, Sweden.
Email: philip.parnamets@ki.se

Funding information

Riksbankens Jubileumsfond, Grant/Award Number: M14-0138:1

Abstract

Consistency in civil servant decisions is paramount to upholding judicial equality for citizens and individuals seeking safety through governmental intervention. We investigated refugee status decisions made by a sample of civil servants at the Swedish Migration Agency. We hypothesized, based on the emotional demands such decisions bring with them, that participants would exhibit a compassion fade effect such that refugee status was less likely to be granted over time. To test this, we administered a questionnaire containing brief presentations of asylum seekers and asked participants to judge how likely they would be to give refugee status to the person. Crucially the first, middle, and final case presented were matched on decision relevant characteristics. Consistent with our hypothesis, we saw a significant decline in ratings. These effects were accentuated by the amount of time a participant had worked at the agency, consistent with depletion of affective resources, and attenuated in workers with greater responsibility and additional training. We conclude that active regulation of empathic and affective responses to asylum seekers may play a role in determining the outcome in refugee status decisions.

KEYWORDS

asylum seekers, compassion fade, decision-making, emotion, organizations

1 | INTRODUCTION

Equality before the law is an important concept in a just society, especially when an individual is subject to a decision or a judgment (cf. article 20, Charter of Fundamental Rights of the European Union 2000/C 364/01; article 7, United Nations Universal Declaration of Human Rights (1948); and article 14 and 26, International Covenant on Civil and Political Rights). Upholding this ideal requires an impartial and consistent application of relevant rules and regulations. A minimal definition of consistency is that a decision-maker should reach the same judgment or decision given the same decision relevant information. Changes in extraneous factors to the decision, including the passage of time and the decision-maker's emotional state should not affect its outcome. Thus, two legal cases should be decided equally if their relative merits are the same.

In the following study, we investigated the consistency of repeated refugee status decisions among civil servants at the Swedish Migration Agency. Refugee status decisions significantly affect the future life of individuals seeking protection and may entail death or severe psychological and physical harm. Understanding how pitfalls and biases of judgment and decision-making affect civil servants making refugee status decisions is an important step towards safeguarding societal and legal norms. Ironically, the nature and importance of these decisions increase also their emotional impact, something that by itself may reduce consistency. Here, we investigate the interplay between legal judgments, emotion regulation, and the maintenance of due process.

The increased number of individuals seeking asylum in Europe during the 2010s, with its peak during the "refugee crisis" of 2015, has brought questions concerning migration and the functioning of

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2020 The Authors. Journal of Behavioral Decision Making published by John Wiley & Sons Ltd

migration boards to the forefront of popular and political debates. Asylum decisions are difficult and involve evaluating the truthfulness of the asylum seeker's account (Granhag, Strömwall, & Hartwig, 2005), the objective dangers of the seeker's country of origin, and how this relates to relevant legal frameworks. In Sweden, individuals seeking asylum do so at the Swedish Migration Agency, and each case is handled by a case worker ("handläggare"). The case worker meets the applicant, gathers information, summarizes the case, and presents it, together with a recommendation, to another, often more senior, legal officer that, in turn, acts as a decision-maker ("beslutsfattare"). The process has many steps, one of which is determining if the applicant is a refugee by relevant international conventions and under national law. This is typically the most important decision, as Swedish law requires individuals given refugee status to be, at a minimum, granted temporary residence.

A salient aspect of both case workers' and decisions makers' work flow is that it concerns people in need. As such, affective and empathic processes are likely to impact their decision processes, as they are known to have considerable influence on judgments and decisions (Damasio, 1994; Haidt, 2001; Lerner, Li, Valdesolo, & Kassam, 2015). Empathic processes appear to be welcomed by the Swedish Migration Agency. Empathy is, for instance, one of three keywords selected to describe the agency's ethical values foundation (the other two are courage and clarity) and is frequently mentioned in both training materials and job advertisements.

There is considerable evidence that encounters with multiple victims or persons in need, compared with single individuals, change valuation processes during decision-making (Fetherstonhaugh, Slovic, Johnson, & Friedrich, 1997; Jenni & Loewenstein, 1997; Kogut & Ritov, 2005; Markowitz, Slovic, Västfjäll, & Hodges, 2013; Västfjäll, Slovic, Mayorga, & Peters, 2014; Slovic, Västfjäll, Erlandsson, & Gregory, 2017; for recent review, see Butts, Lunt, Freling, & Gabriel, 2019) and that prolonged exposure to persons in need can lead to permanent deficits in empathic capacities (Butts et al., 2019; Cocker & Joss, 2016; Figley, 1995; Sabo, 2006). The latter effect is known as compassion fatigue and the former as compassion fade. In one study investigating compassion fade, some participants were faced with a brief vignette describing the plight of a named child depicted in a photograph, whereas other participants were given similar information about two children (Västfjäll et al., 2014). For both conditions, participants were asked to make donations that would benefit the child/children described. More money was donated in the one-child condition compared with the two-child condition, and participants also reported feeling greater amounts of affect in the one-child condition. Similarly, iconic photographs (such as the ones depicting Aylan Kurdi or Phan Thi Kim Phuc) appear to give rise to stronger reactions in the general public than conventional news reports do (Slovic et al., 2017).

The prototypical compassion fade effect is elicited by comparing how participants respond to single persons in need compared with groups of varying sizes. When refugee status is judged, civil servants handle a single individual at a time, even if that person is seeking refugee status together with their family, and thus, the typical contrast

between individuals and groups cannot be implemented in our sample without compromising the representativeness of the stimuli. Instead, we hypothesized that sequentially encountering multiple affective targets—*asylum seekers*—over a short period of time could engender the same need to downregulate empathic processes (cf. Cameron & Payne, 2011; Keyesers & Gazzola, 2014; Zaki, 2014). If so, this would generalize the compassion fade effect temporally. We therefore opted for a repeated measures design by creating a questionnaire containing practically identical cases at the beginning, middle, and end. We hypothesized that if empathic regulation was a factor for our participants, their judgments approving of refugee status would decline over time, an effect comparable to what has been seen in earlier studies on compassion fade.

Compassion fatigue affects people who regularly interact with persons in need or who are traumatized. Examples are individuals in clinical professions such as caregivers and nurses as well as emergency service personnel (Cocker & Joss, 2016; Figley, 1995). It shows itself as lower levels of empathy, or compassion, over time. There is also evidence that immigration judges in the United States have high levels of secondary posttraumatic stress symptoms, indicative of compassion fatigue (Lustig et al., 2008; Lustig, Karnik, Delucchi, & Tenakoon, 2008). It has been argued that legal judgments have a particularly strong effect on compassion fatigue, partly due to the requirement to act impartially (Saakvitne, Pearlman, & Abrahamson, 1996). Unlike compassion fade effects, compassion fatigue is, by definition, difficult to elicit in a controlled study. However, although we did not directly measure compassion fatigue in our sample, we operationalize it through longer work experience at the Migration Agency. If workers at the agency do experience compassion fatigue, we hypothesized that this would translate to greater observed inconsistencies in judgments in our study because those individuals would be quicker to downregulate their empathic responses during the repeated exposures to asylum seekers that our questionnaire entailed. In addition, we asked participants to self-report experienced compassion and effort at the end of the questionnaire. We reasoned that differences in such ratings might provide clues to empathic processes. If empathic down regulation increases inconsistencies, inconsistencies should be smaller for individuals with more self-reported compassion. Similarly, they should be smaller for individuals with less self-reported effort.

Expertise in a relevant domain is known to mitigate biases in judgment and decision-making (Kirk, Harvey, & Montague, 2011; Smith & Kida, 1991), and it is possible that compassion fade effects are mitigated by involvement and knowledge with the case at hand (Markowitz et al., 2013). In one sense, all the civil servants in this study were experts, because they were employed at the Migration Agency and worked with asylum decisions. Nevertheless, decision-makers at the Migration Agency are typically granted extra training and are given larger responsibility than case workers. Legal skill or other similar qualifications is also a reason for experienced case workers to be promoted to decision-makers. Note that legal training is otherwise not a requirement to work at the Migration Agency. Because the judgments and decisions made by the civil servants of

the Agency involve interpretations of statutes and legal conventions, this selection and extra training provide additional subject matter expertise to decision-makers relative to case workers. In the context of this study, we operationalize this as expertise and hypothesize that decision-makers will be less inconsistent compared with case workers. However, an alternative operationalization of expertise could be work experience at the agency, leading to a competing prediction to the one outlined in the previous paragraph. Another possibility is that work experience affects decision-makers and case workers differently, perhaps due to differences in organizational roles and perceived control at the workplace, factors known to affect the emergence of compassion fatigue (Newell & MacNeil, 2010).

Individual differences in information processing and reflexivity might also affect consistency. One common measure of this is the cognitive reflection test (CRT; Frederick, 2005; Bialek & Pennycook, 2018). CRT correlates with performance on multiple tasks and measures, including bullshit detection (Pennycook, Cheyne, Barr, Koehler, & Fugelsang, 2015), recognition of manipulation one's own prior attitudes (Strandberg, Sivén, Hall, Johansson, & Pärnamets, 2018), and a host of indices related to analytical thinking (Pennycook, Fugelsang, & Koehler, 2015). In our second wave of data collection, we added the CRT as an exploratory measure. We reasoned that participants with higher CRT scores—individuals less prone to following their “gut feelings” and more prone to analytic reflexivity—might also be less likely to exhibit inconsistencies in our task.

If civil servants at the Migration Agency are like most other people, we should expect inconsistencies in judgments. Any such inconsistencies are, however, a challenge to legal certainty. It is therefore important to find and adjudicate possible sources of inconsistency. We attempt to single out and identify emotionally regulative influences on the legal judgments of experts, by looking at how work experience, expert knowledge, and involvement affect the consistency in judgments of practically identical refugee status cases.

In sum, we hypothesized that participants would be less likely to grant refugee status for matched cases the later in the questionnaire they encountered them. We expected these effects to be attenuated as a result of reported compassion at the end of the task and accentuated for participants who experienced the task to be effortful. We hypothesized that decision-makers would exhibit smaller inconsistencies compared with case workers and explored if workplace experience moderated any effects observed. Finally, we expected participants with higher CRT scores to exhibit smaller inconsistencies compared with participants with lower CRT scores.

2 | METHOD

2.1 | Participants

The Swedish Migration Agency is organized into units—or teams—of about 15–20 individuals. Each unit consists of both case workers and decision-makers and is relatively specialized when it comes to the cases they handle. In collaboration with the agency, we singled out

units suitable for the study (e.g., units that handled the type of cases used in the stimulus). These were visited in two waves. The first wave of data collection was conducted between March and April 2017, and the second wave was conducted between November 2017 and January 2018. Participants were recruited from units located in southern and western Sweden.

In the first wave, we visited four Migration Agency units. A total of 74 participants (14 males, 60 females) completed the questionnaire. Thirty-nine participants worked as case workers and 35 worked as decision-makers. In the second wave, we originally targeted 150 participants, based on simulations of results from the first wave. However, due to reorganizations at the Agency resulting from the sharp decrease in asylum seekers during the period of our study, this was not possible. The second wave of data collection was slightly delayed due to this reorganization. Eight units were visited in the second wave of data collection, and 88 participants (24 males, 64 females) completed the questionnaire. Sixty-one participants worked as case workers and 27 as decision-makers.

All participants were at the time of participation employees of the Swedish Migration Agency. Participation took place during weekly staff meetings. All participation was voluntary, and workers who did not want to participate were asked to hand in a blank questionnaire to protect their anonymity to the group. No compensation was given for participating in the study.

The study was approved by the Lund University Ethics board, D. nr. 2017-815.

2.2 | Materials

The questionnaire consisted of eight fictional scenarios, each depicting an individual or a family seeking refugee status in Sweden with a length of around 300 words. Scenarios were developed together with experts at the Migration Agency with the aim to be as realistic as possible.

The critical comparison in our study is between Scenarios 1, 5, and 8. Scenarios 1 and 8 were counterbalanced between participants (see Figure S1). The scenarios were constructed to be as similar as possible. Features that are irrelevant to judging refugee status were altered in each scenario, such as name, age (adult), and hometown of the persons presented. The persons depicted in the scenarios came from regions in the same country (Afghanistan) that our experts (and the Migration Agency's expert institution for legal and country of origin information, Lifos) at the time of the study considered to be equally (un)safe. Finally, although all scenarios involved violent persecution, some details were varied to ensure that participants would not react to reading the same story. All details varied were such that experts at the Migration Agency judged that the legal merits of the cases were not affected by the changes. The experts worked with the agency's competence development center and did not take part in the study.

For the remaining scenarios, a similar procedure was followed. Some filler scenarios depicted individuals, and some individuals

arriving with minors were described as families. The people depicted were described as coming from Iraq and Somalia. Order of presentation (families or individuals) was counterbalanced.

All scenarios were paired with a photograph. Photographs were sourced online using Google Image Search and manipulated in an image editing program to standardize size and to remove any background information. Photographs were included to increase the realism of the questionnaire and to maintain methodological similarity with previous work on compassion fade. Photographs for Critical Scenarios 1 and 8 were counterbalanced between participants and texts. The full list of the resulting questionnaire versions (eight in total), is given in Figure S1.

Each scenario had the same two questions for the participant at the bottom of the page. The first was the judgment of refugee status: "How likely is it that you would deem X to be a refugee?," where X was the name of the person presented in the scenario. This question was formulated to capture one of the important parts of the asylum-seeking process in Sweden—determining if the person is a refugee. Finding that the person is means that laws and convention governing refugees are applicable. Hence, this would be a type of judgment that participants would be familiar with in their work. The second was a judgment about the participants' colleagues: "Do you think that your assessment conforms with those of your colleagues at the migration unit?"¹ Both questions were answered through a visual analog scale stretching the width of the questionnaire paper. The scale was anchored in both ends with "not likely at all" and "very likely."

Following the eight scenarios, participants were asked for some demographic information. Participants were also asked to indicate how much compassion they felt with the persons depicted in the scenarios and how effortful they felt responding to the questionnaire. Both questions were answered using a visual analog scale. The compassion question was anchored with (in Swedish) "No compassion" and "A great amount of compassion," whereas the effort question was anchored with (in Swedish) "No effort at all" and "Very effortful." In the second wave of data collection, an additional 3-item standard cognitive reflection test (Frederick, 2005) was added.

2.3 | Procedure

Each session started with the experimenter giving verbal information about the project. The members of the unit visited read the consent form and signed. Participants were told that they would not be allowed to communicate with one another during the session. They were asked to read the instruction page carefully and to base their answers on the information found in the scenarios.

Each participant was given 3 min to address each scenario and was told that they were not allowed to progress to the next page of the questionnaire until instructed to do so. The experimenter

informed the group when 30 s remained on each scenario. Additionally, participants were instructed not to go back in the questionnaire to reread scenarios or change responses. This was implemented to ensure that no participants compared the critical scenarios with each other. The whole procedure took no more than 30 min, which was the time allotted for the experiment.

2.4 | Analysis

All responses from the visual analog scales were converted to 0–100 scales to facilitate comparison. We compared ratings to the first (S1), fifth (S5), and final (S8) scenario. Because the scenarios contained the same decision relevant information, *consistent* participants should judge the scenarios equally, and systematic average deviation would reveal *inconsistencies* in judgments.

Analyses were carried out using hierarchical regression models and Bayesian estimation as implemented in the package *brms* (Bürkner, 2017) in the statistical language R (R Core Team, 2018). Weakly informative, zero-centered priors were used and are given in the supplementary methods. For all models, all slopes and intercepts were varied by participant ID and questionnaire ID, thus fitting the maximal group-level varying structure (Barr, Levy, Scheepers, & Tily, 2013).

2.5 | Open science practices

The data presented in this article, analysis code, and stimulus materials can be found at the Open Science Framework: https://osf.io/62ua8/?view_only=916b09db425f4be1ba4e86e5bb196ca8.

3 | RESULTS

3.1 | Analysis of ratings

We first compared participants' ratings on the three critical scenarios to evaluate if ratings systematically changed over time. We combined our two waves of collection to maximize precision. An analysis of the first and second data collection wave data separately yielded similar results and conclusions and are reported in the supplementary results.

Average ratings on the first scenario (S1) were $M = 52.6$, $SD = 26.6$, on the fifth scenario (S5) $M = 43.4$, $SD = 24.1$, and on the final scenario (S8) $M = 40.7$, $SD = 22.3$, see also Figure 1a. To assess if these changes were statistically robust, ratings were analyzed using Bayesian multilevel regression, with S5 and S8 dummy coded. We found that participants reliably declined in their willingness to assign refugee status from the first to the fifth scenario ($b = -8.5$, $SE = 2.2$, 95% CrI [credible interval] = $[-12.8, -4.1]$) and likewise from the first to the final (eighth) scenario ($b = -11.1$, $SE = 2.3$, 95% CrI = $[-15.4, -6.5]$, see Figure 1b). Although the difference appeared to increase between S5 and S8, the magnitude of this change was small ($b_{diff} = -2.7$, 95%

¹As can be seen in Figure S5, some participants appeared to have interpreted the question as "How likely is it that your colleagues would deem X to be a refugee?" as given by the many responses identical to the participants' own response. We therefore do not analyze responses to this question further. The data are available from our OSF page.

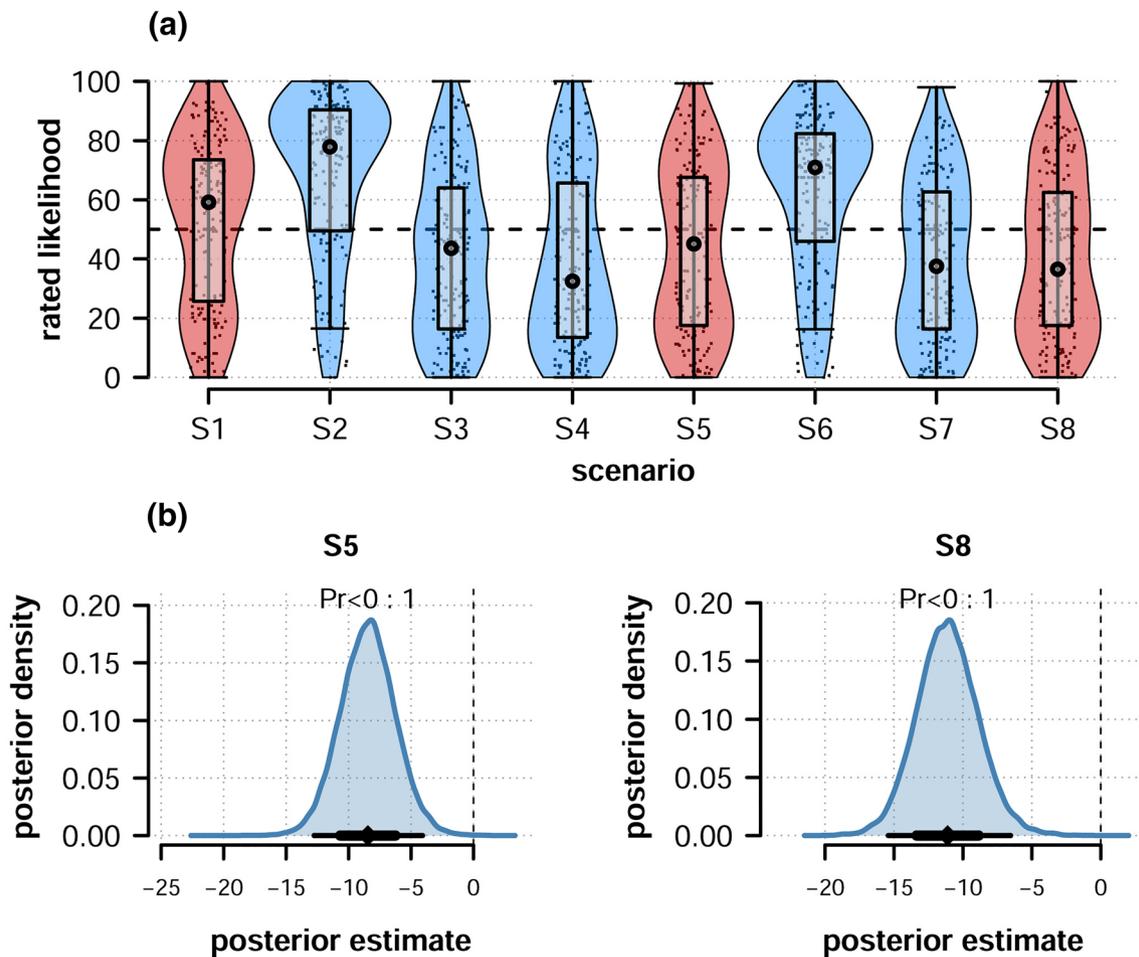


FIGURE 1 (a) Ratings of likelihood to grant refugee status to all scenarios of the questionnaire. The three critical scenarios are colored red. Each point represents ratings from a single participant. Box plots depict median (large circle), 25th and 75th quantile (box edges) values, as well as 1.5*interquartile range (hinges). (b) Density plots of posterior samples for the regression coefficients estimating the population-level effects contrasting S5 and S8 ratings to baseline (S1) ratings [Colour figure can be viewed at wileyonlinelibrary.com]

CrI = $[-7.6, 2.3]$). Another way of gauging the practical significance of the decline in ratings from S1 to S8 is to convert them into binary decisions, taking the midpoint of the scale as a cutoff—although we emphasize that there are no guarantees that this choice of cutoff corresponds to how participants used the scale. Using this metric, we found that 57.4% of decisions to S1 granted refugee status, whereas 38.9% of decisions to S8 did the same—a decline of 18.5%.

An unpredicted aspect of our data, as evident from Figure 1a, was the great dispersion of ratings for each individual scenario. To better understand the sources of this variation, we examined the estimated variance of the group-level intercepts and slopes from the same regression model as above. Consistent with the observed dispersion, the estimated standard deviation was large for the by participant intercepts (estimate = 19.7, SE = 1.5, 95% CrI = $[16.9, 22.6]$). The variation for the effects of each scenario were however much smaller both for S5 (estimate = 4.1, SE = 2.8, 95% CrI = $[0.2, 10.4]$) and for S8 (estimate = 3.2, SE = 2.5, 95% CrI = $[0.1, 9.4]$), indicating that although participants varied strongly in their baseline estimates, there was considerably less variation in the effects of subsequent scenarios. The variation attributable to questionnaire ID was considerably smaller for

the intercepts (estimate = 2.1, SE = 1.6, 95% CrI = $[0.8, 6.1]$) and of similar size for the S5 (estimate = 2.6, SE = 1.9, 95% CrI = $[0.1, 7.2]$) and S8 (estimate = 3.0, SE = 2.1, 95% CrI = $[0.2, 7.7]$) slopes. Importantly, the low estimates for the questionnaire ID terms suggests that the variation in order of scenarios mattered little for the observed responses.

We additionally conducted an exploratory analysis to control for participants' ratings to the filler scenarios (S2–S4 and S2–S4 & S6–S7, respectively for S5 and S8). The motivation for this was that the observed decline in ratings could be the result of a compensatory mechanism such that participants who had given high ratings earlier in the questionnaire gave lower ratings on the critical scenarios. We added the mean filler rating, standardized, to the regression model together with the crucial interactions with the S5 and S8 indicators. The results indicated very small compensatory effect. For S5 the posterior probability was 80.0% of a moderating effect, estimated as $b = 1.8$, SE = 2.2, 95% CrI = $[-2.5, 6.0]$. For S8, the posterior probability was 84.0% of a moderating effect, estimated as $b = 2.2$, SE = 2.3, 95% CrI = $[-2.3, 6.7]$. In both cases, even for highly outlying participants, the estimates were not sizeable enough to completely reduce

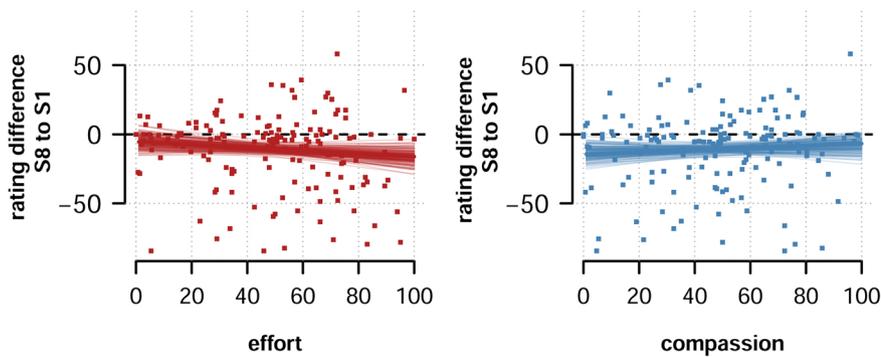


FIGURE 2 Scatter plot depicting rating difference between final (S8) and first (S1) scenario for all participants. Left panel as a function of self-rated effort and right panel as a function of self-rated compassion. Thick lines depict the mean posterior regression estimates, and thin lines show 150 draws from the posterior distribution of regression lines, indicating uncertainty around the mean estimate [Colour figure can be viewed at wileyonlinelibrary.com]

the main effect of scenario, further indicating the robustness of our findings.

Finally, having found declines in ratings on our critical scenarios, we explored if average ratings differed for our four noncritical filler scenarios depending on position. These four scenarios, in positions 2/6 and 3/7, appeared early or late in the questionnaire for different participants, making between-subjects comparisons of average ratings possible. We found that the mean differences were directionally in line with our hypothesis in three of four cases—lower ratings for later compared with early presentations: Asad,² $M_{diff} = -1.4$; Zainab, $M_{diff} = -11.0$; Sayyid, $M_{diff} = 2.8$; and Geesi, $M_{diff} = -5.8$. Next, for these four scenarios, we regressed on ratings including a dummy variable indicating if the scenario was presented in the first half (*early*) or second half (*late*) of the questionnaire. Intercepts and slopes varied both by participants and by scenario ID. The results indicated that there was a small decline in ratings when comparing ratings for late scenarios to ratings for early ones ($b = -3.70$, $SE = 2.3$, 95% CrI = $[-8.2, 1.0]$), with the posterior probability of a negative change in ratings from early to late scenarios being 95.0%.

In sum, we found that participants were inconsistent in their refugee status decisions for identical scenarios and that this inconsistency took the form of a decided decline in willingness to grant refugee status, as predicted by a compassion fade framework. Participants were highly variable in their overall ratings but demonstrated less variability in the lowered ratings between earlier and later scenarios.

3.2 | Self-rated effort and compassion

We tested if the self-rated effort of filling out the questionnaire, or the self-rated compassion felt towards the persons presented in the scenarios, moderated changes in ratings. We hypothesized that effort should increase inconsistencies and compassion decrease them. Small and variable effects along these lines were found (see also Figure 2). The interaction between effort and S8 difference in ratings had 90% posterior probability of negative direction and was estimated as $b = -2.7$, $SE = 2.1$, 95% CrI = $[-6.8, 1.6]$. Similarly, the interaction between compassion and S8 difference in ratings had a more modest 81.5% probability of positive direction and was estimated as $b = 1.9$, $SE = 2.2$, 95% CrI = $[-2.5, 6.2]$. For all model coefficients see supplementary results.

We report additional analyses investigating the relationship between compassion and the magnitude of ratings in the supplementary results, finding that participants who on average gave higher ratings over all scenarios also rated their compassion higher, but that this correlation was not observed for Scenarios 1 and 8 specifically.

3.3 | Effect of role and work experience

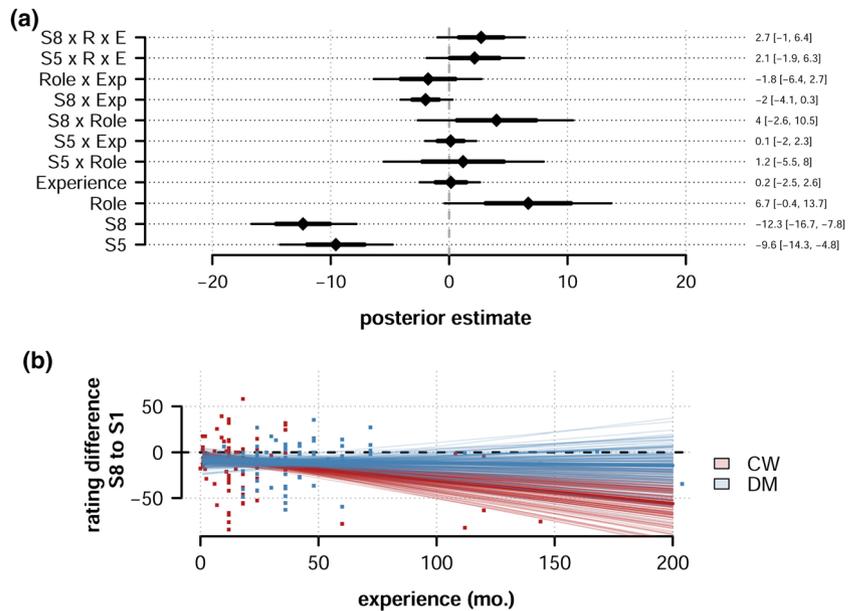
We next investigated if expertise moderated the effects described above. We tested two possible predictors: role at the Migration Agency (“case worker” or “decision-maker”) and time working at the agency. These were entered into a multilevel regression model together with indicators for S5 and S8, as well as their interactions. The resulting model coefficients are plotted in Figure 3a. Although the results are not conclusive, several effects emerged that warrant attention. Overall, the posterior probability that decision-makers gave higher ratings compared with case workers was 96.8%, with the difference between the roles estimated as $b = 6.7$, $SE = 3.6$, 95% CrI = $[0.4, 13.6]$. Decision-makers' ratings at S8 exhibited less decline compared with case workers, with a posterior probability of 88.5% for a positive effect, which was estimated as $b = 4.0$, $SE = 3.3$, 95% CrI = $[-2.5, 10.4]$. Ratings declined more with longer work experience. Each extra year worked was estimated to increase the difference to baseline by $b = -2.0$, $SE = 1.1$, 95% CrI = $[-4.1, 0.3]$, with a 95.8% posterior probability of a negative effect.

The regression results presented as the difference between S8 and S1 are visualized in Figure 3b. For example, for participants with 12 months of work experience, the model predicts an average difference of -9.0 , 95% CrI = $[-14.3, -3.9]$, with the difference between case workers and decision-makers estimated as 0.6, 95% CrI = $[-8.3, 9.5]$. After 48 months, the average difference is instead estimated as -15.0 , 95% CrI = $[-20.5, -9.1]$, with the difference between case workers and decision-makers being -7.5 , 95% CrI = $[-15.7, 0.8]$ (negative numbers indicate more decline in case workers).

We performed several additional analyses to better understand the relationship between work experience and change in ratings. These analyses are reported in full in the supplementary results. First, to check that work experience did not just make employees more likely to grant or refuse refugee status in general, we explored if experience correlated with responses to each scenario rather than just with the change in ratings to the critical scenarios. This was not the

²The names refer to the person depicted in each scenario.

FIGURE 3 (a) Coefficients from model estimating effects of work experience and role on ratings. Points represent the posterior mean, thick bars the standard deviation of the posterior, and thin bars the 95% credible interval. Posterior mean and credible intervals are printed in the right margin. (b) Scatter plot depicting rating difference between final (S8) and first (S1) scenario for all participants as a function of work experience (months) and role. Case workers are shown in red and decision-makers in blue. Thick lines depict the mean posterior regression estimates, and thin lines show 150 draws from the posterior distribution of regression lines, indicating uncertainty around the mean estimate [Colour figure can be viewed at wileyonlinelibrary.com]



case (see Figure S2). Second, we found that work experience correlated with expressed effort and compassion, but that the effect of experience on rating change was not mediated by these variables (see Figures S3–S4).

To summarize, we found that longer work experience tended to mean larger inconsistencies between S1 and S8, and that working as a decision-maker attenuated the inconsistency effects observed.

3.4 | CRT scores

CRT scores were evenly distributed, 23% of participants scored 0, 30% scored 1, 26% scored 2, and 21% scored 3. Because CRT was only part of the questionnaire in the second wave of data collection, and some participants did not complete the CRT questions, 73 participants were included in the following analyses. Despite this reduced sample, the analysis indicated that there was a likely moderating effect of CRT scores on a S8 difference to S1 in the predicted direction. This with an 82% posterior probability of a positive sign, which was estimated as $b = 2.5$, $SE = 2.8$, $95\% \text{ CrI} = [-3.0, 7.9]$ (see Figure 4, full regression table reported in supplementary results). In practice, the model predicted a change in S8 ratings compared with S1 for a participant with a CRT score of 0 to be -15.0 , $95\% \text{ CrI} = [-26.0, -5.0]$. For a participant with a CRT score of 3, the model predicted the change to be -7.8 , $95\% \text{ CrI} = [-18.7, 8.9]$, resulting in an estimated difference between those participants (CRT 0–CRT 3) of -7.5 , $95\% \text{ CrI} = [-23.7, 8.9]$.

4 | DISCUSSION

We investigated repeated judgments of refugee status among civil servants at the Swedish Migration Agency. The judged likelihood of granting refugee status declined for matched cases by position in the

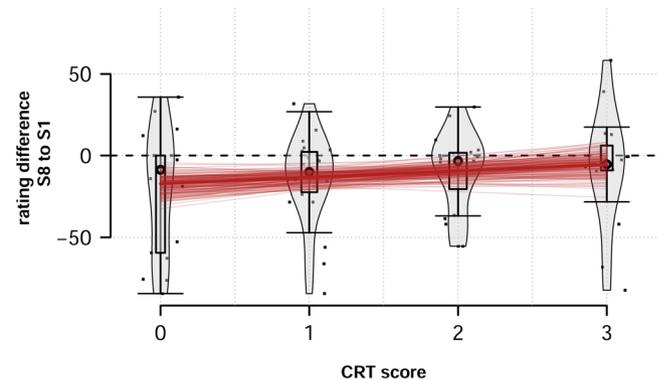


FIGURE 4 Violin plots depicting rating difference between final (S8) and first (S1) scenario for participants who completed the CRT split by their CRT scores. Thick line depicts the mean posterior regression estimate, and thin lines show 150 draws from the posterior distribution of regression lines, indicating uncertainty around the mean estimate [Colour figure can be viewed at wileyonlinelibrary.com]

survey. This decline over time was moderated by expertise and work experience. Finally, we found some indications that individual differences in perceived effort and compassion expended during the survey, as well as differences in cognitive reflection ability, moderated these effects. Our findings broadly support a compassion fade framework, where empathic and affective engagement in a task is a scarce resource that is downregulated during intensive decision-making.

Our primary aim was to investigate repeated judgments, where the outcome of the judgment was normatively expected to be the same. We hypothesized that our participants would become less likely to grant refugee status as the questionnaire progressed. This is precisely what we observed. The effect sizes were notable, with an average difference of 11/100 points on the rating scale. We also observed large, nonhypothesized variability in responses between our participants, suggesting little agreement on how to judge the cases in the

first place. This dovetails with findings from a recent investigation comparing the approval rates of officers at the Regional Asylum Offices in the U.S. Department of Homeland Security. Here, large differences in approval between officers was observed although cases are assigned to them at random (Ramji-Nogales, Schoenholtz, & Schrag, 2009). Nevertheless, those data cannot inform the inconsistency of judgments *within* each officer that we observe.

Why do we find inconsistencies? Although we designed our investigation based on hypotheses derived from theories of affective regulation based on compassion fade, a possibly competing explanation might be gleaned by considering order effects. Order effects occur when judgments or decisions are altered through the sequence of presentation of options and have been observed in both experimental and naturalistic contexts (Bruine de Bruin, 2006; Bruine de Bruin & Keren, 2003; Page & Page, 2010; Wilson, 1977), including legal judgments (e.g., Englich, Mussweiler, & Strack, 2006). It has been proposed that the direction of comparison, that is, whether new options have primarily positive or negative features, determines if order effects take the form of gradually increasing or decreasing judgments over time (Bruine de Bruin & Keren, 2003).

We interpret our findings to primarily support an explanation based on compassion fade rather than direction of comparison. A key reason is that compassion fade yields straightforward directional predictions consistent with our observations. As the questionnaire progresses, judgments are expected to become harsher, which is what our data show. By contrast, predictions from other mechanisms, such as direction of comparison are not clearly directional. Although it is common to observe positive correlation between judgments and serial position (Bruine de Bruin, 2006), negative correlations are also possible (Li & Epley, 2009), depending on if new objects of judgments are evaluated primarily on negative or positive characteristics (Bruine de Bruin & Keren, 2003). It is possible that individual differences in evaluation style gives rise to some of the variation in the effects we observe, and further work is necessary to fully account for our findings.

Another related finding in the literature is the time-order effect documenting that decision-makers change their decisions depending on the amount of hours spent working (Danziger, Levav, & Avnaim-Pesso, 2011; but see Glöckner, 2016; Persson, Barrafreem, Meunier, & Tinghög, 2019). One recent study found that orthopedic surgeons were 33% less likely to assign a patient to surgery if they met them at the end of their day compared with the start of the day (Persson et al., 2019). The precise cause of this effect is not entirely understood, but it is hypothesized that it is partially caused by willpower depletion leading decision-makers to revert to selecting default options. In the case of surgeons, the default (and less cognitively taxing) option is to send patients home rather than refer them for surgery. Our data cannot rule out a time-order explanation. The prediction from depletion theory regarding our participants would depend on what the relevant default is for civil servants at the Migration Agency. This will be an important question for future work to gauge.

It is also possible that affective processes, like the ones hypothesized here, contribute to time-order effects. There are also additional reasons, we argue, for preferring a compassion fade interpretation of our findings. We found correlations between self-rated effort and compassion matching our predictions; participants who experienced the questionnaire as more effortful were also less consistent, and participants who report experiencing more compassion at the end of the questionnaire were more consistent. Jointly, these effects can be accounted for within a compassion fade framework, which predicts that effort should translate into larger fading effects and maintained compassion into smaller fading effects. Finally, although our sample was more limited for the analysis of CRT scores, those findings indicate that participants higher in reflective abilities were more consistent. It is possible that these participants had alternate strategies available to them, perhaps relying less on affective response altogether in their decision-making, and that this enabled them to be more consistent in their judgments.

The compassion fade framework suggests two related mechanisms that explain changes in participants' judgments: identifiability and empathic concern. As participants progress through the questionnaire, they displace previously encountered information about the persons presented with novel information. This likely leads to each subsequent case standing out less, hence becoming less identifiable, which in turn shifts participants' judgments (cf. Jenni & Loewenstein, 1997; Kogut & Ritov, 2005). Relatedly, decreasing identifiability as well as the taxing process of encountering repeated cases affects the empathic resources participants are willing to utilize. Each case presents a person in plight and under dire circumstances. Empathizing to the same degree with the last case as with the first may be difficult or undesirable. Instead, we argue, in line with an emerging view in the affective sciences, that empathy is a motivated process amenable to direct regulation (Cameron & Payne, 2011; Keyesers & Gazzola, 2014; Zaki, 2014). We see participants as rationally downregulating their empathy somewhat analogously to how downregulation of empathic concern can arise from encountering larger victim group sizes in the standard elicitation of compassion fade phenomena (cf. Butts et al., 2019).

The declines in likelihood of granting refugee status that we observed were moderated by expertise and work experience. We found that participants who had been employed longer were more inconsistent, even if the effect size was relatively modest. This finding is in line with an emotion regulation explanation assuming that case workers with long-term exposure to highly emotional cases are more prone to quickly downregulate their emotional responses. In this case, the downregulation affects consistency of their judgments. This effect might be similar to the so-called compassion *fatigue* effect that is well documented in professions such as nursing (Cocker & Joss, 2016; Figley, 1995; Newell & MacNeil, 2010; Sabo, 2006). However, the relationship between compassion fatigue and time working is not straightforward. Results are mixed for research concerning healthcare professionals. Some work reports, or negative relationship between time working and measures of compassion fatigue or burnout (Ray, Wong, White, & Heaslip, 2013; Thompson, Amatea, & Thompson,

2014), whereas other work indicated a positive relationship (Lasalvia et al., 2009; Rossi et al., 2012). In legal professionals—judges—Jaffe, Crooks, Dunford-Jackson, and Town (2003) report higher levels of vicarious trauma with increasing work experience. Additional work specifically designed to target this question will be necessary to address the relationship between work experience and compassion fatigue at the Migration Agency and its role, if any, in explaining inconsistencies of the kind reported here.

We also found that decision-makers were less inconsistent compared with case workers. This suggests that expertise might mitigate some of the risks arising from emotion regulation (cf. Markowitz et al., 2013). It is difficult to say what factor is responsible for this latter effect. It could be due to the extra training decision-makers receive, perhaps providing them with judgment tools that are less emotionally taxing. However, another possibility is that decision-makers tend to have a lower day-to-day involvement with persons seeking refugee status. This should lead them to experience less day-to-day emotional strain and make emotion regulation less central for their well-being, as well as generally lower the risks of compassion fatigue. Decision-makers also have greater workplace responsibility that may shield them from organizational factors affecting workplace stress, including compassion fatigue (Newell & MacNeil, 2010).

Although the effects reported are a cause for concern, the findings are not without their limitations. We were not able to gather enough participants to meet our original sample size target. This means that the evidence for many of the moderating factors we investigated is more limited than would have been desirable. Similarly, our measures were adapted to the real-time constraints of running the study. Future work will want to use more developed self-report measures of effort and compassion, coupled with trait measurements and, ideally, physiological measurement (cf. Västfjäll et al., 2014). Additionally, civil servants at the Migration Agency must make binary decisions about refugees and not graded ones as employed here. Although employing a continuous scale likely increased the probability of observing an effect, this choice somewhat reduced the external validity of our findings. Future work should attempt to replicate our findings using more naturalistic decision categories. The scenarios we used were developed with experts to make them as realistic as possible. Nevertheless, they contain much less information about the persons seeking refugee status than would normally be available, and it is possible that providing richer life histories accentuates or attenuates the effects observed here.

To conclude, using a sample of civil servants, we demonstrated that repeated judgments about refugee status are susceptible to compassion fade effects. These findings highlight the importance of understanding the interplay between affective processes and decision-making. In high-stakes situations, legal consistency is vital, but, as we have demonstrated, the emotions they evoke endanger the rule of law—and this when it matters the most. We observed a tendency of affective regulation with longer work experience, and this means that we cannot trust time and experience to increase the consistency of civil servants' refugee status judgment. Understanding how these results translate into day-to-day work will be of

considerable importance moving forward. Learning more, we might design work environments with sufficient emotional and decision support to stop compassion fade from hindering civil servants in effectively pursuing their work.

ACKNOWLEDGMENTS

We wish to thank the Swedish Migration Agency and their employees for allowing us to conduct the study. In particular, we wish to thank Christian Andersson, Frida Lee, and Cornelia Hagberg for being instrumental in assisting with stimulus design and facilitating contact with the Migration Agency. We also thank Wändi Bruine de Bruin, Petter Johansson, and the VBE-group for the helpful feedback on earlier versions of this manuscript. A. W. gratefully acknowledges the financial support from Riksbankens Jubileumsfond Grant M14-0138:1.

AUTHOR CONTRIBUTIONS

P. P. and A. W. developed the study concept. All authors contributed to the study design. Testing and data collection were performed by A. T. and A. W. The data were analyzed by P. P. All authors drafted and revised the manuscript and approved the final version of the manuscript for submission.

ORCID

Philip Pärnamets  <https://orcid.org/0000-0001-8360-9097>

Alexander Tagesson  <https://orcid.org/0000-0003-4139-7377>

Annika Wallin  <https://orcid.org/0000-0001-6019-8264>

REFERENCES

- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278.
- Bialek, M., & Pennycook, G. (2018). The cognitive reflection test is robust to multiple exposures. *Behavior Research Methods*, 50(5), 1953–1959. <https://doi.org/10.3758/s13428-017-0963-x>
- Bruine de Bruin, W. (2006). Save the last dance II: Unwanted serial position effects in figure skate judgments. *Acta Psychologica*, 123, 299–311.
- Bruine de Bruin, W., & Keren, G. (2003). Order effects in sequentially judged options due to the direction of comparison. *Organizational Behavior and Human Decision Processes*, 92(1–2), 91–101.
- Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1), 1–28.
- Butts, M. M., Lunt, D. C., Freling, T. L., & Gabriel, A. S. (2019). Helping one or helping many? A theoretical integration and meta-analytic review of the compassion fade literature. *Organizational Behavior and Human Decision Processes*, 151, 16–33.
- Cameron, C. D., & Payne, B. K. (2011). Escaping affect: How motivated emotion regulation creates insensitivity to mass suffering. *Journal of Personality and Social Psychology*, 100(1), 1–15. <https://doi.org/10.1037/a0021643>
- Cocker, F., & Joss, N. (2016). Compassion fatigue among healthcare, emergency and community service workers: A systematic review. *International Journal of Environmental Research and Public Health*, 13(6), 618. <https://doi.org/10.3390/ijerph13060618>
- Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York: Putnam.
- Danziger, S., Levav, J., & Avnaim-Pesso, L. (2011). Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences*, 108(17), 6889–6892.

- Englich, B., Mussweiler, T., & Strack, F. (2006). Playing dice with criminal sentences: The influence of irrelevant anchors on experts' judicial decision making. *Personality and Social Psychology Bulletin*, 32(2), 188–200. <https://doi.org/10.1177/0146167205282152>
- Fetherstonhaugh, D., Slovic, P., Johnson, S., & Friedrich, J. (1997). Insensitivity to the value of human life: A study of psychophysical numbing. *Journal of Risk and Uncertainty*, 14(3), 283–300.
- Figley, C. R. (Ed.) (1995). *Brunner/mazel psychological stress series, no. 23. Compassion fatigue: Coping with secondary traumatic stress disorder in those who treat the traumatized*. Philadelphia, PA, US: Brunner/Mazel.
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4), 25–42.
- Glöckner, A. (2016). The irrational hungry judge effect revisited: Simulations reveal that the magnitude of the effect is overestimated. *Judgment and Decision making*, 11(6), 601–610.
- Granhag, P. A., Strömwall, L. A., & Hartwig, M. (2005). Granting asylum or not? Migration board personnel's beliefs about deception. *Journal of Ethnic and Migration Studies*, 31(1), 29–50.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814–834. <https://doi.org/10.1037/0033-295x.108.4.814>
- Jaffe, P. G., Crooks, C. V., Dunford-Jackson, B. L., & Town, J. M. (2003). Vicarious trauma in judges: The personal challenge of dispensing justice. *Juvenile and Family Court Journal*, 54(4), 1–9.
- Jenni, K., & Loewenstein, G. (1997). Explaining the identifiable victim effect. *Journal of Risk and Uncertainty*, 14(3), 235–257.
- Keyesers, C., & Gazzola, V. (2014). Dissociating the ability and propensity for empathy. *Trends in Cognitive Sciences*, 18(4), 163–166. <https://doi.org/10.1016/j.tics.2013.12.011>
- Kirk, U., Harvey, A., & Montague, P. R. (2011). Domain expertise insulates against judgment bias by monetary favors through a modulation of ventromedial prefrontal cortex. *Proceedings of the National Academy of Sciences*, 108(25), 10332–10336.
- Kogut, T., & Ritov, I. (2005). The "identified victim" effect: An identified group, or just a single individual? *Journal of Behavioral Decision Making*, 18(3), 157–167.
- Lasalvia, A., Bonetto, C., Bertani, M., Bissoli, S., Cristofalo, D., Marrella, G., ... Marangon, V. (2009). Influence of perceived organisational factors on job burnout: Survey of community mental health staff. *The British Journal of Psychiatry*, 195(6), 537–544. <https://doi.org/10.1192/bjp.bp.108.060871>
- Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and decision making. *Annual Review of Psychology*, 66, 799–823. <https://doi.org/10.1146/annurev-psych-010213-115043>
- Li, Y. E., & Epley, N. (2009). When the best appears to be saved for last: Serial position effects on choice. *Journal of Behavioral Decision Making*, 22(4), 378–389.
- Lustig, S. L., Delucchi, K., Tennakoon, L., Kaul, B., Marks, D., & L. & Slavin, D. (2008). Burnout and stress among United States Immigration Judges. *Benders' Immigration Bulletin*, 13(2), 22–30.
- Lustig, S. L., Karnik, N., Delucchi, K., & Tennakoon, L. (2008). Inside the judges' chambers: Narrative responses from the National Association of Immigration Judges stress and burnout survey. *Georgetown Immigration Law Journal*, 23, 57–83.
- Markowitz, E. M., Slovic, P., Västfjäll, D., & Hodges, S. D. (2013). Compassion fade and the challenge of environmental conservation. *Judgment and Decision making*, 8(4), 397–406.
- Newell, J. M., & MacNeil, G. A. (2010). Professional burnout, vicarious trauma, secondary traumatic stress, and compassion fatigue. *Best Practices in Mental Health*, 6(2), 57–68.
- Page, L., & Page, K. (2010). Last shall be first: A field study of biases in sequential performance evaluation on the Idol series. *Journal of Economic Behavior & Organization*, 73(2), 186–198.
- Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J., & Fugelsang, J. A. (2015). On the reception and detection of pseudo-profound bullshit. *Judgment and Decision making*, 10, 549–563.
- Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015). Everyday consequences of analytic thinking. *Current Directions in Psychological Science*, 24(6), 425–432.
- Persson, E., Barrafreem, K., Meunier, A., & Tinghög, G. (2019). The effect of decision fatigue on surgeons' clinical decision making. *Health Economics*, 28, 1194–1203. <https://doi.org/10.1002/hec.3933>
- R Core Team. (2018). R: A language and environment for statistical computing. R Foundation for statistical computing, Vienna, Austria. URL: <https://www.R-project.org/>.
- Ramji-Nogales, J., Schoenholtz, A. I., & Schrag, P. G. (2009). *Refugee roulette: Disparities in asylum adjudication and proposals for reform*. New York: New York University Press.
- Ray, S. L., Wong, C., White, D., & Heaslip, K. (2013). Compassion satisfaction, compassion fatigue, work life conditions, and burnout among front-line mental health care professionals. *Traumatology*, 19(4), 255–267.
- Rossi, A., Cetrano, G., Pertile, R., Rabbi, L., Donisi, V., Grigoletti, L., ... Amadeo, F. (2012). Burnout, compassion fatigue, and compassion satisfaction among staff in community-based mental health services. *Psychiatry Research*, 200(2–3), 933–938. <https://doi.org/10.1016/j.psychres.2012.07.029>
- Saakvitne, K. W., Pearlman, L. A., & Abrahamson, D. J. (1996). *Transforming the pain: A workbook on vicarious traumatization*. New York: WW Norton.
- Sabo, B. M. (2006). Compassion fatigue and nursing work: Can we accurately capture the consequences of caring work? *International Journal of Nursing Practice*, 12(3), 136–142. <https://doi.org/10.1111/j.1440-172X.2006.00562.x>
- Slovic, P., Västfjäll, D., Erlandsson, A., & Gregory, R. (2017). Iconic photographs and the ebb and flow of empathic response to humanitarian disasters. *Proceedings of the National Academy of Sciences*, 114(4), 640–644.
- Smith, J. F., & Kida, T. (1991). Heuristics and biases: Expertise and task realism in auditing. *Psychological Bulletin*, 109(3), 472–489.
- Strandberg, T., Sivén, D., Hall, L., Johansson, P., & Pärnamets, P. (2018). False beliefs and confabulation can lead to lasting changes in political attitudes. *Journal of Experimental Psychology: General*, 147(9), 1382–1399. <https://doi.org/10.1037/xge0000489>
- Thompson, I., Amatea, E., & Thompson, E. (2014). Personal and contextual predictors of mental health counselors' compassion fatigue and burnout. *Journal of Mental Health Counseling*, 36(1), 58–77.
- United Nations Universal Declaration of Human Rights (1948), General Assembly Resolution 217A.
- Västfjäll, D., Slovic, P., Mayorga, M., & Peters, E. (2014). Compassion fade: Affect and charity are greatest for a single child in need. *PLoS ONE*, 9(6), e100115. <https://doi.org/10.1371/journal.pone.0100115>
- Wilson, V. E. (1977). Objectivity and effect of order of appearance in judging of synchronized swimming meets. *Perceptual and Motor Skills*, 44, 295–298.
- Zaki, J. (2014). Empathy: A motivated account. *Psychological Bulletin*, 140(6), 1608–1647.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Pärnamets P, Tagesson A, Wallin A. Inconsistencies in repeated refugee status decisions. *J Behav Dec Making*. 2020;1–10. <https://doi.org/10.1002/bdm.2176>