Sample Instructions and questionnaires for Study 1

The instructions in normal font were presented to all three experimental groups. *Instructions for participants in the pre-accountable also included the text in italics.* These sample instructions and questionnaires were given for reports in which the truthful speaker was female. For the reports in which the truthful speaker was male, the gender of the speakers was switched in the instructions.

Today's task requires you to assume the role of a judge. You will be listening to the police reports of two separate offenses committed here in the Brussels area within the last few years. Your task is to listen to these reports. Each of the two reports will be presented only once, so you will have only one opportunity to learn what happened -- the gravity of the offenses, the circumstances surrounding them, etc. After you have listened to both reports, we will ask you to judge the perpetrators of the offenses and to recall details of the crimes. [*You will have to make your judgments attentively, as we will later record your while you justify those judgments orally to the experimenter.*]

There is one feature that makes this task a little more demanding. For each of the reports that you will listen to, you will listen to two speakers: a woman and a man. The woman sticks to reality and the information she provides is true. By contrast, the information provided by the man is false, taken from other, unrelated police reports and.

Let's review. You will soon be listening to a pair of crime reports, trying to learn and remember as much as you can about the two criminals and their actions. Mixed in with the verified facts (uttered by the woman) around the crimes will be several falsehoods (pronounced by the man). After you're done listening to the reports, you will be asked to judge the perpetrators of the crimes and recall details of the stories [*while you also need to orally justify those judgments to the experimenter.*]

Feel free to ask any questions now.

Questionnaire used in Study 1.

A. Now that you have listened to both crime reports, we would like you to think over the facts concerning each of the two crimes and decide upon the punishment that each of the two criminals should be assigned. [*You should make these judgments carefully since you will be asked to justify them. At the end of the experiment, you will be asked to justify your judgment orally to the experimenter explaining on which facts your decisions were based. Your explanations will be registered.*]

First of all, indicate the severity of the punishment that Etienne and Dimitri deserve. Please answer by typing the number that better matches your opinion on each scale.

Etienne, the perpetrator of the first report should be assigned a punishment

extremely light 1---2---3---4---5---6---7---8---9---10 extremely heavy

Dimitri, the perpetrator of the second report should be assigned a punishment

extremely light 1---2---3---4---5---6---7---8---9---10 extremely heavy

Secondly, we would like to have an estimation of the actual prison term that Etienne and Dimitri should be sentenced to, in your opinion. You may sentence each of the criminals to anywhere from 0 to 5 years in prison. Please take your time.

Etienne, the perpetrator in the first crime report, should be sentenced to ____ years of actual prison time. (Write a number between 0 and 5 in the blank.)

Dimitri, the perpetrator in the second crime report, should be sentenced to ____ years of actual prison time. (Write a number between 0 and 5 in the blank.)

B. Moreover, we would also like to have your opinion on the following questions. Please answer by typing the number that best matches your opinion on each scale.

Please reply by choosing the number that more accurately reflects your opinion.

1) What are your general feelings towards Etienne?

neutral 1---2---3---4---5---6---7---8---9---10 extreme dislike

2) How dangerous do you think Etienne is?

slightly dangerous 1---2---3---4---5---6---7---8---9---10 extremely dangerous

3) How likely is Etienne to slip back?

not likely at all 1---2---3---4---5---6---7---8---9---10 extremely likely

1) What are your general feelings towards Dimitri?

neutral 1---2---3---4---5---6---7---8---9---10 extreme dislike

2) How dangerous do you think Dimitri is?

slightly dangerous 1---2---3---4---5---6---7---8---9---10 extremely dangerous

3) How likely is Dimitri to slip back?

not likely at all 1---2---3---4---5---6---7---8---9---10 extremely likely

C. You have listened to several statements about the two crimes. Some of these statements were true (uttered by the woman) and some were false (uttered by the man). Please read each of the phrases below and decide, first of all, if you remember reading the phrase. If you remember reading it, try to remember also if the phrase was true or false.

Please answer by writing one of the following letters in the dedicated space. Here are the letters and their meaning:

V = The phrase was true.

F = The phrase was false.

N = The phrase did not appear in the reports.

(Then the statements in Table S3 appeared in a randomized order).

D. Since you were finally attributed to a control group, you won't have to orally justify your responses, only to write them down. Therefore, we now ask you to explain, in a few phrases, your decisions concerning the two crimes.
Please justify your judgments of the two perpetrators, explaining the facts on which your decisions were based. Did you judge the two perpetrators differently? If yes, on what grounds?

Power Calculation for Study 1

We calculated the contrast interaction that represented the accountability effect reported in Tetlock (1983b). This corresponded to the contrast between the non-significant difference of the two pre-accountable groups – incriminating first vs. exculpating first– on the one hand, and the significant differences of the averaged – incriminating first vs. exculpating first– post-accountable and control groups). The contrast reflected an effect size of $f = 0.49$. Like Tetlock's our judgments design was a mixed *group* (pre-accountable vs. post-accountable vs. control) x *false evidence* (aggravating vs mitigating). According to G*power, to detect a within-between interaction effect of $f = 0.49$ on our participants' judgments, given our 3 groups, 2 measurements, a correlation of $r = .30$ between the judgments for the aggravated and mitigated defendants (as reported by Pantazi et al., 2018) we would need 27 participants with a .95 power at the .05 alpha level.

Accordingly, the design for the memory analysis was a mixed *group* (pre-accountable vs. post-accountable vs. control) × *statement type* (true vs. false). To detect an effect similar to that of Tetlock on participants' memory, given our three groups, 2 measurements, a correlation of .237 between false evidence misremembered as true and true evidence misremembered as true (as reported by Pantazi et al., 2018), with a .95 power at the .05 alpha level we would need 24 participants.

To detect the within-subjects main effects of *false evidence* and *statement type* (an average of $f = 0.34$ for the former and $f = 0.22$ for the latter across the studies in Pantazi, 2018 that used the same materials and measures as we did), given our three groups and the aforementioned correlations among our repeated measures, we would need 36 participants for the judgment and 87 participants for the memory measures with a .9 power at the .05 alpha level.

Power Calculation for Study 2

The average within-subject *false statement* and *statement type* effects on judgments and memory respectively in Pantazi et al. (2018) amounts to $d = .69$ and $d = .50$ respectively. According to G*power, we needed 25 and 45 participants accordingly to replicate these within-subject mean differences with .9 power at the .05 alpha level.

In the analyses comparing the judges to the mock-jurors from Study 1, the designs were group (mock-jurors vs. judges) × false evidence (aggravating vs. mitigating), and group (mock-jurors vs. judges) × statement type (true vs. false). To detect interaction effects of medium size ($f = .25$), suggesting differential truth bias effects in the two samples, with a .95 power at the .05 alpha level, we would need 76 participants overall for the judgment, and 82 participants overall for the memory measures (half of which would be judges), given the .30 and .237 respective within measures correlations. In sum, 41 judges would be needed maximum to detect any of our hypothesized effects.

Memory analysis for new items

In both studies, the memory test included 32 new items in total, 12 of which were aggravating, 12 of which were attenuating and 8 of which were neutral. We analyzed memory patterns for the new statements using a Generalized Linear Mixed Model for binomial data (GENLINMIXED procedure in SPSS; see Quené & van den Bergh, 2008). Participants could either correctly identify a new statement as new, misidentify it as true or misidentify it as false. We treated these three possible responses as repeated measures at three different levels of a fixed factor *identification* (correct response vs. confounded true vs. confounded false), on a target binomial variable. The target variable was thus repeated three times for each of the 32 new items. Target was coded 1 at the level of the identification factor that represented participants' actual response, and 0 at the other two levels. The target variable was thus binominal, and we ran our analysis after it was logit transformed (LOGIT link Target_Option in SPSS; see Quené and van den Bergh, 2008).

New Items Study 1

Besides the identification factor just described, in the analysis of the new items in Study 1, we included two additional fixed factors: *statement type* reflected whether each statement was aggravating, attenuating or neutral and *group* (pre-accountable vs. post-accountable vs. control). We included in the analysis all two- and three-way interactions. We also included intercepts per subject and items.

The memory patterns for the new statements in Study 1 can be found in Fig. S1. The *identification* factor ($F(2, 6961) = 1045.65$, $p < .001$), *statement* × *identification* interaction ($F(4, 6861) = 26.39$, $p < .001$) and *identification* × *group* interaction ($F(4, 6861) = 6.12$, $p < .001$) were significant. We run Bonferroni corrected pairwise comparisons to decompose these effects. New statements were

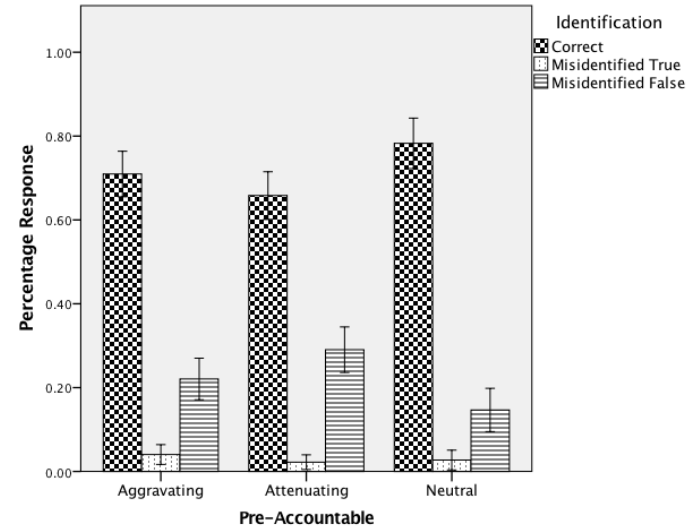
significantly better identified than misidentified as true ($t(1, 6861) = 73.08$, $p < .001$) or as false ($t(1, 6861) = 51.89$, $p < .001$), while new statements were also more likely to be misidentified as false than as true ($t(1, 6861) = 10.89$, $p < .001$). The *statement* × *identification* interaction signaled that new attenuating statements were less correctly identified than both new attenuating ($t(1, 6861) = -4.76$, $p < .001$) and the new neutral ones ($t(1, 6861) = -7.38$, $p < .001$), while neutral new items were more correctly identified than the aggravating new ones as well ($t(1, 6861) = 2.91$, $p = .004$). This pattern was mirrored in that new aggravating statements were less misidentified as false than the new attenuating statements ($t(1, 6861) = 3.42$, $p = .001$) and more misidentified as false than the new neutral statements ($t(1, 6861) = 3.42$, $p = .001$), while the new attenuating statements were more likely to be misidentified as false than the new neutral statements ($t(1, 6861) = -3.42$, $p < .001$). Finally, the *identification* × *group* was due to the fact that the pre-accountable group identified fewer new statements correctly than both the post-accountable ($t(1, 6861) = -2.04$, $p = .041$) and the control group $t(1, 6861) = -2.40$, $p = .016$), since it misidentified more new statements as false than both the post-accountable ($t(1, 6861) = 2.80$, $p = .005$) and the control ($t(1, 6861) = 3.76$, $p < .001$).

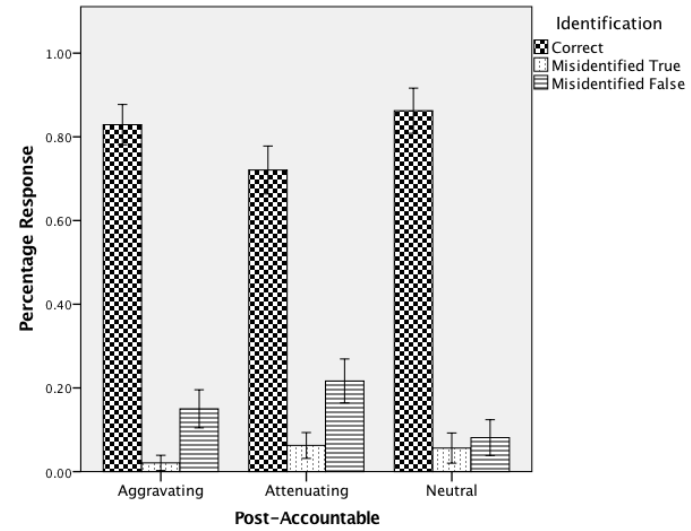
The analyses of new statements in Study 1 suggest that if in doubt, participants would tend to misremember new statements as false rather than as true. While the aggravating new statements were less correctly identified than the other two categories (as they were more misidentified as false), the pre-accountable group had an overall worse performance than the other two groups. This result is in line with the memory patterns for the experimental statements and corroborates the idea that, if anything, the accountability manipulation back-fired rendering participants who encoded information under accountability more error-prone.

New Items Study 2

In Study 2 on top of the *identification* factor we included *statement* (aggravating vs. attenuating vs. neutral). The memory patterns for the new statements in Study 1 can be found in Fig. A2. Both the *identification* factor ($F(2, 4008) = 618.28$, $p < .001$) and the *identification* × *statement* interaction ($F(4, 4008) = 17.72$, $p < .001$) were significant. Bonferroni corrected pairwise comparisons showed that, overall, new statements were significantly more correctly identified than misidentified as true ($t(1, 4008) = 48.93$, $p < .001$) or false ($t(1, 4008) = 42.64$, $p < .001$), while they were more likely to be misidentified as false than as true ($t(1, 4008) = 2.56$, $p = .010$). The significant interaction was due to the fact that neutral new statements were not more likely to be misremembered as false than as true ($t(1, 4008) = .79$, $p = .057$).

The memory patterns for new statements in Study 2, as in Study 1, reveal that, if anything, participants who did not remember the identity of a statement, they tended to misidentify it as false than as true. Thus, although the responses for new and old statements are expected to differ, the new items analyses do not provide any evidence for a general guessing bias explanation of our truth-bias results.





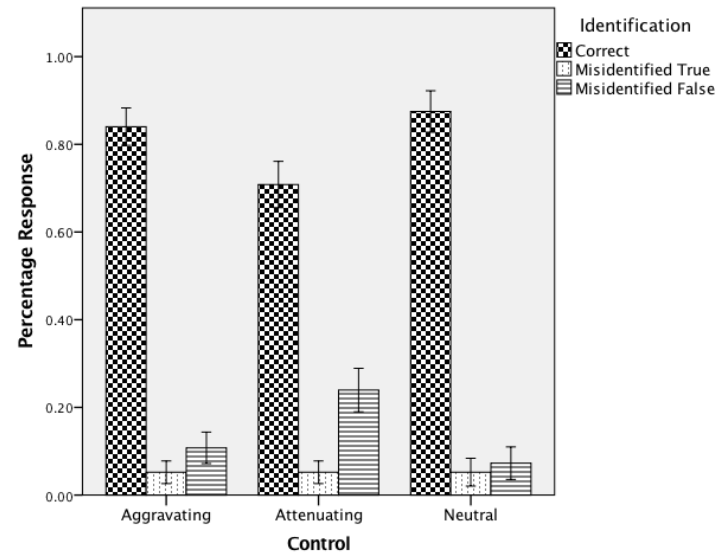


Figure S1. Study 1: Percentage responses for the three categories of new statements in the memory test, separately for each group. Error bars represent 95% CIs.

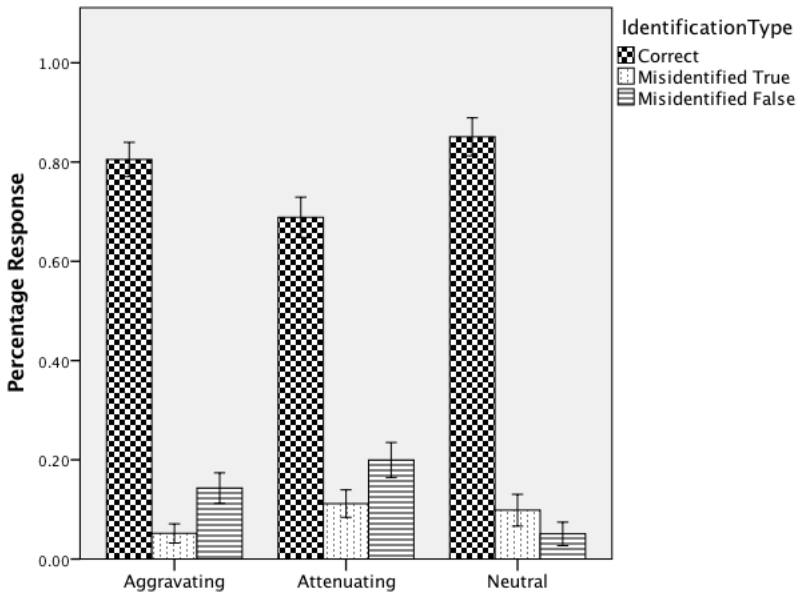


Figure S2. Study 2: Percentage responses for the three categories of new statements in the memory test. Error bars represent 95% CIs.

Table S1

Mean Judgment severity, and SDs for the aggravated and mitigated defendants per group in Study 1.

False Information	Pre-Accountable		Post-Accountable		Control	
	M	SD	M	SD	M	SD
Aggravated	5.70	1.93	6.06	2.03	6.37	1.85
Mitigated	4.99	2.04	4.91	1.76	5.40	2.06

Table S2

Mean percentage error rates for the true and false statements in the memory test and SDs, per group in Study 1.

Statement Type	Pre-Accountable		Post-Accountable		Control	
	M	SD	M	SD	M	SD
True	.11	.31	.06	.26	.02	.14
False	.27	.44	.17	.37	.22	.41

Table S3

The statements and their status in the memory test per report and version, in Experiments 1 &2

	Version	
Dimitri	Aggravating	Attenuating
Dimitri's brother-in-law tried to prevent him from leaving his house.	Old T.1	Old T.1
Dimitri had a friend who lived nearby	Old T.2	Old T.2
Dimitri went to the night shop to buy cigarettes.	Old T.3	Old T.3
Dimitri had a bleu Renault.	Old T.4	Old T.4
The argument between Dimitri and his wife erupted due to his affairs with other women.	Old F. Aggr.1	New Aggr.1
Dimitri bought and used a considerable amount of cocaine.	Old F. Aggr.2	New Aggr.2
Dimitri violently kicked a magazine rack out of his way	Old F. Aggr.3	New Aggr.3
Dimitri threatened to sexually assault a client in the night shop.	Old F. Aggr.4	New Aggr.4
Dimitri explained to the clerk that he needed the cash to cover a serious operation that his daughter underwent.	New Att.1	Old F. Att.1
Dimitri apologized for his action.	New Att.2	Old F. Att.2
Dimitri left his house to protect his kids from witnessing such incidents.	New Att.3	Old F. Att.3
The argument between Dimitri and his wife erupted due to his affairs with other men	New Att.4	Old F. Att.4
Dimitri threatened to kill his brother-in-law if he ever again interfered in his life.	New Aggr.1	New Aggr.5
Dimitri hit his son while he interfered in his parents' argument.	New Aggr.2	New Aggr.6
Dimitri was a lazy man who had never worked in his life.	New Aggr.3	New Aggr.7
Dimitri was extremely jealous of his wife	New Aggr.4	New Aggr.8
Dimitri told the clerk he would not hurt him	New Att.5	New Att.1
During the robbery, Dimitri started crying.	New Att.6	New Att.2
Neighbors report that Dimitri was a good family guy.	New Att.7	New Att.3
Dimitri did not finally take the money that the clerk handed him.	New Att.8	New Att.4
Dimitri was working in a supermarket.	New Neutr.1	New Neutr.1

Dimitri and his brother-in-law were old enemies from school.	New Neutr.2	New Neutr.2
Leaving his house, Dimitri explained to his wife that he needed to get some air.	New Neutr.3	New Neutr.3
The clerk was a university student.	New Neutr.4	New Neutr.4
	Version	
Etienne	Aggravating	Attenuating
Etienne directed Victor through a residential area.	Old T.1	Old T.1
After stealing Victor Etienne went back to his place	Old T.2	Old T.2
In the car, Victor started recounting a funny story.	Old T.3	Old T.3
Etienne told Victor that he wanted to get off downtown.	Old T.4	Old T.4
Etienne told Victor that he found crippled people really disgusting.	Old F. Aggr.1	New Aggr.1
Etienne's kids said he committed infamies frequently.	Old F. Aggr.2	New Aggr.2
Etienne threatened to slit Victor's throat if he tried to follow.	Old F. Aggr.3	New Aggr.3
After stealing Victor Etienne committed a burglary.	Old F. Aggr.4	New Aggr.4
Before leaving Etienne gave back to Victor all his valuables.	Old F.Att.1	Old F. Att.1
The night of the crime Etienne had found that his wife cheated on him.	Old F.Att.2	Old F. Att.2
Before stealing him, Etienne told Victor that he was ashamed of what he had to do.	Old F.Att.3	Old F. Att.3
Etienne's kids said that up to that point he was a good family guy.	Old F.Att.4	Old F. Att.4
After stealing Victor Etienne broke the window display of a shop.	New Aggr.1	New Aggr.5
Etienne stole a valuable from his neighbor's apartment.	New Aggr.2	New Aggr.6
When leaving his neighbor's apartment Etienne attacked a passer-by	New Aggr.3	New Aggr.7
With his valuables in his possession, Etienne punched Victor.	New Aggr.4	New Aggr.8
Victor tried to steal Etienne first.	New Att.5	New Att.1
Etienne wanted to take revenge from Victor who had attacked his son	New Att.6	New Att.2
Etienne regretted committing this crime.	New Att.7	New Att.3
Before leaving Etienne threw his ring back to Victor's lap.	New Att.8	New Att.4
When leaving his neighbor's apartment, Etienne stopped to chat with some old friends.	New Neutr.1	New Neutr.1
Victor was going to join his friends in a strip-tease bar.	New Neutr.2	New Neutr.2

Victor was going to join his family in the theater.	New Neutr.3	New Neutr.3
Etienne had a daughter and a son.	New Neutr.4	New Neutr.4