

Politicians, the Representativeness Heuristic and Decision-Making Biases

Sjoerd Stolwijk^a & Barbara Vis^b

^a NPO, Hilversum

^b Corresponding author: Utrecht University School of Governance / Utrecht University / b.vis@uu.nl /www.barbaravis.nl / ORCID ID: 0000-0003-2323-3862.

Pre-print of an article accepted for publication in Political Behavior

Abstract

Do politicians use the representativeness heuristic when making judgements, that is, when they appraise the likelihood or frequency of an outcome that is unknown or unknowable? Heuristics are cognitive shortcuts that facilitate judgements and decision making. Oftentimes, heuristics are useful, but they may also lead to systematic biases that can be detrimental for decision making in a representative democracy. Thus far, we lack experimental evidence on whether politicians use the representativeness heuristic.

To contribute to and extend the existing literature, we develop and conduct a survey experiment with as main participants Dutch elected local politicians from the larger municipalities (n=211). This survey experiment examines whether politician participants display two decision-making biases related to the representativeness heuristic: the conjunction error and scope neglect. We also run the experiment with a student sample (n=260), mainly to validate the experimental design.

Our findings show that politician participants neglect scope in one scenario and that they display the conjunction error in two of three scenarios. These results suggest that politician participants use the representativeness heuristic. Conversely, our third conjunction error

scenario does not find evidence for politician participants displaying this bias. As we discuss in the paper, the latter may be an artifact of our experimental design. Overall, our findings contribute fundamentally to our understanding of how politicians process information and how this influences their judgements and decision making.

Key words: Politicians; Representativeness heuristic; Judgement; Decision making

Replication files are available on the *Political Behavior* Dataverse website,

<https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/GAX8JV>.

1. Introduction

Politicians—such as members of parliament, ministers, or presidents—often have to make *judgements*, that is ‘appraisals about the likelihood, probability or frequency of an outcome which is unknown or unknowable’ (McDermott 2001: 7). Judgements typically take place prior to a decision, although they do not necessarily lead to one. They are assessments about external events and usually occur under conditions of uncertainty (McDermott 2001: 7). As Hammond (1996: 5) states, ‘(h)uman judgment is the hidden, mysterious link in the process that forms the policies and plans that directly affect, if not control, the nature of our society, as well as its interaction with other societies’. There is a large literature that shows that in making judgements, people regularly apply *heuristics*: cognitive shortcuts that facilitate judgement and decision making (see e.g., Gilovich, Griffin, and Kahneman 2002). Using heuristics may, however, also lead to systematic biases, that is deviations from comprehensive rationality or normative rational choice theory (Gilovich and Griffin 2002).

In this paper, we concentrate on a heuristic that is particularly relevant for political judgement and decision making: *representativeness*. The representativeness heuristic is one of the so-called general purpose heuristics from Kahneman and Tversky’s heuristics and biases tradition (for an overview, see Gilovich et al. 2002; for a discussion of this tradition and the so-called fast & frugal one, see Kelman 2011). The representativeness heuristic is at work when people judge probabilities ‘by the degree to which A is representative of B, that is, by the degree to which A resembles B’ (Tversky and Kahneman 1974: 1124). Oftentimes, representative events will be frequent events, but ‘they may better be thought of as ones that fit a stereotype or as diagnostic of group membership’ (Kelman 2011: 23). The second general purpose heuristic—that we do not focus on here, but that we also examined in our survey experiment—is *availability* (Tversky and Kahneman 1974). People employ the availability heuristic when they assess how likely it is that something occurs by focusing on the ease with which they can think

of instances or occurrences of it. The availability heuristic can lead to biases because salient and attention-grabbing events will more readily come to mind (such as political scandals) or because a dramatic event (like a plane crash) temporarily increases the availability of its category (Kahneman 2011). Both representativeness and availability thus answer the difficult question of probability with the answer to the simpler question of resemblance (Kahneman and Frederick 2002: 50). The difference between the two is that representativeness works through resemblance to for instance a prototype or stereotype, whereas availability works through resemblance to specific instances.

What are the effects of the representativeness heuristic on the functioning of democracy? Drawing on similarity between the instance at hand and a general category—which is what people do when using the representativeness heuristic—facilitates judgements and decision making. This makes it easier for politicians to do their job, which may be good for the functioning of democracy. However, drawing on such similarity may also lead to biases (Tversky and Kahneman 1974). Such biases occur, for example, when politicians judge a proposal’s likelihood of success by thinking only of successful proposals, thereby overestimating the proposal’s likely success. What is more, politicians who think negatively about whole groups of people—say, Muslims—because of the activities of some of them rely on the representativeness heuristic (Vis 2019). In this latter case, bias occurs because politicians may take policy measures that (negatively) affect the whole group, while only a few of them are in fact problematic. Since politicians’ judgements and subsequent decisions have far-reaching consequences for the lives of many people, it is important to investigate how politicians arrive at them. In this article, we therefore ask: Do politicians use the representativeness heuristic when making judgements and taking decisions? There are several observational studies suggesting that they do (Böhmelt, Ezrow, Lehrer and Ward 2016; Vis 2019; Weyland 2007, 2014, see below), but these studies do not offer a systematic experimental test. What is more, the

incentives for politicians to “get it right” are high, indicating that they may process information systematically instead of heuristically. There is evidence suggesting that experienced elites revise their heuristics more efficiently (Alevy, Haigh and List, 2007; Chi 1978; Feltovich, Prietula and Ericsson, 2006). To date, existing studies do not offer conclusive evidence on these largely conflicting expectations.

We fill this lacuna by developing and conducting a survey experiment with as main participants Dutch local elected politicians from the larger municipalities (between 100k and 230k inhabitants). We also run the experiment on a student sample,¹ primarily to validate our experimental design; as we discuss in more detail below, we do not expect any substantive differences between the politician and student participants. Like their national counterparts, politicians from the larger municipalities are selected competitively and make decisions that influence the lives of large numbers of citizens. Similar to many of their national counterparts, many local politicians operate under (severe) time pressure (Pommer and Boelhouwer 2017: 397). The Netherlands is a relevant context because powers have increasingly shifted from the national to the local level. For example, child protection, labor market policies, chronic health care and care for the elderly are decided upon by local politicians since 2015. The annual budget of a typical municipality included in our sample frame, Apeldoorn, amounts to about 600 million euros (Apeldoorn 2016), indicating that these politicians preside over significant public funds.

By studying the potential biases that result from the representativeness heuristic, we contribute to the growing literature on the role of biases in political decision making. Keller and Yang (2016), for example, show that initial problem representation shapes foreign policy

¹ Our student sample can be seen as drawn from the population of non-politicians. Since we are interested in the basic processes that may be true for all humans, using such a student sample makes sense (Druckman and Kam 2011; Morton and Williams 2010: chapter 9).

decision making processes. Since the representativeness heuristic yields the associations constituting such initial representations and given that option generation is also important in many other policy domains, Keller and Yang’s findings suggest how the representativeness heuristic may influence political decision making. Moreover, Baekgaard, Christensen, Dahlman Mondrup, Mathiasen and Grund Petersen (2019) show that politicians are just as likely as non-politicians to rely on motivated reasoning when facing new information. And Pedersen (2017) demonstrates how ratio bias—that is, people’s tendency to focus on numerators and pay insufficient attention to denominators—may influence citizens’ interpretations of numbers in political issues.

We complement and extend the existing literature by examining whether politicians display the biases related to the representativeness heuristic by adjusting existing scenarios to a contemporary and novel context. The representativeness heuristic has wide applicability to many domains of decision making, extending beyond the specific biases we investigate here (see e.g. Kahneman 2011; Kelman 2011). In addition, the representativeness heuristic’s foundation offers a concrete theoretical starting point for how to address these biases and potentially improve political decision making.² Overall, our findings thus contribute fundamentally to our understanding of how politicians process information and how this influences their judgements and decision making.

2. How do politicians process information?

Do politicians use the representativeness heuristic when processing information? The relatively scant existing empirical evidence typically suggests that *yes*, they do (Weyland 2007, 2014, see

² For instance Morewedge et al. (2015) show that a single “debiasing” training intervention—a video or a (serious) game—reduced bias from over-relying on representativeness immediately and three months later.

Vis, 2019). Weyland (2007), for example, demonstrated that key policy makers' use of the representativeness heuristic explains why so many countries in Latin America followed Chile's bold pension reform; they saw the Chilean success as representative of a larger set of reform successes (see also Weyland 2008, 2014). Conversely, in their quantitative analysis of how party policy positions diffuse across countries, Böhmelt et al. (2016) did not find support for the representativeness heuristic.³ Notwithstanding the latter finding, existing studies typically reveal that if politicians differ from ordinary citizens in relying on heuristics, this is mainly a difference in degree not in kind. Also politicians are "normal mortals" (Weyland 2014: 58, see p. 54) who resort to heuristic-processing when the environment is highly uncertain and complex (see also Jacobs 2011: 41-42). And uncertainty and complexity is precisely what characterizes many political judgements and decisions (Bursens, De Landtsheer, Braekmans and Segaelert 2017). Therefore, we do not expect any substantive differences between the politician and student participants.

Nonetheless, there are indications that politicians might process information differently, contra our expectation. Politicians are typically more experienced than are ordinary citizens, and experience usually results in expertise (Hafner-Burton, Hughes and Victor 2013). This could make politicians—including local ones—less susceptible to the biases related to the representativeness heuristic. However, experience in one specific domain does not automatically improve decision making in another domain, even if it is closely related (Hafner-Burton et al. 2013). Since the range of judgements and decisions that politicians typically have to make is wide, this might limit the effect of experience. Most politicians in our sample are no expert in

³ Note that Böhmelt et al. (2016) do not discuss what their empirical findings imply regarding the representativeness heuristic. Because the representativeness heuristic is included in their theorizing, this is—at a minimum—a missed opportunity. Against this backdrop, we consider it reasonable to interpret what the findings imply regarding the representativeness heuristic.

the specific domains of our survey. Tetlock (2005) furthermore demonstrated that people who are experts in making political judgement generally do not better—and oftentimes worse—than do non-experts. This means that it is unclear whether experience will lead politicians to rely less on the representativeness heuristic.

Another potential reason why politicians might differ in their reliance on heuristics is that the amount of information politicians receive is much larger than the amount of information ordinary citizens receive (Walgrave et al. 2013). The former typically have more resources available to process this information, for instance in the form of support staff. Much of the information politicians receive is also pre-structured by the institutional structures (Baumgartner and Jones 2015). Moreover, politicians' information processing has consequences for themselves (e.g., their political career) and for citizens, which may make them more motivated to use the information in the best way possible (Anderson and Harbridge 2014; Jacobs 2011; Jones 2001; Sheffer, Loewen, Soroka, Walgrave and Shaefer 2018; Weyland 2014). The upshot of this is that, at least under some conditions and contra our main expectation, politicians may not use the representativeness heuristic in their judgement and decision making. We extend the existing literature's findings by assessing these—partly—conflicting expectations directly and experimentally.

3. Experimental design and hypotheses

To assess whether politicians use the representativeness heuristic, we follow most extant work and Tversky and Kahneman's own approach and infer heuristics 'by measuring the biases that their use is supposed to evoke' (Bellur and Sundar 2014: 121).⁴ In the case of the

⁴ Bellur and Sundar (2014) present an approach to establish directly whether a heuristic has been used. An approach to quantify heuristic bias in clinical decision making is proposed by Richie and Josephson (2017).

representativeness heuristic, these biases are the conjunction error and scope neglect (discussed in more detail below). To this end, we adjust four existing tests so that they fit the contemporary Dutch context (see Online Appendix 1 for a description of our four-stage procedure, which included pre-registration of the larger project from which the data for this article is drawn;⁵ see Online Appendix 6 for an overview of the full set of hypotheses included in the pre-registration plan; see Online Appendix 3 for the text of the survey experiment). We invited the politician participants in their role as municipal council member (see Online Appendix 2a for the text of the invitation email).⁶ Most of our scenarios are closer to the judgements and decisions local politicians truly make than really abstract scenarios would be. Still, also these scenarios remain a simplification of reality and are not intended to be truly mundane (Iyengar 2011).

We also include a number of questions in our survey to capture the participants' background characteristics, particularly gender, level of education, expertise (measured by being a spokesperson on a for this survey experiment relevant area, like safety or public management), experience (measured by among other questions how long someone has been a local council member), self-reported math skills, and self-placement on a left-right scale. The relevant questions can be found in Online Appendix 3.⁷

Conjunction error I: The Linda/Vera problem

A first bias related to the representativeness heuristic is the *conjunction error*. People make a conjunction error when they consider the conjunction $A \wedge B$ —with the “ \wedge ” indicating logical-

⁵ We preregistered the design and our expectations for the full project at AsPredicted.org on April 4, 2017 (<https://aspredicted.org/5ye4c.pdf>). We uploaded a so-called add-on to our pre-registration at AsPredicted.org on May 9, 2017 (<https://aspredicted.org/pi28u.pdf>). More information is available in Online Appendix 1.

⁶ We invited the students in their role as citizens (see Online Appendix 2b for the text of the invitation email to the students).

⁷ Recall that our full survey also includes several questions intended to evaluate the use of the availability heuristic.

AND—more likely than, for instance, *A*. The latter is logically speaking at least as large, and probably larger, than the conjunction $A \wedge B$. To test whether the politician participants make a conjunction error, we adapt the classic test for this bias, the so-called Linda problem (Tversky and Kahneman 1983), to the contemporary Dutch context. In the original test, Linda is described as a strong-willed woman who was stereotypically related to someone active in the feminist movement. Based on this description, 85 per cent of Tversky and Kahneman's (1983) respondents judged it more likely that Linda was both a bank-teller *and* active in the feminist movement than only a bank-teller.⁸ As Lu (2016: 510) states, ‘the accumulated evidence (...) has suggested that the violation [i.e. the conjunction error] is highly robust to variations in response modes and is very easy to replicate in a variety of contexts’.

While the Linda problem differs from the type of judgements local politicians truly make, we include it as a benchmark for the other, more mundane, scenarios. If we find that politician participants display the conjunction error in this scenario, in which they are arguably least likely to respond in their role as politicians, this means that the sample is not *a priori* different from samples that have been used previously. What is more, since this test is often replicated, it also provides a benchmark for the student sample. If we would not find the conjunction error in this latter sample, this would call into question the student sample’s quality. In addition to this more methodological justification for including this scenario, the Linda-problem is also substantively relevant for politics. When making a conjunction error, people make a judgement that is logically incorrect by (over-)relying on a stereotype (for more on the relation between the representativeness heuristic and stereotypes, see Gilovich and Griffin 2002). This

⁸ Even though this percentage dropped, a majority still made a conjunction error when the test was made more transparent. Despite criticisms on the Linda problem (see Kahneman 2011: chapter 15), the evidence in favor of the conjunction error is still strong (Kahneman and Frederick 2002: 66, see also pp. 65-68 for a discussion of the critique).

means that the resulting decision may be sub-optimal from the standpoint of means–goal efficiency. A hypothetical, yet realistic, example is when politicians (or voters) consider it more likely that people who cause a nuisance in a municipality (such as drugs dealers, or some addicts or homeless people)⁹ have a low socioeconomic status *and* a migrant background ($A \wedge B$) than that they have a low socioeconomic status only (A). Politicians may then spend more resources to address $A \wedge B$ (or that voters ask of them to spend more resources on this), whereas A truly is the larger problem.¹⁰ Indeed, this example reveals that stereotypes are often at the forefront of many political issues (for instance ethnicity or gender). When such stereotypes trigger biases, this might lower the quality of the subsequent political decision, that is, this decision might not be in line with the politician’s stated goals and/or values (for examples on how stereotypes influence political decision making by voters, see Arceneaux 2008; Herrmann and Tepe 2017). To avoid any potential confusion with a popular Dutch TV-host named Linda, we used the name Vera. In line with earlier findings on ordinary citizens (see above), we expect that also politician participants make the conjunction error, indicating that they use the representativeness heuristic:

H1. Politician participants will generally judge the conjunction (“Vera is a bank employee and is active in the feminist movement”) more likely than one or both of its parts

⁹ For Dutch municipalities of the size we focus on (see Section 4), these are the type of people who cause a nuisance that Dutch local politicians face. Note that this, for instance, does not mean that all homeless people or addicts cause a nuisance.

¹⁰ We chose not to use this specific scenario to (a) stay close to Tversky and Kahneman’s (1983) original formulation of the problem and to (b) avoid a strong political connotation.

(“*Vera is a bank employee*” or “*Vera is active in the feminist movement*”).¹¹

Conjunction error II: Making the headlines scenario

A second scenario to assess whether the politician participants display biases related to the representativeness heuristic is the causal conjunction test (Tversky and Kahneman 1983). In the original scenario, different groups of participants were asked to rate the likelihood of a suspension of diplomatic relations between the USA and the USSR sometime in 1983, compared to the likelihood of that same suspension but then due to a Russian invasion of Poland. If people make a causal conjunction error, the adding of a probable cause (Russia’s invasion of Poland) to an unlikely event (the suspension of diplomatic relations) makes this event appear more likely. Note that adding a specific cause *decreases* the logical odds, since other causes are excluded. But a combination of cause and event is *more representative* of something that might happen than the event by itself.

Our adjusted scenario is closer to home for the local politician participants than diplomatic relations would be. One group of participants rates the likelihood of the event that their municipality makes the headlines of all major newspapers next year. The other group rates the same likelihood but will be giving an exact cause for this event: “due to a terrorist attack on King’s day”—a public holiday in the Netherlands at which all municipalities have activities in the public space. Adding this cause makes the “making the headlines” event more representative of something that might happen. In 2009, a terrorist attack happened in Apeldoorn on

¹¹ This hypothesis differs somewhat from the hypothesis as included in the pre-registration plan from May 2017 (see Online Appendix 1). Recall that Online Appendix 6 presents an overview of all hypotheses from this pre-registration and includes information on whether and if so why we diverted (somewhat) from a specific hypothesis. In this appendix, we also list those hypotheses that are part of the larger project and that are or will be discussed elsewhere.

Queen's day,¹² the then host of the Queen and her family. A terrorist drove into the crowd toward an open bus carrying the royal family. Eight people were killed because of the attack, putting a dark mark on this public holiday. This incident was covered by many major international news agencies. This was such a salient event, that most people who were of age in 2009 still remember it. For most of our local politicians, the scenario is thus—unfortunately—representative. It could also be representative for the (much) younger students, because every year around King's Day, the media discusses safety, sometimes referring to Apeldoorn 2009. Both politicians and students may also think about the attacks at Christmas markets in Berlin, Germany in 2016 and Straatsburg, France in 2018—both of which are relatively close to the Netherlands and have been widely covered in Dutch media. This leads to the causal conjunction error hypothesis (H2):

H2. Politician participants will judge it more likely that a terrorist attack will lead to their municipality making the headlines of all major newspapers, compared to making those headlines in general. ¹³

The causal conjunction error has important implications for politics. Most importantly, if at work, it privileges outcomes of conjunctive scenarios in which the causal arguments are apparent—such as chemical or biological warfare—over outcomes that emerge from causally less clear sequences of events—such as accidents or mistakes (McDermott 2001: 10-12; see also Norman and Delfin 2012).

Conjunction error III: The earthquake scenario

¹² In 2009, the Netherlands had a Queen.

¹³ See note 11.

Our third scenario also tests a variant of a conjunction error, in this case related to *scope neglect*, i.e. people's tendency to neglect a representative event's scope. The scenario is based on Tversky and Kahneman's (1983) California earthquake scenario in which one group of participants assesses the likelihood of an earthquake hitting California next year and causing a massive flood, while the other group assesses the likelihood of a massive flood somewhere in North America next year. Participants judged the, representative, image of the earthquake and flood in California as more likely than a flood somewhere in North America, even though California is part of North America, and floods can originate from different causes.

We adjusted this scenario into the earthquake scenario. The damage caused by earthquakes in Groningen (a province in the North of the Netherlands) is currently an important theme in Dutch politics. At the same time, natural disasters like earthquakes are generally rare in the Netherlands. We therefore expect that Dutch citizens—politicians and students alike—will view Groningen more representative of earthquakes than is the eastern half of the Netherlands as a whole of natural disasters. However, Groningen is geographically much smaller than is the much larger eastern half of the Netherlands. Comparing participants' assessment of the likelihood on an earthquake versus a natural disaster thus constitutes a test similar to the California earthquake scenario. If our participants neglect scope, which we expect them to, the following hypothesis (H3) holds:

H3. Politician participants will judge the likelihood of an earthquake in Groningen to be higher than of a natural disaster in the eastern half of the Netherlands.¹⁴

Also scope neglect has important political implications. If politicians consider a specific event

¹⁴ See note 11.

(like an earthquake hitting California and causing a massive flood) more likely than they would a more general event (flood in North America), it means that they concentrate on the more specific event and will probably be willing to spend more resources to avoid this specific event than they would to avoid the more general event. This could result in policy underreaction to the more general problem (Maor, 2014).

Scope neglect I & II: The nuisance scenario

Our fourth and final scenario also assesses whether politician participants neglect scope. Scope neglect may result from a conjunction error, but also from other applications of representative prototypes, as is the case in this scenario (see Kahneman and Frederick 2002). Scope neglect is thus slightly different from, but strongly related to the use of the representativeness heuristic. Politicians deal with scope both in their estimates of the likelihood of events (as in the previous scenario) as well as in their decisions on, for instance, budget allocation. Scope neglect may thus influence how much of the budget politicians are willing to spend on a problem. If politicians would neglect scope, this means that they either under- or overestimate the severity of problem. This, in turn, may result in policy underreaction (Maor, 2014) or policy overreaction (Maor, 2013). We adjust a scenario from a study that found that Toronto residents were willing to pay the same amount to clean up polluted lakes in a small region of Ontario as they were to clean up lakes throughout—the larger region of—Ontario (Kahneman and Knetsch 1992; see Desvouges, Mathews and Train 2012; Kahneman and Frederick 2002). This finding indicates that the problem's scope did not relate to their willingness to pay.

To examine whether politicians are sensitive to this type of scope neglect, we let one group of participants judge the importance of the problem of 23 people who cause a nuisance in their municipality, while the other group judges the importance of 53 people who cause a nuisance. To obtain the numbers of people causing a nuisance that in this context would be

qualitatively different, we consulted an expert on policing and Dutch municipalities who advised using 23 and 53 because these represent categorically different sizes (an average number of people who cause a nuisance for a municipality this size versus record numbers of people who cause a nuisance for a municipality this size). We restricted the politician sample to municipalities of a comparable size (see below). If participants—politicians and students alike—appraise the problem via representative prototypes, they think of how *important in general* they find it to deal with people who cause a nuisance, so whether they are asked about 23 or 53 such people should not matter. This yields the following hypothesis (H4):

*H4. Politician participants will generally consider it equally important to deal with 23 people who cause a nuisance as with 53 ones.*¹⁵

Moreover, and more directly comparable to Kahneman and Knetsch's (1992) original scenario, if participants neglect scope, they would allocate the same budget to a group of 23 and to a group of 53 people who cause a nuisance.

*H5. Politician participants will generally allocate an equal budget to deal with 23 people who cause a nuisance as they do to deal with 53 ones.*¹⁶

4. Data

Our sample frame of elected politicians¹⁷ from the 27 larger Dutch municipalities is 1,063 and

¹⁵ See note 11.

¹⁶ See note 11.

¹⁷ Since mayors in the Netherlands are appointed rather than elected, they are not included in our sample frame.

The same holds for aldermen.

excludes the four largest municipalities ($\approx 838k$, $\approx 631k$, $\approx 520k$ and $\approx 340k$ inhabitants). Note that we do not expect any differences across these largest four municipalities' politicians and the ones we focus on here in how they make judgements. However, politicians in the largest four municipalities probably face qualitatively different problems, which may influence their responses. We send out invitation emails first through each municipality's official channel (the "griffier"), followed by two reminders to each politician's official e-mail address. This resulted in 211 completed surveys (from 264 registered (partial) responses, $\approx 20\%$ response rate). Item non-response is low, as we show in Online Appendix 8.

Table S4 in Online Appendix 7 displays the politician sample's descriptives. This sample is representative of the full population of Dutch elected council members (all municipalities, $n=8,931$) in terms of age—average 52.7 years—and gender—28 per cent female. Our sample is somewhat more highly educated; in our sample, almost 90 per cent self-report holding an applied higher college or university degree compared to 67 per cent on average for the full population (Ministerie van Binnenlandse Zaken en Koninkrijksrelaties 2016). Also in terms of party membership, our sample is largely representative of the full population of council members (see Table S3 in Online Appendix 7). Only council members from local parties are somewhat underrepresented in our sample (18.5 per cent versus 27.8 per cent in the population). Randomization tests indicate that gender, age, municipality, party and education level were not significantly different between the different conditions (see Table S9 in Online Appendix 12).

Online Appendix 5 presents information on recruitment, response rates and characteristics of the student sample. Also for the student sample, item non-response is low (see Online Appendix 8). Many student participants live in one of the larger municipalities that we exclude in our sample of municipalities. To make the student participants' results as comparable as possible to the politician participants', we ask the student participants in the nuisance scenario to imagine being a municipal council member of Apeldoorn. In the making the headlines

scenario, we also ask the student participants to think of Apeldoorn, but not as a municipal council member (see Online Appendix 3 for the question wording in the survey).

5. Results

Conjunction error I: The Linda/Vera problem

Recall that we asked the participants to judge the likelihood that Vera was a bank employee, active in the feminist movement, or both. In line with the conjunction error hypothesis (H1), we found that 134 politicians (72%, n = 186)¹⁸ committed the conjunction error by ranking the likelihood of Vera being both active in the feminist movement and a bank employee to be higher than one of its constituents, see Figure 1. The results for the student sample were very similar to those of the politician sample (see Table S8 and Figure S1 in Online Appendix 11).

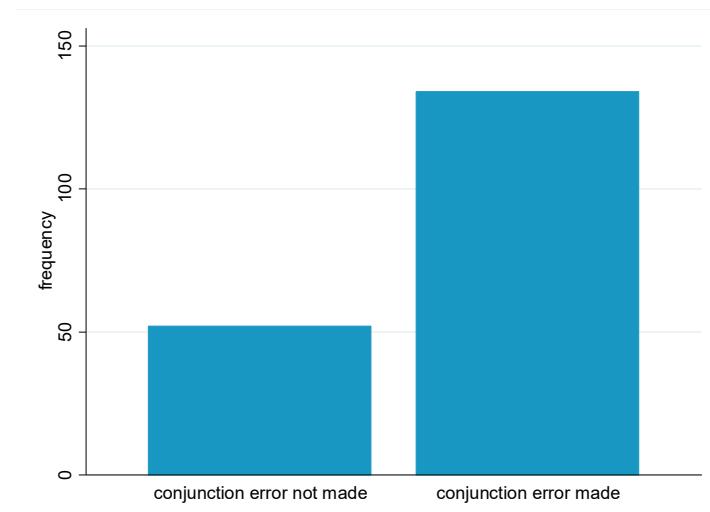


Figure 1. Frequency of politician participants committing the conjunction error (n = 186).

¹⁸ The number of respondents is lower than the full sample due to item non-response. Table S5 in Online Appendix 8 displays the non-response per item, which is low for all the substantively relevant items.

Table 1. Politician participants selection of each ordering for the Vera-problem.

Ordering	Frequency of politician participants choosing this ordering (%)
a. Most likely bank employee who is active in the feminist movement, than someone who is active in the feminist movement, than bank employee	23 (12%)
b. Most likely someone who is active in the feminist movement, than bank employee who is active in the feminist movement, than bank employee	56 (30%)
c. Most likely bank employee, than someone who is active in the feminist movement, than bank employee who is active in the feminist movement	24 (13%)
d. Most likely bank employee who is active in the feminist movement, than bank employee, than someone who is active in the feminist movement	14 (8%)
e. Most likely someone who is active in the feminist movement, than bank employee, than bank employee who is active in the feminist movement	28 (15%)
f. Most likely bank employee, than bank employee who is active in the feminist movement, than someone who is active in the feminist movement	41 (22%)
Total	186

Table 1 shows that 56 politician participants (30 per cent) choose order (b)—the stereotypical ordering of Vera being active in the feminist movement as most likely, followed by a bank employee who is active in the feminist movement, and her being a bank employee as least likely. The reverse ordering—order (f)—was chosen by 41 politician participants (22 per cent).¹⁹ The two orderings that do *not* imply the conjunction error—orderings c and e—, were chosen by 28 politician participants (15 per cent) and 24 (13 per cent), respectively. These results confirm H1 that politician participants make the conjunction error. Still, the findings also show that they did not implement stereotypical/representativeness reasoning in a 1:1 fashion: its reverse—which is still a conjunction error—was picked by almost a quarter of all participants.

¹⁹ Consequently, slightly less than half of all participants chose another ordering (45 per cent, NB: numbers do not add to 100 per cent due to rounding).

Conjunction error II: Making the headlines scenario

Interestingly, the results of the second conjunction error test—the making the headlines scenario—are quite different from those of the Linda/Vera problem. If politician participants make the conjunction error, we should find support for H2 that they will judge it more likely that a terrorist attack will lead to their municipality making the headlines of all major newspapers, compared to making those headlines in general. This is not what we find. Politician participants' likelihood estimates of the headlines (either due to a terrorist attack or in general) averaged to about 16 per cent (standard deviation $\pm 28\%$), spanning the full range of the scale from 0 to 100 per cent. The distribution of judged likelihood estimates is quite skewed to the left and contains quite some outliers (i.e. estimations over two standard deviations above the mean). In line with existing work (Lodge and Taber 2005; Wolsiefer, Westfall and Judd 2017), including willing-to-pay studies like ours (Tubeuf et al., 2015), we adjust for the skewed distribution by applying a log-transformation (see Online Appendix 14 for a more extensive discussion). After doing so, we find evidence to reject H2. As Figure 2 shows, politicians judged the likelihood of making the headlines in general on average considerably higher than making them due to a terrorist attack on King's day ($t=12.53$; $p<0.001$ single tailed, $N(\text{total})=195$: $n(\text{general})=107$, $n(\text{terrorism})=88$).²⁰ The results of this test obtained in the student sample are in the same direction as those on the politician sample (Figure S2, Online Appendix 10) suggesting that this might not be a valid causal conjunction error test.

²⁰ Adding control variables like gender, political experience, education, left-right self-placement, math-skill, or number of statistics courses followed has no influence on this result (see Table S9 in Online Appendix 13).

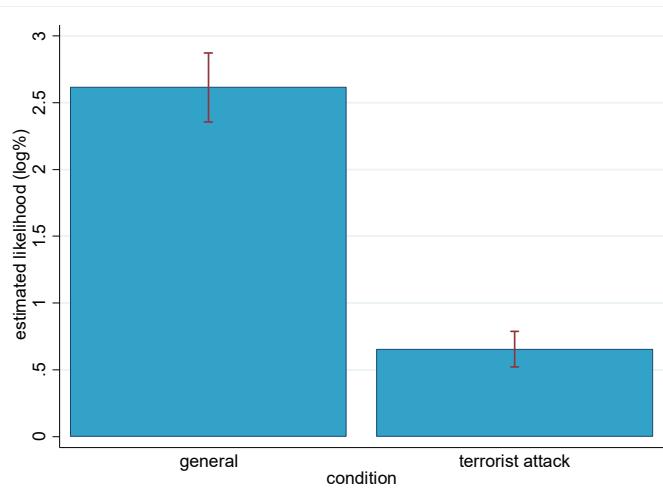


Figure 2. Estimated log-likelihood of a council member's municipality making the headlines in 2018 in general (on the left), versus due to a terrorist attack (on the right) ($n=195$ politician participants).

Conjunction error III: Earthquake scenario

In our third conjunction error test, we examined if participants neglect scope. If they do, we should find support for H3 that participants judge the likelihood of an earthquake in Groningen to be higher than that of a natural disaster in the Netherlands' eastern half. This is indeed what we find. The participants' estimates were quite skewed to the left, with a significant number of outliers at the higher end of the scale ($n=201$, $M=5.30$, $SD=10.22$, range [0; 89]). Therefore, we again apply a log-transformation (see Online Appendix 14 for a more extensive discussion). As Figure 3 shows, the politician participants judged the earthquake's likelihood to be higher than that of the natural disaster ($N(\text{total})=201$: $n(\text{earthquake})=103$, $n(\text{natural disaster})=98$, $t=1.69$, $p=0.05$ single tailed). On average, the politician participants thus neglected the larger scope of the East of the Netherlands compared to the province of Groningen (which it contains) as well as the larger class of events of natural disasters versus earthquakes (which it again contains). This result suggests that the politician participants display this conjunction error. The results based on the student sample match those presented here (see Figure S3 in Online Appendix 10), supporting the validity of this test.

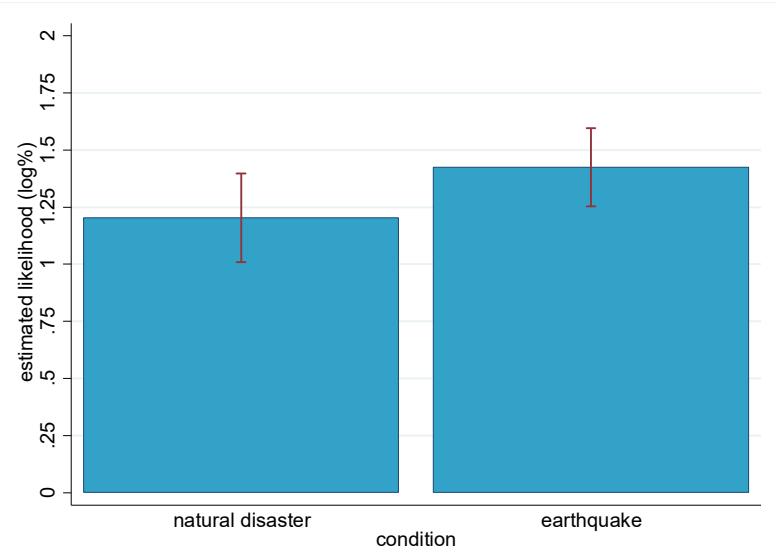


Figure 3. Estimated log-likelihood of dozens of people getting injured in 2018 due to a natural disaster in the eastern half of the Netherlands (on the left), versus due to an earthquake in Groningen (on the right) ($n=201$ politician participants).

Scope neglect I & II: Nuisance scenario

Turning to the nuisance scenario, we find support for H4 that politician participants will generally consider it equally important to deal with 23 people who cause a nuisance (4.88, $n=105$) as they do to deal with 53 ones (4.75, $n=106$) (see Figure 4). This indicates that the politician participants neglect scope, confirming H4. Note that given that the scale runs from [1-6], this corresponds to in-between “important” and “very important” ($t=-1.15$, $p=0.87$ single tailed in the direction that 53 people who cause a nuisance are more important than 23 ones).

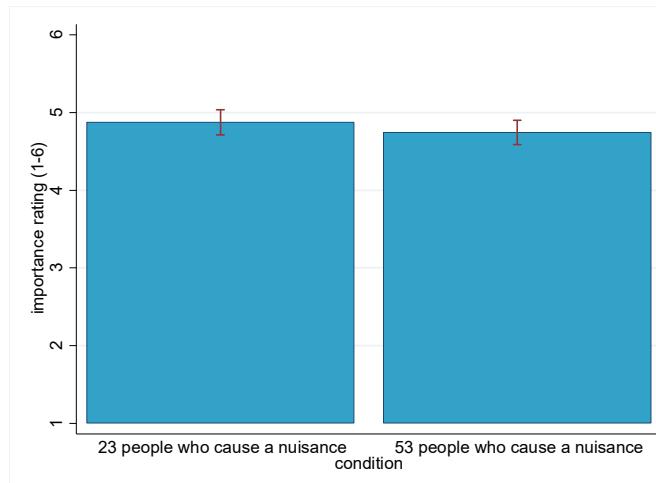


Figure 4. Importance of dealing with 23 people who cause a nuisance (on the left) versus 53 ones (on the right) in a council member's municipality on a scale of 1 “not important at all” to 6 “utmost important” ($n=211$ politician participants).

When looking at the budgets allocated in each condition, we find further support for politicians neglecting scope. Figure 5 shows that there was no difference between the budgets that politician participants allocated to 23 who cause a nuisance and 53 ones when adjusting for the disproportional influence of outliers through a log-transformation ($N(\text{total})=210$: $n(23 \text{ people who cause a nuisance})=105$, $n(53)=105$, $t=0.02$, $p=0.49$ single tailed in the direction that 53 people who cause a nuisance receive a larger budget than 23 ones). This confirms H5 that politician participants will generally allocate an equal budget to deal with 23 people who cause a nuisance than they do to deal with 53 ones. Interestingly, the difference in importance (H4) and budget allocated (H5) is significant in the student sample (see Table S7 & Figure S4, Online Appendix 10). A possible explanation for the difference between the politician participants and the student participants is that the former were asked about *their own* municipality whereas the latter were asked about Apeldoorn.²¹ However, the difference in findings across the two samples is

²¹ The choice of city—Apeldoorn, which witnessed a terrorist attack in 2009—could have influenced the results of the nuisance scenario. However, this was not the case. If student participants would have been primed with

substantively not that large, since the difference between the student sample and the politician sample itself is not significant.

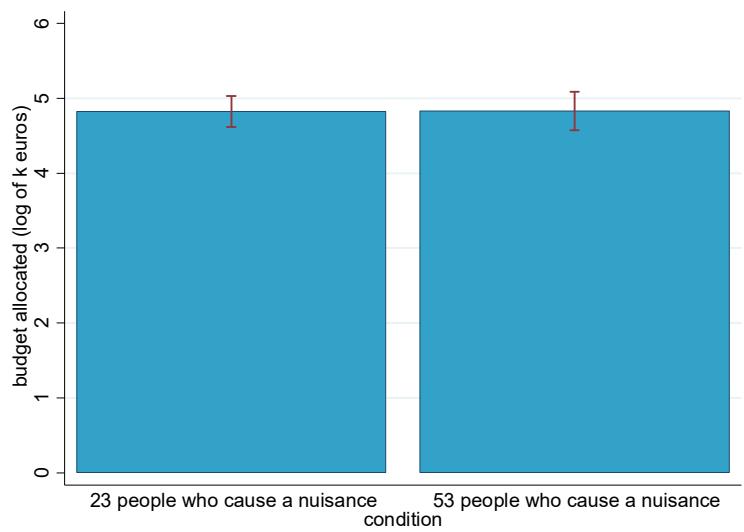


Figure 5. Budget allocated to deal with 23 people who cause a nuisance (on the left) versus 53 ones (on the right) in a council member's municipality ($n=210$ politician participants).

Additional and robustness analyses

We conducted several additional and robustness analyses. First, as an additional analysis, we examined if politician participants thought that the difference in scope (23 versus 53 people who cause a nuisance) should be reflected in the allocated budget (Online Appendix 9). We find that they thought it should, which contradicts the between-subjects' findings we report

danger, this would have increased the importance of safety, which should have made them willing to spend *more* on people who cause a nuisance. However, we find that they are willing to spend less to address this problem. Also in the making the headlines scenario, the choice of city could have influenced our findings. If the King's day attack in Apeldoorn would have been "available" to the student participants, the availability heuristic would predict them to find this attack to be *more* likely, thus bolstering our expectation from H2. Since we do not find that students believed this to be more likely (see Table S7 in Online Appendix 10), there is little reason to suspect that our choice of city influenced the results we obtained in this scenario.

above. We return to the implications of this finding in the discussion.

As an exploratory analysis, we examined whether left wing politicians will rank the conjunction (“ $A \wedge B$ ”—see above) as more probable for Vera compared to her being active in the feminist movement (“ B ”), while right wing politicians will rank the conjunction (“ $A \wedge B$ ”) as more probable for Vera compared to her being a bank teller (“ A ”). This hypothesis was part of our pre-registration plan (see Online Appendices 1 & 6). The results of this analysis, which Online Appendix 11 reports, are contradictory. Specifically, right wing politicians were more likely to choose the reverse stereotypical ordering, but left wing politicians were less likely to make the conjunction error. This finding is difficult to interpret and suggests that further research is needed.

We also conducted several robustness analyses. To examine whether politicians might be less prone to display the biases related to the representativeness heuristic because of specific traits, we first ran a regression analysis with the following control variables added: gender, education level, political experience, math skill, right-left self-placement.²² The results in Table S10 in Online Appendix 12 are similar to the results presented in the main text. Furthermore, we ran four additional analyses to explore whether politicians with different levels of experience (Table S11 in Online Appendix 12), expertise (Table S12), education level (Table S13) and self-reported math skill (Table S14) are sensitive to displaying the biases related to the representativeness heuristic. For none of these variables did we find a significant interaction effect, suggesting that politician participants who differ on these traits were not more or less sensitive to

²² Age could not be included as a control variable because of too many missing data.

the difference between the conditions in the various scenarios.²³

6. Discussion and conclusion

Are politicians' judgements, which shape their decisions, influenced by the representativeness heuristic? In this article, we presented a new survey questionnaire to answer this question. Our results demonstrate that in most scenarios politicians—in this case: elected local politicians from the larger Dutch municipalities—displayed the biases related to the representativeness heuristic. Specifically, the politician participants made the conjunction error in two scenarios and neglected scope in one scenario. Conversely, we did not find support for the conjunction error in a third scenario. The latter finding conflicts with existing findings on a similar scenario. We also failed to find support for the conjunction error in this scenario in the student sample (see Online Appendix 10), which suggests that there might be an issue with the validity of this item. This could be a relevant avenue for further research.

Our results are likely generalizable to “real” political decisions, where deliberation and counter-arguing is possible. In our scenarios, our politician participants made an individual judgement, while in practice deliberation with advisors and other politicians—that is, group decision making—may help to dissipate the biases before a final decision is made (Finucane, Alhakami, Slovic and Johnson 2000; Kaplan and Miller 1978). Still, the positive effect of deliberation requires the absence of strong time pressure (Bang and Frith 2017: 6), whereas politicians—including local ones—typically operate under severe time pressure (Pommer and Boelhouwer 2017; Goetz 2014). Moreover, even without time pressure, these biases will not disappear when a group decision is made by people who are similar. Given that most individuals display biases, they appear similar in this respect. We confirmed the presence of these biases

²³ As we explain in Online Appendix 13 in more detail, while the sample size for experience is too small to draw firm conclusions, the results still provide first evidence that experience has no significant effect.

for most of our politician participants, not just for specific individuals. Furthermore, if a group polarizes—which regularly happens with group decision making—the bias at the individual level may even been amplified at the group level (Bang and Frith 2017: 10). Group decision making thus does not necessarily resolve biases (Kerr, MacCoun and Kramer 1996). It would be an interesting avenue for future research to examine the extent to and conditions under which political deliberation and polarized debate may contribute to or hinder bias reduction.

The elected local council members who were the main participants in our survey experiment are part time rather than full time politicians: on average, they spend about two or three days a week on their work as a council member. This means that they are less “elite” than politicians at the national level are. Still, our results are probably generalizable to the national level, as well as to politicians from other countries. For one, we replicated Tversky and Kahneman’s findings (1983) which were based on a US sample in the 1980s in another country (the Netherlands) in another period (2017), illustrating their generic character. In addition, our survey experiment included a scenario on framing and risk preferences that was fielded previously among a sample of Dutch national-level politicians (members of parliamentarian, MPs), thereby allowing us to compare local and national politicians directly.²⁴ We found similar results in both cases (Linde and Vis 2017), suggesting that our results generalize to the national level. Moreover, our finding for Dutch MPs on this item was similar to those of Sheffer et al. (2018) for Belgian and Canadian MPs, indicating that our results based on samples from the Netherlands are likely generalizable to these other systems.

We conducted various additional analyses to better understand what drives our results. Together, our results support the conclusion that politicians display the biases related to the representativeness heuristic, but they do not necessarily imply that all politicians are equally

²⁴ Note that we did not discuss this item elsewhere in this article, because, although it is related, it concerned the availability heuristic and not the representativeness one that we focus on here.

likely to display the biases under all circumstances. Further research might explore whether there are certain politicians who are in general more likely to base their judgements on heuristics like representativeness and, if so, which characteristics increase or decrease that probability. Are there different “types” of politicians, with some more likely to use System 1 and others more likely to use System 2 (Kahneman 2011)? While the sample size does not allow us to draw strong conclusions here, we conducted a preliminary analysis to answer this question. Specifically, we examined whether we could predict a politician participant’s answer in a scenario based on whether s/he made the conjunction error in the Vera scenario. If we could, this would offer preliminary evidence that these politician participants systematically used System 1. Our results—see Table S15 in Online Appendix 13—show, conversely, that this is not the case; there is no systematic relationship between making the conjunction error in the Vera scenario and biases in other scenarios.

Another line of research might explore the contextual circumstances that increase or decrease the probability of politicians to use heuristic reasoning. In our additional analyses, reported in the results section in the main text, we found that politician participants indicated themselves that the difference in scope (23 versus 53 people who cause a nuisance) *should be* reflected in the allocated budget (see also Online Appendix 9). Our main findings, however, showed that when asked between subjects, the difference in scope did not influence the allocated budget. Since both findings are based on the same set of politicians, it appears that the context under which an issue is evaluated plays an important role. This finding suggests that politicians in general would like to use System 2, while in practice they oftentimes use System 1 instead. Learning more about the conditions under which the preference for using System 2 is stronger than the inclination to go with System 1 would also be a worthy topic to explore.

Summing up, we presented the results of a newly developed survey experiment to evaluate the influence of the representativeness heuristic on political judgement, and thereby to

better understand how politicians process information. Our finding that politicians display the conjunction error and neglect scope indicate that they use the representativeness heuristic, which has important implications for political judgement and decision making. For one, the responses to our Vera-question illustrate how politicians bias their likelihood assessment when confronted with a stereotype. Note that they did so despite being aware of the stereotypical nature of the question.²⁵ So even though stereotypes are often at the forefront of political debates (e.g. discrimination, gender) and politicians can thus be expected to deal with them carefully, they were still influenced by such stereotypes in non-obvious ways. Biased likelihood assessments related to stereotypes might support political decision making, whether pro or counter, which re-affirms that stereotype by giving stereotypical groups more (policy) attention than warranted (cf. Forbes and Schmader 2010).

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the research committee of the Faculty of Social Sciences of the Vrije Universiteit Amsterdam and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Acknowledgements The research project on which this article is based was funded by a VIDI grant awarded to Barbara Vis from the Netherlands Organization for Scientific Research (NWO, grant nr. 452-11-005). Earlier drafts of this article have been presented at the 2017 NIG Work Conference, Maastricht, the Netherlands; the 2017 General Conference in Oslo, Norway; at the 2017 ECPR Joint Sessions of Workshops in Nottingham, the UK; the 2017 IMEBESS conference in Barcelona, Spain; the 2017 Politicologenetaal conference in Leiden, the

²⁵ See also Online Appendix 1.

Netherlands; and the 2017 ISPP annual meeting, Edinburgh, Scotland. We thank all participants of these events for their valuable comments and suggestions. We also thank the four anonymous reviewers and the editor of *Political Behavior* for their useful comments.

References

- Alevy, J. E., Haigh, M. S., & List, J. A. (2007). Information Cascades: Evidence from a Field Experiment with Financial Market Professionals. *The Journal of Finance*, 62(1), 151–180.
- Anderson, S. E., & Harbridge, L. (2014). The Policy Consequences of Motivated Information Processing Among the Partisan Elite. *American Politics Research*, 42(4), 700–728.
- Apeldoorn, G. (2016). *Doorwerken: Meerjaren Programma Begroting 2016-2019*. Apeldoorn: Gemeente Apeldoorn.
- Arceneaux, K. (2008). Can Partisan Cues Diminish Democratic Accountability? *Political Behavior*, 30(2), 139–160.
- Baekgaard, M., Christensen, J., Dahlman Mondrup, C., Mathiasen, A., & Grund Petersen, N. B. (2019). The Role of Evidence in Politics: Motivated Reasoning and Persuasion among Politicians. *British Journal of Political Science*, 49(3), 1117-1140.
- Bang, D., & Frith, C. D. (2017). Making Better Decisions in Groups. *Royal Society Open Science*, 4(170193), 1–22.
- Baumgartner, F. R., & Jones, B. D. (2015). *The Politics of Information: Problem Definition and the Course of Public Policy in America*. Chicago and London: The University of Chicago Press.
- Bellur, S., & Sundar, S. S. (2014). How Can We Tell When a Heuristic Has Been Used? Design and Analysis Strategies for Capturing the Operation of Heuristics. *Communication Methods and Measures*, 8(2), 116–137.

- Böhmelt, T., Ezrow, L., Lehrer, R., & Ward, H. (2016). Party Policy Diffusion. *American Political Science Review*, 110(2), 1–14.
- Bursens, P., De Landtsheer, C., Braekmans, L., & Segaert, B. (eds. . (2017). *Complex Political Decision-Making: Leadership, Legitimacy and Communication*. London: Routledge.
- Chi, M. (1978). Knowledge Structures and Memory Development. In R. Siegler (Ed.), *In Children's Thinking: What Develops?* (pp. 73–96). Hillsdale: Erlbaum.
- Desvouges, W., Mathews, K., & Train, K. (2012). Adequate Responsiveness to Scope in Contingent Valuation. *Ecological Economics*, 84(December), 121–128.
- Druckman, J. N., & Kam, C. D. (2011). Students as Experimental Participants: A Defense of the “Narrow Data Base.” In J. N. Druckman, D. P. Green, J. H. Kuklinski, & A. Lupia (Eds.), *Cambridge Handbook of Experimental Political Science* (pp. 41–57). New York: Cambridge University Press.
- Evert Pommer, & Boelhouwer, J. (eds. . (2017). *Overall Rapportage Sociaal Domein 2016: Burgers (De)centraal*. Den Haag: Sociaal en Cultureel Planbureau.
- Feltovich, P. J., Prietula, M. J., & Ericsson, K. A. (2006). Studies of Expertise from Psychological Perspectives. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge Handbook of Expertise and Expert Performance* (pp. 41–68). Cambridge: Cambridge University Press.
- Finucane, M. L., Alhakami, A., Slovic, P., & Johnson, S. M. (2000). The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making*, 13(1), 1–17.
- Forbes, C. E., & Schmader, T. (2010). Retraining Attitudes and Stereotypes to Affect Motivation and Cognitive Capacity Under Stereotype Threat. *Journal of Personality and Social Psychology*, 99(5), 740–754.
- Gilovich, T., & Griffin, D. (2002). Introduction - Heuristics and Biases: Then and Now. In T.

- Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and Biases: The Psychology of Intuitive Judgment* (pp. 1–18). New York: Cambridge University Press.
- Gilovich, T., Griffin, D., & Kahneman, D. (eds). (2002). *Heuristics and Biases: The Psychology of Intuitive Judgment*. New York: Cambridge University Press.
- Goetz, K. H. (2014). A Question of Time: Responsive and Responsible Democratic Politics. *West European Politics*, 37(2), 379–399.
- Hafner-Burton, E. M., Alex Hughes, D., & Victor, D. G. (2013). The cognitive revolution and the political psychology of elite decision making. *Perspectives on Politics*, 11(2), 368–386.
- Hammond, K. R. (1996). *Human Judgment and Social Policy: Irreducible Uncertainty, Inevitable Error, Unavoidable Injustice*. Oxford etc.: Oxford University Press.
- Herrmann, M., & Tepe, M. (2018). Does Exposure to Stereotype-Disconfirming Politicians Reduce the Effect of Stereotypes on Voting? Evidence From Seven Plagiarism Scandals in Germany. *Political Psychology*, 39(2), 303–324.
- Iyengar, S. (2011). Laboratory Experiments in Political Science. In J. N. Druckman, D. P. Green, J. H. Kuklinski, & A. Lupia (Eds.), *Cambridge Handbook of Experimental Political Science* (pp. 73–88). New York: Cambrdige University Press.
- Jacobs, A. M. (2011). *Governing for the Long Term: Democracy and the Politics of Investment*. New York: Cambridge University Press.
- Jones, B. D. (2001). *Politics and the Architecture of Choice: Bounded Rationality and Governance*. Chicago and London: The University of Chicago Press.
- Kahneman, D. (2011). *Thinking, Fast and Slow*. London: Penguin Books.
- Kahneman, D., & Frederick, S. (2002). Representativeness Revisited: Attribute Substitution in Intuitive Judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and Biases: The Psychology of Intuitive Judgment* (pp. 49–81). New York: Cambridge

University Press.

- Kahneman, D., & Knetsch, J. L. (1992). Valuing Public Goods: The Purchase of Moral Satisfaction. *Journal of Environmental Economics and Management*, 22(1), 57–70.
- Kaplan, M. F., & Miller, L. E. (1978). Reducing the Effects of Juror Biases. *Journal of Personality and Social Psychology*, 36(12), 1443–1455.
- Keller, J., & Yang, Y. E. (2016). Problem Representation, Option Generation, and Poliheuristic Theory: An Experimental Analysis. *Political Psychology*, 37(5), 739–752.
- Kelman, M. (2011). *The Heuristics Debate*. Oxford: Oxford University Press.
- Kerr, N. L., MacCoun, R. J., & Kramer, G. P. (1996). Bias in judgment: Comparing individuals and groups. *Psychological Review*, 103(4), 687–719.
- Ministerie van Binnenlandse Zaken en Koninkrijksrelaties (2016). *Staat van het Bestuur*.
- Linde, J., & Vis, B. (2017). Do Politicians Take Risks Like the Rest of Us? An Experimental Test of Prospect Theory under MPs. *Political Psychology*, 38(1), 101–117.
- Lodge, M., & Taber, C. S. (2005). The Automaticity of Affect for Political Leaders, Groups, and Issues: An Experimental Test of the Hot Cognition Hypothesis. *Political Power and Social Theory*, 26(3), 455–482.
- Lu, Y. (2016). The Conjunction and Disjunction Fallacies: Explanations of the Linda Problem by the Equate-to-Differentiate Model. *Integrative Psychological and Behavioral Science*, 50(3), 507–531.
- Maor, M. (2013). Policy Overreaction. *Journal of Public Policy*, 32(3), 231–259.
- Maor, M. (2014). Policy Persistence, Risk Estimation and Policy Underreaction. *Policy Sciences*, 47(4), 425–4443.
- McDermott, R. (2001). The Psychological Ideas of Amos Tversky and Their Relevance for Political Science. *Journal of Theoretical Politics*, 13(1), 5–33.
- Morewedge, C. K., Yoon, H., Scopelliti, I., Symborski, C. W., Korris, J. H., & Kassam, K. S.

- (2015). Debiasing Decisions: Improved Decision Making With a Single Training Intervention. *Policy Insights from the Behavioural and Brain Sciences*, 2(1), 129–140.
- Morton, R. B., & Williams, K. C. (2010). *Experimental Political Science and the Study of Causality: From Nature to the Lab*. Cambridge: Cambridge University Press.
- Norman, E. R., & Delfin, R. (2012). Wizards under Uncertainty: Cognitive Biases, Threat Assessment, and Misjudgments in Policy Making. *Politics and Policy*, 40(3), 369–402.
- Pedersen, R. T. (2017). Ratio Bias and Policy Preferences: How Equivalency Framing of Numbers Can Affect Attitudes. *Political Psychology*, 38(6), 1103–1120.
- Richie, M., & Josephson, S. A. (2018). Quantifying Heuristic Bias: Anchoring, Availability, and Representativeness. *Teaching and Learning in Medicine*, 30(1), 67–75.
- Sheffer, L., Loewen, P. J., Soroka, S., Walgrave, S., & Shaefer, T. (2018). Nonrepresentative Representatives: An Experimental Study of the Decision Making of Elected Politicians. *American Political Science Review*, 112(2), 302–321.
- Tetlock, P. E. (2017). *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton University Press.
- Tubeuf, S., Willis, T. A., Potrata, B., Grant, H., Allsop, M. J., Ahmed, M., ... McKibbin, M. (2015). Willingness to Pay for genetic testing for inherited retinal disease. *European Journal of Human Genetics*, 23(3), 285–291.
- Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157), 1124–1131.
- Vis, B. (2019). Heuristics and Political Elites' Judgment and Decision Making. *Political Studies Review*, 17(1), 41–52.
- Walgrave, S., Epping, L., Sevenan, J., Vos, D., Van Camp, K., Andre, A., ... Dejaeghere, Y. (2013). *Het Informatiemenu van politici en hun medewerkers: Resultaten van Interviews met Federale Parlementsleden, Ministers, Partijvoorzitters en hun Persoonlijke*

- Medewerkers*. Antwerp: University of Antwerp.
- Weyland, K. (2007). *Bounded Rationality and Policy Diffusion: Social Sector Reform in Latin America*. Princeton/ Oxford: Princeton University Press.
- Weyland, K. (2008). Towards a New Theory of Institutional Change. *World Politics*, 60(2), 281–314.
- Weyland, K. (2014). *Making Waves: Democratic Contention in Europe and Latin America since the Revolutions of 1848*. New York: Cambridge University Press.
- Wolsiefer, K., Westfall, J., & Judd, C. M. (2017). Modeling Stimulus Variation in Three Common Implicit Attitude Tasks. *Behavior Research Methods*, 49(4), 1193–1209.

TABLE OF CONTENTS ONLINE APPENDICES

ONLINE APPENDIX 1	Four-stage procedure for developing the survey experiment
ONLINE APPENDIX 2A	Invitation letter municipal council members
ONLINE APPENDIX 2B	Invitation letter students
ONLINE APPENDIX 3	Question wording questionnaire
ONLINE APPENDIX 4	Information on the pre-testing
ONLINE APPENDIX 5	Recruitment, response rates, and characteristics of the student sample
ONLINE APPENDIX 6	Overview of hypotheses included in the pre-registration
ONLINE APPENDIX 7	Sample distribution of party membership politician participants
ONLINE APPENDIX 8	Item non-response
ONLINE APPENDIX 9	Additional analysis nuisance scenario
ONLINE APPENDIX 10	Results for student participants
ONLINE APPENDIX 11	Results for exploratory analysis left-right orientation
ONLINE APPENDIX 12	Influence of randomization, left-right self-placement, gender, experience, expertise, education and math skill
ONLINE APPENDIX 13	Are some politicians more likely to use the representativeness heuristic across scenarios?
ONLINE APPENDIX 14	Results without log transformations

ONLINE APPENDIX 1: FOUR-STAGE PROCEDURE FOR DEVELOPING THE SURVEY EXPERIMENT

To develop our survey experiment, we used a four-stage procedure. First, we pre-registered a first version of the design and expectations for the full project,²⁶ and obtained approval from the Faculty’s research ethics’ committee. Since we wanted to make sure that our tests yielded the interpretations/associations in line with the original tests, we tested and improved the first version of the survey experiment using successive cognitive interviews (Willis 2004). The latter are qualitative “thick” interviews on a pilot sample that enable examining respondents’ thinking in response to our questions. The cognitive interviews were conducted with about 20 faculty members from the University of Amsterdam and the Vrije Universiteit Amsterdam, specialized in political science and communication science respectively; the peer pre-testers (PPTs). We asked several PPTs to participate in our survey experiment and then questioned them one by one, usually in person and in some cases per email. We asked how the survey could be improved in general, what their thoughts were while answering our questions, and how they had interpreted the intentions of our questions. We then explained our intentions to them and discussed ways to improve the survey experiment. The resulting, improved, version of the survey experiment was then presented to the next (couple of) faculty member(s), again going through the same process. This allowed us to test out multiple versions of our questions.²⁷ In Online Appendix S4, we discuss the pre-testing of several of the scenarios in more detail. To reflect our updated expectations in view of the responses of the PPTs, we uploaded what is called an “add-on” to our initial pre-registration.²⁸ This add-on includes the hypotheses of our full project (see also Online Appendix 6) and listed our new estimations of the final sample sizes needed based on the power calculations using the pre-test data. Information on whether data had already been

²⁶ We preregistered the design and our expectations for the full project at AsPredicted.org on April 4, 2017 (<https://aspredicted.org/5ye4c.pdf>).

²⁷ We particularly would like to thank Ronald van Steden for his advice on the number of people who cause a nuisance to use in this scenario, which were meant to represent categorically different sizes (relatively many people who cause a nuisance for a municipality this size, versus record numbers of people who cause a nuisance for a municipality this size). Note that we restricted the sample of municipal councilmembers to include only those from municipalities of a comparable size (see main text).

²⁸ We uploaded the add-on to our pre-registration at AsPredicted.org on May 9, 2017 (<https://aspredicted.org/pi28u.pdf>).

collected, the study's main question and key dependent variables is available in the initial pre-registration document.

Second, we fielded the improved survey among a pre-test student-sample to evaluate whether their answers were in line with previous research and to assess whether our answer scales were sensitive enough to grasp the variation to achieve enough power for our survey experiment.

Third, we used the peer feedback in response to these pre-test results to construct the final survey experiment.²⁹ Fourth and finally, we fielded the survey.

²⁹ We presented our results at the 2017 ECPR Joint Sessions of Workshops in Nottingham, the UK; at the 2017 IMEBESS conference in Barcelona, Spain and send our report to several colleagues. We like to thank everyone for their constructive comments.

ONLINE APPENDIX 2: INVITATION LETTERS

ONLINE APPENDIX 2A. Invitation letter municipal council members, translated to English; the original invitation letter is in Dutch

Subject: Invitation for participation in study on making judgements

Dear madam, sir X (NB: the invitations were personalized),

At the Vrije Universiteit Amsterdam, a large research project is currently running that is financed by the Netherlands Organisation for Scientific Research. I—Barbara Vis, Professor of Political Decision Making—am the principal investigator of this project. One of the questions in this project is how elected representatives, such as you as a municipal council member, make judgements. And whether they do this similar to or different from citizens who are not elected representatives. Thus far, little is known about these questions in scientific research. By answering these questions by means of this study, we expand our knowledge of the decisions taken by municipal council members. This knowledge is not only relevant scientifically but will also generate useful information for you as an elected representative. Therefore, we will be happy to share the results of this study with you.

Hereby we warmly invite you, as a council member of one the (moderately) large municipality, to participate in our study. Filling in the online questionnaire will take maximum 10 minutes of your time. Participation is possible until XXX.

You can participate via think link: XXX.

Of course, participation in this study is fully on voluntary basis. The answers will be treated confidentially, whereby we safeguard anonymity. The results will be processed and analyzed at an aggregate level and will not be linkable to the individual level. At the website of this research project, we discuss this information in some more detail.

In case of any further questions and/or comments, you can contact me at all time. I am usually easiest to research by email [XXX], but by phone is of course also possible [XXX].

I very much hope that you are willing to cooperate in our great project, for which already much thanks.

Yours sincerely,

Prof. dr. Barbara Vis & Dr. Sjoerd Stolwijk

ONLINE APPENDIX S2B. Invitation letter students, translated to English; the original invitation letter is in Dutch (note that we used three slightly different invitation emails to cater to the specific student populations. The emails did not differ substantively).

Subject: Invitation for participation in study on making judgements

Dear students,

In the day-to-day governance of Dutch municipalities, many decisions are taken of which the exact consequences are unknown. Elected representatives, the local council members, thus need to make judgements, and we want to find out how they do this.

Do elected representatives make different judgements than “normal” citizens like students because of their experience? To test this, we have prepared a series of questions in which a judgement needs to be made about a range of subjects. You would help our study enormously if you could spend 10 minutes of your time on this study. This enables us to compare your questions with those of the elected representatives. Please note that the questions are about judgements, which means that there are no right or wrong answers. The people who have already participated liked the survey!

So please go to XXX [link to survey].

You can participate until XXX.

Thank you in advance!

Best wishes,

Prof. dr. Barbara Vis & Dr. Sjoerd Stolwijk

ONLINE APPENDIX 3: QUESTION WORDING QUESTIONNAIRE, TRANSLATED TO ENGLISH—the original questionnaire is in Dutch

Opening questions:

In which year were you born?

What is your highest level of education completed?

Are you currently a municipal council member?

How long have you served as a municipal council member in your current municipality?

Have you worked for a municipality before the start of your municipal council-membership?

For example, as a council member, or in another position at a different municipality, or your current municipality.

How many years have you worked at a municipality preceding your current municipal council-membership?

On which of the following policy areas do you spend most time?

How much time do you spend in an average week on your work as a council member?

How often do you check the news?

Making the headlines scenario

What do you think is the likelihood that [your municipality/for students: Apeldoorn] will be on the front pages of almost all major newspapers [next year (in 2018) / by a terrorist attack on King's Day next year (in 2018)]?

Drag the slider to the correct percentage.

(also choose 25 per cent if you believe the odds are larger than 25 per cent)

Potential follow up question:

How likely do you think this is on a scale from 0 to 100 per cent?

Earthquake problem

What is the probability that, in the coming year (in 2018) an [earthquake / natural disaster] will occur in [Groningen / the eastern half of the Netherlands (everywhere to the east of Utrecht)] where dozens of people get injured? Give the answer that comes closest to your estimate.

Drag the slider to the correct percentage.

(Also select 25 per cent if you believe the chances are larger than 25 per cent.)

[this question displayed a slider ranging from 0 to 25 per cent]

[if 25 per cent is selected another slider appeared ranging from 0 to 100 per cent]

The Linda/Vera problem

Following are a few scenarios asking you to indicate what you think is most likely or how much you estimate a certain chance.

Vera is 31 years old and single. She is extravert and smart. As a student she was opposed to globalization, she had dreadlocks and often listened to reggae music. She occasionally smoked a joint.

What does Vera do in daily life?

Drag the following options in order of least likely (1) most likely (3)

- 1 Vera is active in the feminist movement.
- 2 Vera works for a bank and is active in the feminist movement.
- 3 Vera works for a bank.

[options appear in a random order]

Nuisance scenario

Here are some questions about policy choices:

For politicians: [Your municipality] / For students: [Apeldoorn] / belongs to the approximately 50 Dutch municipalities with > 100,000 inhabitants. [For students: Imagine being a municipal council member of the municipality of Apeldoorn.]

There is a group of [23/53] people who cause a nuisance in your municipality (e.g. drug dealers, some addicts and / or homeless people).

How important do you think it is to address this problem?

- not important at all
- not important
- not important, but not unimportant either
- important
- very important
- utmost important

Follow up question on nuisance scenario

How much can the municipality spend to address this problem if it were [23/53] instead of [53/23] people (for example: drug dealers, some addicts and / or homeless people) who are causing a nuisance. (**Indicate your answer in thousands of euros**, also select 500 thousand if you believe the municipality should be able to spend more than 500 thousand)

You previously indicated that your municipality should be able to spend [X] thousand euros to [53/23] people who cause a nuisance.

Potential follow up question:

How much can the municipality spend to address this problem if it were [23/53] instead of [53/23] people (for example: drug dealers, some addicts and / or homeless people) who are causing a nuisance, **on a scale of 0 to 3 million euros?**

Closing questions:

Are you a man or a woman? [choice options: man / woman / different, i.e.:]

Where would you locate yourself on a political left-right scale?

Would you be willing to tell us in which municipality you are a council member?

Would you be willing to tell us for which party you are a municipal council member?

How many courses in statistics did you follow during your education?

If you had to give yourself a mark (scale from 1 [very bad] to 10 [excellent]), how good do you believe you are in mathematics?

Which of the following examples do you consider to be a natural disaster? (multiple answers possible)

Can you indicate who filled out this questionnaire? [choice options: the council member him/herself / a co-worker of the council member / different, i.e.:]

ONLINE APPENDIX 4: INFORMATION ON THE PRE-TESTING

The Linda/Vera problem

We pre-tested—see Online Appendix 1—the Vera-problem with about 20 academic colleagues (our peer pre-testers [PPTs]), asking them about their response to this question. Some of them complained that they felt a little uncomfortable because this question asked them too much about stereotypes, which we believe supports our confidence in our description of Vera.

Making the headlines scenario

Since our pre-test student-sample was drawn from a considerable larger municipality—Amsterdam ($\approx 838k$ inhabitants)—than the municipalities that are part of our politician-sample (between $\approx 225k$ and ≈ 100 inhabitants), we found in our pre-tests that they rated the odds of making the world's headlines as reasonably large. The latter suggests that the representative cause for this news attention is this large city's importance rather than any incidents like a terrorist attack. Since this would induce an unwanted difference between our student sample and politician sample in the form of municipality importance, we decided to ask our student participants to imagine they are from a specific municipality similar to those of the politicians (Apeldoorn, $\approx 160k$ inhabitants).

The nuisance scenario

We would like to thank Ronald van Steden for his advice on the number of people who cause a nuisance to use in this scenario, which were meant to represent categorically different sizes (an average number of people who cause a nuisance for a municipality this size, versus record numbers of people who cause a nuisance for a municipality this size). Note that we restricted the sample of municipal council members to include only those from municipalities of a comparable size.

ONLINE APPENDIX 5: RECRUITMENT, RESPONSE RATES, AND CHARACTERISTICS OF THE STUDENT SAMPLE

Recruitment and response rates

We recruited the student sample from the three large research universities in the Netherlands. Bachelor students in communication science at the University of Amsterdam could participate in return for 0.5 student participation credits; one invitation was sent out via BlackBoard ($n[\text{completed}] = 120$, from $n[\text{returned}] = 127$, from $n[\text{sample frame}] = 475$, response rate = ≈ 25 per cent). Bachelor thesis students in political science and public administration at the Vrije Universiteit Amsterdam could participate on a voluntary basis; one invitation and two reminders were sent out via BlackBoard ($n[\text{completed}] = 37$, $n[\text{returned}] = 40$, $n[\text{sample frame}] = 230$, response rate = ≈ 16 per cent). And students from Utrecht University could participate on a voluntary basis; one invitation was sent out via email ($n[\text{completed}] = 103$, $n[\text{returned}] = 105$, $n[\text{sample frame}] = 590$, response rate = ≈ 17.5 per cent). Together this results in a student sample of 260 students.

Table S1: Descriptives student sample.

Variable	n	Mean	S.D.
Woman	261	59 %	n.a.
Age	209	22.04	3.37
Study year	262	2.25	1.04
# statistics courses	261	1.94	1.55
Left-right political self-placement [scale: 1-7] ^a	261	4.36	1.24
Interest in politics [scale: 1(no interest)-7(very interested)]	264	5.28	1.31

^a Higher scores indicate more leftwing.

ONLINE APPENDIX 6: OVERVIEW OF HYPOTHESES INCLUDED IN THE PRE-REGISTRATION PLAN

Table S2. *Overview of hypotheses included in the pre-registration plan*

Pre-registration plan	Included in the main text?	Explanation
H1. Both politicians and students will generally rank the conjunction (“Vera is a bank teller and is active in the feminist movement”) as more likely for Vera than one of its parts (“Vera is a bank teller” or “Vera is active in the feminist movement”).	In the main text, we report the findings for the politician participants; in Online Appendix 10, we report those for the student participants.	Before collecting our data, we expected that the comparison between politicians and students would be meaningful and useful. However, after having collected the data, we observed differences across the two groups that could be attributed to a variety of factors. We found, for instance, large variation in the distribution of budgets allocated by the politician participants and student ones, which could be due to differences in background knowledge of the two groups, but also by factors like age or political experience. The large base-line differences across the two groups made a direct comparison less relevant than we had initially thought. We therefore decided to write a separate paper that delves more deeply in the differences between these groups and to zoom in on the politician participants in this paper.
H2. Left wing politicians will rank the conjunction (“A & B”) as more probable for Vera compared to her being active in the feminist movement (“B”), while right wing politicians will rank the conjunction (“A & B”) as more probable for Vera compared to her being a bank teller (“A”).	The results of this exploratory analysis are reported in Online Appendix 12.	
H3. Both politicians and students will generally believe it is more likely that an earthquake will hit Groningen than that a natural disaster will hit the eastern half of the Netherlands.	In the main text, we report the findings for the politician participants; in Online Appendix 10, we report those for the student participants.	See H1.
H4. Both politicians and students will generally believe it is more likely that a terrorist attack will lead to their municipality making the world’s headlines, compared to	Note: In the paper, this is H2. Based on the results of the pre-test and the feedback of our peer pre-testers, we believed the formulation “making the world’s	See H1.

making the world's headlines in general.	<p>headlines" was too strict a formulation, so we adapted the questionnaire to refer to "all major newspapers" instead. Unfortunately, the pre-test formulation ended up in the preregistration.</p> <p>In the main text, we report the findings for the politician participants; in Online Appendix 10, we report those for the student participants.</p>	
H5. Politicians will give more similar estimates to making the headlines in general and to making the headlines due to a terrorist attack, compared to students.	We did not include this test.	H5 was meant as a further specification to H4. Since H4 was rejected for both the politician participants and student ones, H5 could no longer be tested in accordance with the idea behind it.
H6. Both politicians and students will generally believe it is equally important to deal with 23 people who cause a nuisance as it is to deal with 53 ones. ^a	Note: In the paper, this is H4.	See H1.
H7. Both politicians and students will generally allocate an equal budget to deal with 23 people who cause a nuisance as they do to deal with 53 ones, when asked between subjects. ^a	<p>Note: In the paper, this is H5. We removed "when asked between subjects" from this hypothesis, but did test between subjects.</p> <p>In the main text, we report the findings for the politician participants and refer to those of the student participants; in Online Appendix 10, we report those for the student participants.</p>	See H1 for why we concentrate on the politician participants' findings.
H8. Both politicians and students will generally allocate a smaller budget to dealing with 23 people who cause a nuisance as they do to dealing with 53 ones, when asked in a direct test. ^a	The results of this additional analysis are reported in Online Appendix 9.	
H9. Both politicians and students will be more likely to prefer "more city council watchmen and encouragement of voluntary neighborhood watchmen" over "more police" when these options are contrasted with "more city council watchmen", compared to when they are not.	The results of this analysis will be reported elsewhere.	
H10. Both politicians and students will generally prefer the sure amount when choosing among gains, while preferring the lottery when choosing among losses.	The results of this analysis will be reported elsewhere.	

H11. Both politicians and students will be more likely to prefer “more city council watchmen and encouragement of voluntary neighborhood watchmen” as a policy response to people who cause a nuisance when first asked about the amount of budget allocated, compared to when the budget questions follow the policy question. ^a	The results of this analysis will be reported elsewhere.	
H12. Both politicians and students will generally allocate a larger budget to deal with 53 people who cause a nuisance than to deal with 23 ones if they are asked about the budget per person who causes a nuisance first, rather than if they are asked about the total budget first. ^a	The results of this analysis will be reported elsewhere.	
RQ1. Do politicians with more experience in office will be more or less likely to allocate a larger budget to deal with 53 people who cause a nuisance than to deal with 23 ones, regardless of whether or not they are asked about the budget per person who causes a nuisance first? ^a	The results of this analysis will be reported elsewhere.	
H13. Both politicians and students will generally allocate a different budget to deal with people who cause a nuisance after being asked about their preferred policy, compared to before they are asked about their preferred policy. ^a	The results of this analysis will be reported elsewhere.	

^a In the pre-registration, “people who cause a nuisance” are referred to as “troublemakers”. Because this latter term proved to be problematic by not capturing the original Dutch term and by being unnecessarily insulting to some, we have changed it.

In addition to these hypotheses, we also tested for the effect of the possible moderators listed in the pre-registration plan (education, knowledge of statistics, self-rated math skill, and finance as a politicians main focus area), but—as we report in the paper (see note 20 in the main paper and the results in Table S9 in Online Appendix 13)—did not find a significant effect for any of them.

The pre-registration plan also included two secondary analyses. The first of these related to an additional hypothesis, namely: Politicians will act more in line with rational choice for scenario's more alike to everyday political decision making, but not for abstract decision tasks, compared to a student sample. When formulating this expectation, we still planned to include more truly abstract tests, such as the letter

task that is often used in studies on the representativeness heuristic. However, based on the input from the cognitive interviews with our peer pre-testers (see Online Appendix 1), we decided to remove the abstract tasks and stick to the tests that are at least somewhat related to political judgment and decision making. This means that it would have been cleaner if we would have removed the secondary expectation “We specifically expect that politicians will act more in line with rational choice for scenario’s more alike to everyday political decision making, but not for abstract decision tasks, compared to a student sample” from the pre-registration add-on of May 2017.

We present the results of the second secondary analysis from the pre-registration plan—that left-wing politicians to make use of positive discrimination when asked to decide on a stereotype eliciting scenario, while right-wing politicians do not—in Online Appendix 11. These results are also discussed briefly as an exploratory analysis in the main text.

ONLINE APPENDIX 7: SAMPLE DISTRIBUTION OF PARTY MEMBERSHIP POLITICIAN PARTICIPANTS

Table S3. *Party membership of Dutch municipal council members (population) versus the sample used in this study.*

Variable	Population	Sample	Population	Sample
	%	%	#	#
CDA	14.4	17.9	26.6	33
ChristenUnie / SGP	7.3	4.4	13.4	8
D66	12.1	16.3	22.2	30
GroenLinks	5.4	7.0	9.9	11
Local parties	27.8	18.5	51.1	34
PvdA	10.3	12.0	18.9	22
SP	6.6	8.7	12.1	16
VVD	12.2	15.8	22.5	29
Partij voor de Dieren	0.5	0.5	1	1
Other party	3.5	0	6.5	0

Notes. Source: Nederlandse Kiesraad, <https://www.verkiezingsuitslagen.nl/verkiezingen/detail/GR20140319> (accessed December 2017). Population (#) represents the number of council members per party if our sample would have been a perfect representation of the distribution in the population. Sample (#) represents the actual number of council members per party in our sample. Percentages are rounded to one decimal.

The χ^2 -test for systematic differences between the sample and the population is not significant ($\chi^2 = 22.43$, df = 183, p = 1.00), although local party politicians appear to be under-represented in the sample.

Table S4: *Descriptives politician sample.*

Variable	n	Mean	S.D.	Min	Max
Woman	210	31%	N.A.	0	1
Age	87	50.31	12.37	26	73
Experience	169	5.97	4.67	0	30
Right-left self-placement (left)	211	4.25	1.33	2	7

Note. N.A.=not applicable since woman is a dichotomous variable.

ONLINE APPENDIX 8: ITEM NON-RESPONSE

Item non-response among politician participants

Table S5: Non-response per item.

Variable	Non-	Sample	Non-
	response (n)	(n)	response (%)
Party	27	211	12.8
Gender	1	211	0.5
Age	87	211	41.2
Education	0	211	0
Vera-question	25	211	11.9
Likelihood headlines general	2	109	1.8
Likelihood headlines terrorism	14	102	13.7
Likelihood headlines both conditions	16	211	7.6
Likelihood natural disaster	7	110	6.4
Likelihood earthquake	3	101	3.0
Likelihood natural disaster/earthquake (both conditions)	10	211	4.7
Importance of the nuisance issue	0	211	0
Total budget allocated to deal with the number of people who cause a nuisance as specified in the condition (either 23 or 53)	1	211	0.5
Total budget allocated to deal with the number of people who cause a nuisance in the other condition (i.e. 23 rather than 53)	4	211	1.9

Note: Percentages rounded to one decimal.

Overall, item non-response is low. The highest item non-response is found for age. We phrased the age, gender and party questions with a view to avoid eliciting anonymity-concerns amongst our participants, and we did not require a response.

The sample of 211 municipal council members is not always exactly equally divided across conditions. This is because 264 people started out with the survey questionnaire, but only 211 answered the question “Are you a municipal council member” confirmatively both at the onset and end of the questionnaire. The analysis in this paper is based on these 211 council members.

Item non-response among student participants

Table S6. Item non-response student sample.

Variable	Non-response (n)	Sample (n)	Non-response (%)
Gender	5	266	1.9
Age	57	266	21.4
Vera-question	6	266	2.3
Likelihood headlines general	3	121	2.5
Likelihood headlines terrorism	15	140	10.7
Likelihood headlines both conditions	23	266	8.7
Likelihood natural disaster	1	131	0.8
Likelihood earthquake	2	133	1.5
Likelihood natural disaster/earthquake (both conditions)	5	266	1.9
Importance of the nuisance issue	3	266	1.1
Total budget allocated to deal with the number of people who cause a nuisance as specified in the condition (either 23 or 53)	8	266	3.0
Total budget allocated to deal with the number of people who cause a nuisance in the other condition (i.e. 23 rather than 53)	7	266	2.6

Note: Percentages rounded to one decimal.

Overall, item non-response for the student-sample is low. The highest item non-response is found for age. We phrased the age, gender and party questions with a view to avoid eliciting anonymity-concerns amongst our participants, and we did not require a response.

The sample of 266 students is not always exactly equally divided across conditions. This is because 272 people started our student-questionnaire, but 266 answered the question “Are you a student?” confirmatively. The analysis in this paper is based on these 266 students. In addition, some students dropped out during the questionnaire, so were not allocated to a specific condition for (some of) the tests. Therefore, the non-response for separate conditions does not always add to the non-response per test.

ONLINE APPENDIX 9. ADDITIONAL ANALYSIS NUISANCE SCENARIO

Our tests of the nuisance scenario in the main text are between-subject. As an additional test, we conducted a within subjects test. This latter test allows us to assess if participants thought the difference in scope (23 versus 53) *should be* to be relevant for the budget allocated. If they would, participants who are asked about the 23 nuisance-budget *after* being asked about the 53 nuisance-budget would indicate that the former budget should be smaller (and vice versa).

The results show that politician participants indeed thought that the difference in scope of 23 versus 53 people who cause a nuisance should be relevant for the budget allocated. Those participants who were initially asked about the budget for 53 people who cause a nuisance, subsequently indicated that they would spend a smaller amount had there be only 23 ones (budget log-transformed to deal with outliers: n=103, t=5.45, p<0.0001). Likewise, politician participants who were initially asked about the budget for 23 people who cause a nuisance, subsequently indicated that they would spend a larger amount had there be 53 ones (budget log-transformed to deal with outliers: n=104, t=-6.53, p<0.0001). This means that politician participants think that the budgets to deal with 23 and with 53 people who cause a nuisance should be different, indicating that the problem's scope should matter. Yet, between-subjects, this is not what we find: the politician participants neglected the problem's scope. In the discussion, we reflect on the implications of this finding.

ONLINE APPENDIX 10. RESULTS FOR STUDENT PARTICIPANTS

Table S7. *Results for each test for the student participants.*

Hypothe-sis	Test	Result
H1	Conjunction error I: The Linda/Vera problem	Like the politician participants, most of the student participants committed the conjunction error: 204 students (78%) committed the conjunction error versus 56 (22%) who did not (n[students]=260).
H2	Conjunction error II: The making the headlines scenario	Like the politician participants, the student participants estimated the likelihood of making the newspaper in general larger than making the headlines due to a terrorist attack (log transformed likelihood, $t=3.41$, $p<0.001$, n(students)=243).
H3	Conjunction error III: The earthquake scenario	Like the politician participants, the student-participants estimated the likelihood of an earthquake higher than a natural disaster (log transformed likelihood, $t=2.49$, $p=0.007$ single tailed, n(students)=261).
H4	Scope neglect I: Nuisance scenario (importance)	In contrast to the politician participants, the student participants on average believed the problem of 53 people who cause a nuisance to be slightly more important than when asked the problem of 23 ones (4.20 vs 4.40 on a 1-6 scale, $t=1.68$, $p=0.05$, n[students]=263).
H5	Scope neglect II: Nuisance scenario (total budget)	In contrast to the politician participants, the student participants on average allocated a larger budget to deal with 53 than with 23 people who cause a nuisance (log transformed likelihood, $t=3.83$, $p<0.001$ single tailed, n[students]=258).

Figure S1. Number of students committing the conjunction error ($N=260$ students).

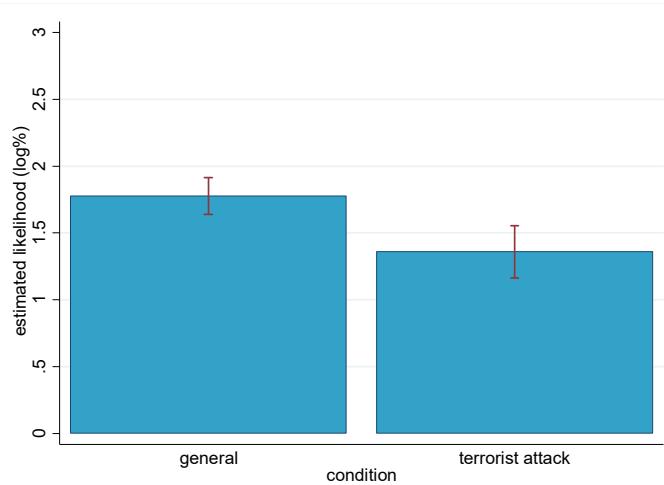


Figure S2. Log of estimated likelihood of Apeldoorn making the headlines in 2018: In general, versus due to a terrorist attack ($N=243$ students).

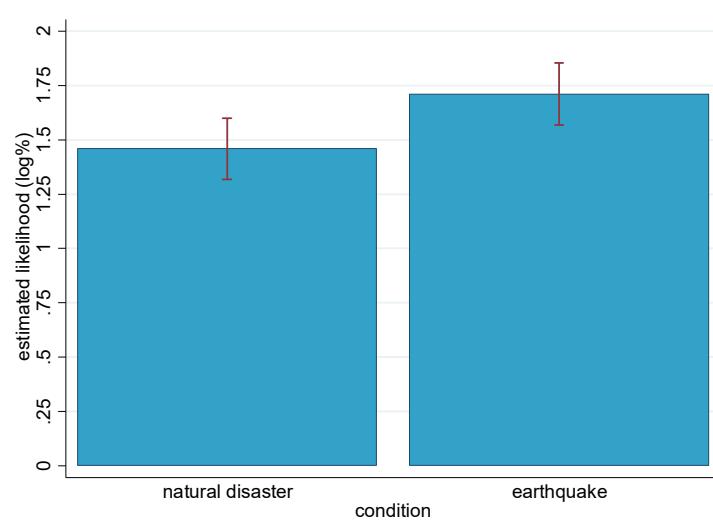


Figure S3. Log of estimated likelihood of dozens of people getting injured in 2018: Due to a natural disaster east of Utrecht, versus due to an earthquake in Groningen ($N=261$ students).

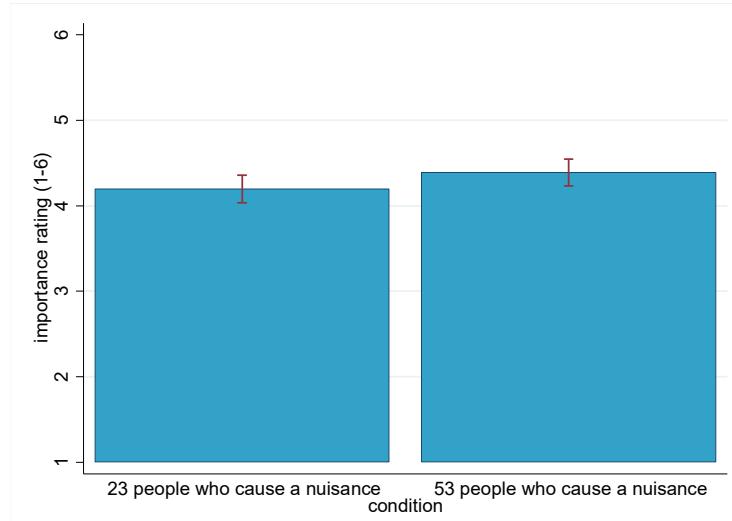
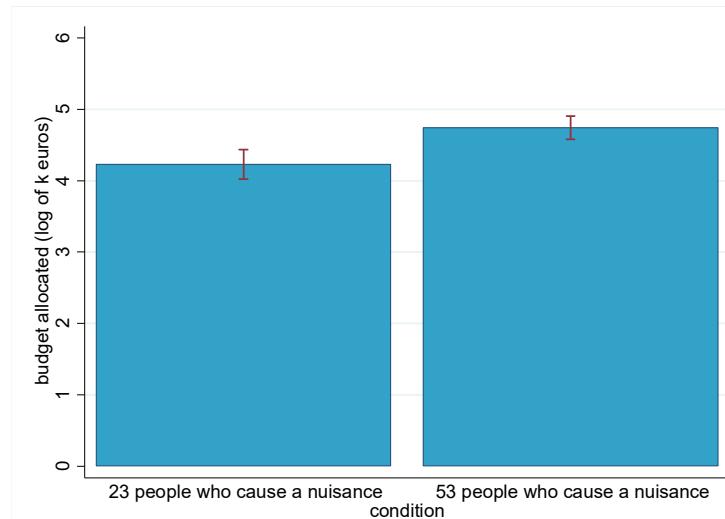


Figure S4. Judged importance of dealing with 23 versus 53 people who cause a nuisance in Apeldoorn on a scale of 1 “not important at all” to 6 “utmost important” ($n=263$ students).



ONLINE APPENDIX 11. RESULTS EXPLORATORY ANALYSIS LEFT-RIGHT ORIENTATION

This online appendix reports the results on an exploratory analysis of the effect of left-right orientation that was included in the pre-registration plan of May 2017 (see Online Appendix 1). Specifically, H2 of this plan read: Left wing politicians will rank the conjunction (“A & B”) as more probable for Vera compared to her being active in the feminist movement (“B”), while right wing politicians will rank the conjunction (“A & B”) as more probable for Vera compared to her being a bank teller (“A”). Note that in the fielded survey experiment, we changed the phrasing of “being a bank teller” into “works for a bank” because the former no longer is a regular type of job.

The spirit of H2 of the pre-registration was to evaluate whether left wing politicians were less likely than other politicians to make the conjunction error *because* their political inclination might lead them to “overcompensate”: they might recognize the stereotypical nature of the Linda-question, and respond with positive discrimination, i.e. by indicating it more likely that Linda is a bank employee who is active in the feminist movement (“A & B”) than active in the feminist movement (“B”) (ordering *a*, *d* or *e* from Table 1 in the main text). On the other hand, right wing politicians might not be inclined to apply positive discrimination and therefore opt more often than other politicians for a stereotypical ordering, i.e. by indicating it more likely that Linda is a bank teller who is active in the feminist movement (“A & B”) than a bank teller (“A”) (ordering *a*, *b* or *d* from Table 1 in the main text).³⁰

We define politicians as “right wing” if they self-rated themselves as 1, 2 or 3 on our right-to-left 7-point scale and as “left wing” if they rated themselves as 5, 6 or 7 on that scale.

We conducted a series of logistic regressions to test these expectations, the results are shown in Table S8 below. The first two columns of Table S8 show that indeed, in the spirit of H2 of the pre-registration, left wing politicians are less likely to make the conjunction error in the Linda scenario, and right wing politicians are more likely to make it. However, the results of the exact tests of H2 of the pre-registration plan—shown in columns 3 to 6—do not support the explanation for this effect as stated in the hypothesis. Left wing politicians were not more likely to opt

³⁰ Note that both expectations differ only in including ordering *e* for the reverse stereotypical choice and ordering *b* for the stereotypical choice.

for a reverse stereotypical ordering (and right wing politicians were *more* rather than less likely to do so).³¹ And right wing politicians were not more likely to opt for a stereotypical ordering (and left wing politicians were not less likely to do so). We thus do not find support for our hypothesis. We do find some support that left wing politicians are less likely to commit the conjunction error. The findings thus contradict each other: right wing politicians were more likely to choose the reverse stereotypical ordering, but left wing politicians were less likely to make the conjunction error. These findings are difficult to interpret and suggest that further research is needed.

Table S8. *Logistic regressions predicting expectation orders for left wing and right wing politicians in the Linda/Vera scenario.*

	conjunction	conjunction	rank the conjunction ("A & B") as more probable for Vera com- pared to her being active in the femi- nist move- ment ("B")	rank the conjunction ("A & B") as more probable for Vera com- pared to her being active in the femi- nist move- ment ("B")	rank the conjunction ("A & B") as more probable for Vera com- pared to her being active in the femi- nist move- ment ("A")	rank the conjunction ("A & B") as more probable for Vera com- pared to her being a bank teller ("A")
Left wing	-0.83*		-0.41		-0.26	
	(0.33)		(0.30)		(0.30)	
Right wing		1.05*		0.63*		0.15
		(0.41)		(0.32)		(0.32)
Constant	1.34***	0.67***	-0.15	-0.53**	0.11	-0.05
	(0.24)	(0.19)	(0.19)	(0.18)	(0.19)	(0.18)
N (Politicians)	186	186	186	186	186	186

Notes. * $p<0.05$; ** $p<0.01$; *** $p<0.001$

³¹ If we only test for the odds of choosing ordering *e* (most likely she is a bank employee, than a bank employee who is active in the feminist movement, than someone who is active in the feminist movement) thus without the other two reverse stereotypical orderings, similar results are found.

ONLINE APPENDIX 12. INFLUENCE OF RANDOMIZATION, LEFT-RIGHT SELF-PLACEMENT, GENDER, EXPERIENCE, EXPERTISE, EDUCATION AND MATH SKILL

This online appendix presents the results of randomization tests and regression analyses for each of the tests in the main text to allow adding control variables.

Table S9: *Randomization tests politician sample.*

Variable	Test statistic	n	Scenario		
			Making the headlines	Earthquake	Nuisance
Woman	Prtest	210	P=0.16	P=0.87	P=0.11
Age	Ttest	87	P=0.27	P=0.17	P=0.52
Municipality	Chi2	187	P=0.11	P=0.54	P=0.55
Party	Chi2	184	P=0.94	P=0.56	P=0.70
Education level [scale: 1-3]	Spearman	211	P=0.26	P=0.44	P=0.44

Note. All tests are two-tailed, p-values indicate the probability that there is no difference in the variable specified between the conditions allocated; Prtest = equality of proportions test; Spearman = Spearman's rho.

Table S9 shows that gender, age, municipality, party and education level were not significantly different between the different conditions.

The results in Table S10 below show that results of the regression analyses with control variables added are similar to the results presented in the main text. Adding gender, education level, political experience, math skill and right-left self-placement as control variables does not change whether the treatment in each scenario is significant and in which direction. The exception is the estimated likelihood for an earthquake rather than a natural disaster, additional tests show that (although not significant in Table S10, column 3), math skill is the reason for this. Without math skill the treatment effect is similar to presented in the paper. Perhaps, a higher skill in math helps participants avoid bias in this scenario. However, this is not the most probable explanation, since math skill has no effect when using a different method to correct for outliers: trimming (see Angrist and Pischke 2008; Crump, Hotz, Imbens and Mitnik 2006; Shor and McCarty 2011). When we trim the top and bottom 10 per cent of all answers from the estimated likelihoods for both the earthquake and natural disaster conditions, we focus on those participants who provided answers in the same order of magnitude. In this way, outliers are removed without changing the measurement scale of the data (i.e. likelihood percentages). In

addition, math skill has no effect on the other scenarios. It thus appears more likely that math skill interferes with this operationalization (i.e. log transformation) for the earthquake scenario. Because of the high number of missing data (see Online Appendix 8), we did not include age as a control variable. The first and fourth column of Table S10 show that right-left self-placement has a significant effect on committing the conjunction error (larger chance for those leaning more to the right), and on the perceived importance of the nuisance issue (more important for those leaning more to the right). The third column shows that those with higher education levels perceive it as less likely that a natural disaster/earthquake will lead to dozens of people getting injured in 2018.

Table S10. Adding control variables to the regression/logit results for each test.

	Conjunction (logit)	Likelihood of mak- ing the headlines (log transformed)	Likelihood of doz- ens of people get- ting injured (log transformed)	Issue importance	Total budget allocated (log transformed)
Woman	-0.60 (0.38)	0.12 (0.19)	0.07 (0.15)	0.03 (0.14)	0.16 (0.19)
Education level	-0.18 (0.31)	-0.20 (0.13)	-0.26* (0.11)	-0.00 (0.10)	0.05 (0.13)
Political expe- rience (years)	-0.01 (0.04)	0.01 (0.02)	0.00 (0.02)	-0.01 (0.01)	-0.03 (0.02)
Math skill	0.03 (0.12)	0.06 (0.06)	-0.09 (0.05)	0.02 (0.04)	0.11 (0.06)
Right-left self- placement (left)	-0.30* (0.15)	0.06 (0.07)	0.09 (0.06)	-0.16** (0.05)	-0.10 (0.07)
Due to terrorism		-1.90*** (0.18)			
Due to an earthquake			0.26 (0.15)		
53 people who cause a nuisance				-0.06 (0.13)	0.16 (0.18)
Constant	2.66* (1.31)	2.27*** (0.59)	1.95*** (0.49)	5.49*** (0.42)	4.55*** (0.58)
N	147	155	160	168	167

Note. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

To examine whether politicians with different levels of political experience are sensitive to displaying the biases related to the representativeness heuristic, Table S11 shows the results of

interacting the treatments in the various scenarios with political experience (years in office). Table S11 shows no significant interaction effects for any of the scenarios. This should not be interpreted as strong evidence that experience has no effect, since the many missing values on this variable make the sample size for these analyses rather small.

Table S11. *The effect of experience on the use of the representativeness heuristic in the various scenarios.*

	Conjunction (logit)	Likelihood of making the headlines (log transformed)	Likelihood of dozens of peo- ple getting in- jured (log transformed)	Issue impor- tance	Total budget allocated (log transformed)
Political expe- rience	0.01 (0.04)	0.01 (0.03)	-0.03 (0.03)	-0.01 (0.02)	-0.02 (0.03)
Due to terro- rism		-1.92** (0.29)			
Due to terror- ism* Political experience		0.00 (0.04)			
Due to an earthquake			-0.02 (0.24)		
Due to an earthquake * Political expe- rience			0.04 (0.03)		
53 People who cause a nui- sance				-0.16 (0.21)	0.18 (0.28)
53 People who cause a nui- sance* Political experience				0.01 (0.15)	-0.01 (0.20)
Constant	0.82** (0.29)	2.51*** (0.20)	1.34*** (0.18)	4.93*** (0.03)	4.93*** (0.04)
N	148	156	161	169	168

Note. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table S12 presents results of an analysis similar to that in Table S11, but now for expertise (being a spokesperson on an area related to the troublemaker scenario, i.e. order/security/safety [*veiligheid*] or use/management of public spaces [*openbare orde*]). Table S12 again shows no significant interaction effects. In this case the sample sizes are larger, but, as we explain in the main text, based on the literature, expertise (defined in this way) should not be expected to

influence the use of the representativeness heuristic in a major way.

Table S12. *The effect of expertise on the use of the representativeness heuristic in the nuisance scenario.*

	Issue importance	Total budget allocated (log transformed)
Expertise	0.01 (0.17)	-0.11 (0.24)
53 People who cause a nuisance	0.08 (0.18)	-0.07 (0.26)
Expertise *53 People who cause a nuisance	-0.37 (0.23)	0.11 (0.34)
Constant	4.87*** (0.13)	4.89*** (0.19)
N	211	210

Note. * $p<0.05$; ** $p<0.01$; *** $p<0.001$

Tables S13 and S14 present results of an analysis similar to that in Table S11 and S12, but now for education level and self-rated math skill respectively. Again, neither analysis shows significant interaction effects. This indicates that politician participants with different levels of math skill or education were not more or less sensitive to the difference between the conditions in the various scenarios.

Table S13. *The effect of education on the use of the representativeness heuristic in the various scenarios.*

	Likelihood of making the headlines (log transformed)	Likelihood of dozens of people getting injured (log transformed)	Issue importance	Total budget allocated (log transformed)
Education	-0.18 (0.16)	-0.14 (0.13)	-0.08 (0.12)	-0.06 (0.18)
Due to terrorism	-1.59** (0.58)			
Due to terrorism* Education	-0.14 (0.23)			
Due to an earthquake		0.48 (0.50)		
Due to an earthquake *		-0.10 (0.20)		
Education				
53 People who cause a nuisance			-0.30 (0.43)	-0.44 (0.63)
53 People who cause a nuisance* Education			0.07 (0.17)	0.18 (0.25)
Constant	3.05*** (0.39)	1.55*** (0.32)	5.07*** (0.31)	4.98*** (0.45)
N	195	201	211	210

Note. * $p<0.05$; ** $p<0.01$; *** $p<0.001$

Table S14. *The effect of math skill on the use of the representativeness heuristic in the various scenarios.*

	Likelihood of making the headlines (log transformed)	Likelihood of dozens of people getting injured (log transformed)	Issue importance	Total budget allocated (log transformed)
Math skill	0.12 (0.06)	-0.11 (0.06)	0.04 (0.05)	0.06 (0.07)
Due to terrorism	-0.88 (0.66)			
Due to terrorism* Math skill	-0.18 (0.10)			
Due to an earthquake		0.11 (0.55)		
Due to an earthquake *		0.01		
Math skill				
53 People who cause a nuisance		(0.08)		
			0.66	-0.07
53 People who cause a nuisance* Math skill			(0.48)	(0.67)
Constant			-0.12 (0.07)	0.02 (0.10)
<i>N</i>	1.87*** (0.40)	1.92*** (0.42)	4.60*** (0.31)	4.48*** (0.44)
Math skill		201	210	209

Note. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

ONLINE APPENDIX 13. ARE SOME POLITICIANS MORE LIKELY TO USE THE REPRESENTATIVENESS HEURISTIC ACROSS SCENARIOS?

This online appendix provides the results of an analysis to test whether politician participants who committed the conjunction error in the Vera scenario were also more likely to give biased answers in the other scenarios. None of the interaction estimates presented in Table S15 are significant, which means that politician participants who committed the conjunction error in the Vera scenario did not give different answers in the other scenarios compared to politician-participants who did not commit the conjunction error in the Vera scenario. Together these results provide no indication that some politician participants are more likely to use the representativeness heuristic in general. Instead, the findings suggest that whether the heuristic is used differs from decision to decision.

For issue importance and total budget, the hypothesis for those using the representativeness heuristic was that this did not differ across conditions; a result that we found and presented in the main paper. Consequently, it would be difficult to show for these scenarios that those participants committing the conjunction error in the Vera scenario gave answers which were even more similar across conditions. However, it could have been the case that there were some politicians that avoided bias in all scenarios. If this was the case, they should have given different responses to the issue importance and total budget questions depending on the condition (23 or 53 people who cause a nuisance) they were in, the results in columns 3 and 4 from Table S15 shows that this was not the case.

Table S15. *The likelihood to use of the representativeness heuristic in other scenarios when shown to make the conjunction error in the Vera scenario.*

	Likelihood of making the headlines (log transformed)	Likelihood of dozens of people getting injured (log transformed)	Issue importance	Total budget allocated (log transformed)
Made conjunction error in Vera scenario	-0.25 (0.25)	-0.30 (0.22)	0.01 (0.20)	0.07 (0.30)
Due to terrorism	-2.16*** (0.31)			
Made conjunction error in Vera scenario * Due to terrorism	0.20 (0.36)			
Due to an earthquake		-0.09 (0.27)		
Made conjunction error in Vera scenario * Due to an earthquake		0.32 (0.32)		
53 People who cause a nuisance			-0.11 (0.23)	0.48 (0.34)
Made conjunction error in Vera scenario * 53 people who cause a nuisance			-0.05 (0.28)	-0.75 (0.40)
Constant	2.84*** (0.21)	1.47*** (0.18)	4.91*** (0.18)	4.77*** (0.26)
N	171	176	186	185

Note. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

ONLINE APPENDIX 14. RESULTS WITHOUT LOG TRANSFORMATIONS

The article tests H2, H3 and H5 using a log-transformed dependent variable. In this appendix, we present some additional analyses that show why this is necessary: the untransformed dependent variables violate the basic assumption of the t-test that the data are distributed normally (see any statistics textbook, for example Healey 2014). As a further robustness check, we present the untransformed results of all hypothesis tests in Table S16. Only the result for H3 differs. Figures S5-S8 below show that the non-log-transformed result for H3 is due to the disproportionate influence of outliers (the dots above the box and whiskers in the left-hand side of Figure S7, some of which are over 5 to 10 SD's above the median, compared to the right-hand side). Figure S5 displays the skewed distribution of the reported probabilities without a log-transformation; Figure S6 shows that the skewed distribution is much reduced after the log-transformation. Figures S7 and S8 reveal that the log-transformation substantially reduces the number of outliers and their distance to the median (which indicates the amount of disproportionate influence they have on the mean relative to the bulk of respondents within one SD from the median). These latter two figures thus show that the log-transformation we apply improves the distribution of the dependent variable (likelihood of an earthquake/natural disaster) for H3, illustrating why this result is presented in the main manuscript.

Table S16. Results for each test for the politician participants without log-transformation.

Hypothesis	Test	Result
H1	Conjunction error I: The Linda/Vera problem	See article, no log-transformation was applied.
H2	Conjunction error II: The making the headlines scenario	Similar to the log-transformed results reported in the article, the politician participants estimated the likelihood of making the newspaper in general larger than making the headlines due to a terrorist attack ($t=7.60$, $p<0.001$, $n(\text{politicians})=195$).
H3	Conjunction error III: The earthquake scenario	Without log-transforming or trimming the probability estimations, the politician participants did not estimate the likelihood of an earthquake to be significantly higher (or lower) than a natural disaster ($t=0.45$, $p=0.67$ single tailed in the direction of the earthquake being more likely, $n(\text{politicians})=201$).
H4	Scope neglect I: Troublemaker scenario (importance)	See article, no log-transformation was applied.
H5	Scope neglect II: Troublemaker scenario (total budget)	Similar to the log-transformed results reported in the article, the difference between the budget indicated by politician participants to deal with 53 troublemakers was not significantly higher than that for 23 troublemakers ($t=0.63$; $p=0.26$ single tailed in the direction of the budget for 53 troublemakers being higher, $n(\text{politicians})=210$).

Figure S5. Histograms of the untransformed estimated probability of each disaster occurring, according to the politician participants.

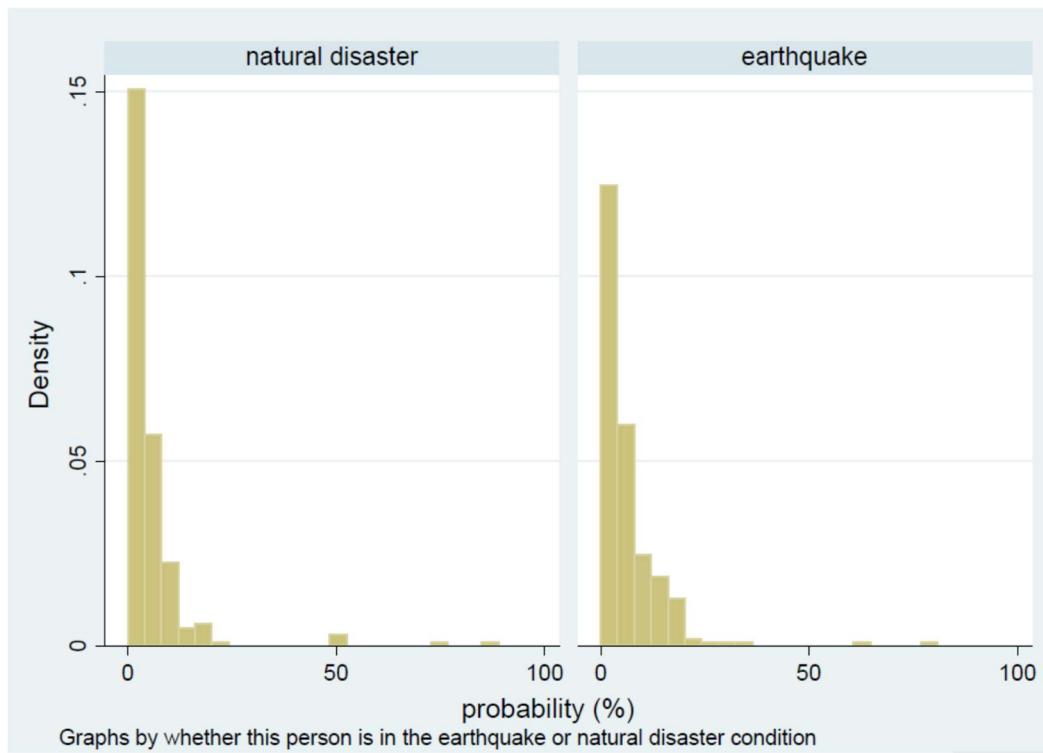


Figure S6. Histograms of the log-transformed estimated probability of each disaster occurring, according to the politician participants.

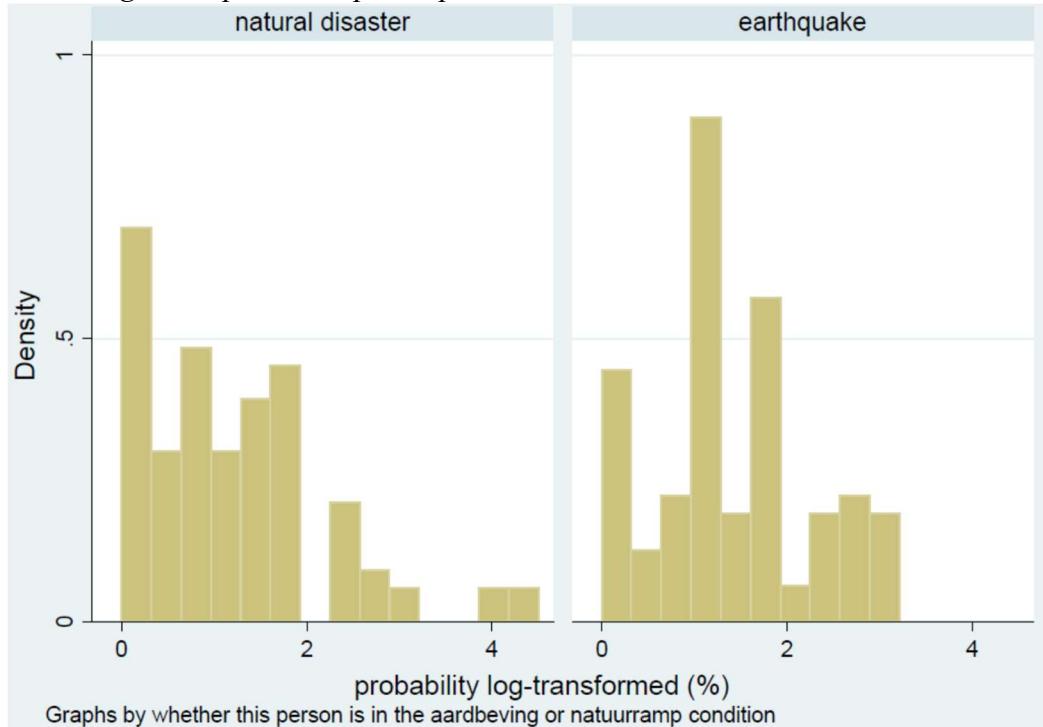


Figure S7. Boxplot of the untransformed estimated probability of each disaster occurring, according to the politician participants.

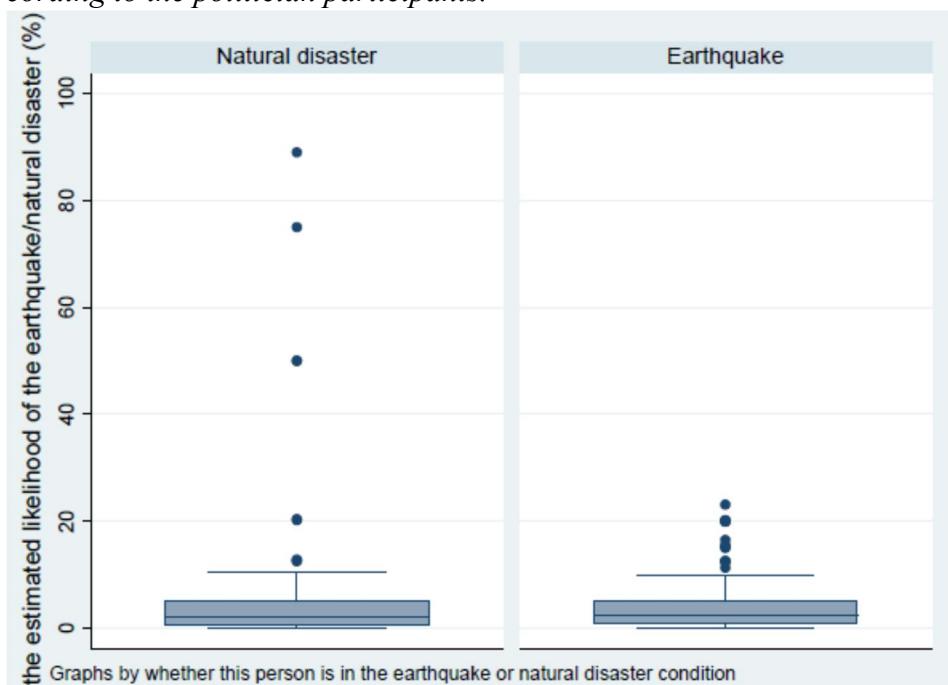


Figure S8. Boxplot of the log-transformed estimated probability of each disaster occurring, according to the politician participants.

