Psychological Topics, 28 (2019), 1, 73-91

Original Scientific Paper UDC: 159.955.072

doi:https://doi.org/10.31820/pt.28.1.4

The Relationship between Consistency and Consensuality in Syllogistic Reasoning

Igor Bajšanski and Valnea Žauhar

University of Rijeka, Faculty of Humanities and Social Sciences,
Department of Psychology, Rijeka, Croatia

Abstract

In this study, we examined the effects of response consensuality in syllogistic reasoning on patterns of answer change by using the two-response paradigm. Participants evaluated 24 syllogistic problems previously found to differ in consensuality, including consensually correct (CC), consensually wrong (CW), and nonconsensual (NC) items. Each problem was presented two times and participants were required to provide an initial quick answer to the first presentation, to rethink the problem, and to provide their second and final response without time limits to the second presentation. Participants reported the feeling-of-rightness (FOR) following the initial response, and the final judgment of confidence (FJC) after the final response. Following the assumptions of Koriat's (2012) Self-Consistency Model of confidence, we expected higher probability of answer change for initial nonconsensual responses than for initial consensual responses. The results showed that patterns of answer change, as well as metacognitive judgments and response times, were related to item consensus and response consensuality. Nonconsensual responses were more likely to be changed than consensual responses, and the probability of answer change correlated negatively with item consensus. Faster response times and higher FORs and FJCs were obtained for consensual and consistent responses than for nonconsensual and inconsistent responses. The obtained results indicate that answer change may in part be a consequence of random fluctuations in representation sampling, or in generating evidence that supports each of the two response options.

Keywords: syllogistic reasoning, confidence, Self-Consistency Model, consensuality, two-response paradigm

✉ Igor Bajšanski, Department of Psychology, Faculty of Humanities and Social Sciences, University of Rijeka, Sveučilišna avenija 4, 51000 Rijeka, Croatia. E-mail: *sibajsan@ffri.hr*

This work was supported by the University of Rijeka Research Grant under Grant number 13.04.1.3.11.

Introduction

Two-response paradigm, introduced by Thompson and her colleagues (Shynkaruk & Thompson, 2006; Thompson, Prowse Turner, & Pennycook, 2011), is used to explore the relationships between quick intuitive responses (Type 1) and deliberate responses (Type 2) to reasoning tasks. The paradigm requires participants to provide two responses to a series of reasoning problems. First, they are asked to provide an initial quick answer to a reasoning problem. Second, immediately after the initial response is given, the problem is presented again, and participants are allowed to rethink the problem and to provide a second and final response without time limits. After each response is provided, participants report their confidence in response accuracy. Confidence following the initial response is labelled feeling-of-rightness (FOR), and confidence given after the final response is labelled final judgment of confidence (FJC) (Thompson et al., 2011; see also Ackerman & Thompson, 2015, 2017).

Research using the two-response paradigm revealed some important findings, with clear theoretical implications for the psychology of reasoning, and in particular for dual-process accounts of reasoning (Evans, 2003; Evans & Stanovich, 2013; Sloman, 1996; Thompson, 2009). As it was suggested by the Metacognitive Reasoning Theory (Thompson et al., 2011; see also Thompson, 2009), metacognitive feelings about the accuracy of initial responses determine the amount of Type 2 engagement. Therefore, lower FOR accompanying initial responses should be associated with higher probability of answer change and with longer rethinking times. Both assumptions are strongly supported by the experimental findings (Bago & De Neys, 2017; Pennycook & Thompson, 2012; Shynkaruk & Thompson, 2006; Thompson et al., 2011; Thompson, Evans, & Campbell, 2013; Thompson & Johnson, 2014). Furthermore, it was found that, in general, reasoners are reluctant to change their initial responses, even if those responses are incorrect. Typically, the same response is given in a subsequent presentation of an item in more than 70% of trials, and in some studies this percentage is about 90% (e. g., Bago & De Neys, 2017). At the trial level, the combinations of two responses can be classified into four categories with respect to their accuracy. Bago & De Neys (2017) labelled these four categories of responses as 00, 01, 10, and 11. The first number indicates the accuracy of the first response, and the second number indicates the accuracy of the second response. Therefore, category 00 includes trials on which both responses were incorrect, category 01 includes trials on which the first response was incorrect and the second response was correct, etc. Trials on which the initial response is changed typically comprise of similar number of changes from initial incorrect responses to final correct responses (01 category) and changes from initial correct responses to final incorrect responses (10 category). As a consequence, accuracy of final responses is not substantially higher than accuracy of initial responses.

One important problem concerns the conditions which affect the likelihood of answer change. Previous studies have demonstrated that initial responses were more likely to be changed when a) they were followed by lower FOR, b) they were not generated fluently, and c) they were given to conflict problems which cue two different responses, one normatively correct and the other believable but incorrect (Bago & De Neys, 2017; Thompson & Johnson, 2014; Thompson et al., 2011). However, besides conflict/no conflict manipulations little is known about the properties of reasoning problems that could affect the probability of answer change and the direction of that change. The question is which factors determine whether initial incorrect answers will be corrected after rethinking, and which factors affect the probability of change of an initial correct response to a final incorrect response. In this study, we explored self-consistency as a possible factor that affects the probability of answer change and the direction of that change.

Self-Consistency Model of Confidence

Our predictions are based on Koriat's (2012) Self-Consistency Model of confidence (SCM). According to the SCM, when answering two-alternative forced choice questions (2AFC), people randomly sample the representations that are outputs of the cognitive processes involved in the decision about the options. The decision about the options is based on the outcome of the sampling process. The option supported by stronger evidence, or a larger number of representations favouring the option, is chosen. Since response decisions are based on small samples of representations derived from the item-specific pool of representations, variations in responding reflect, in part, random fluctuations in the sampling process. Furthermore, confidence assigned to the chosen option is based on the hypothetical cue labelled self-consistency. Self-consistency represents overall agreement among sampled representations, or, the proportion of representations favouring the selected option.

Two important predictions are related to this account. First, the population of representations associated with an item is shared among people, as long as we assume that there are common experiences and processes that underlie decisions about the options. Various items used to test human memory, knowledge, and judgments, typically differ reliably in item consensus or the proportion of participants who chose the consensual option. For binary choices, item consensus can vary between 50% and 100%. There is strong empirical evidence that confidence correlates with item consensus, and that it is higher for consensual responses (options chosen by the majority of participants) than for nonconsensual responses. Accordingly, when the consensual response to an item is incorrect, confidence for this answer is higher than for the correct one. Second, since it is assumed that on each presentation of an item representations are sampled from the same pool of representations associated with that item, items should differ in the consistency of responding, or, in the likelihood of repeating the same response on subsequent presentations of the item. This within-person consistency is correlated with confidence. Predictions about the relationships between consensuality and confidence, as well as between consistency and confidence have strong empirical support (for a review see Koriat, 2012; Koriat & Adiv, 2016). It should be also noted that response times track consensuality in a similar way as confidence judgments (for details see Koriat, 2012).

A further prediction of central importance for this study is the prediction about the correlation between cross-person consensus and within-person consistency. In a series of studies using various tasks including general knowledge questions (Koriat, 2008), perceptual judgments (Koriat, 2011), personal preferences (Koriat, 2013), judgments of category membership (Koriat & Sorka, 2015), attitudinal judgments (Koriat & Adiv, 2011), and social beliefs (Koriat & Adiv, 2012), Koriat and his colleagues demonstrated that within-person consistency correlated with cross-person consensus. The option chosen by the majority of participants on the first presentation of an item was more likely to be consistently chosen on subsequent presentations of that item. Furthermore, initial responses followed by high confidence judgments were more likely to be repeated on subsequent trials.

Consensuality, Consistency, and Answer Change

Findings about the relationships between within-person consistency and cross-person consensus can be used to examine the patterns of responses in two-response paradigm using the conclusion evaluation task in a domain of syllogistic reasoning. The main hypothesis of this study is that the probability of answer change in two-response paradigm should be related to consensuality of initial responses. This hypothesis is based on several notions.

First, conclusion evaluation task is similar to 2AFC questions, since it includes two response options. Various items elicit similar rates of acceptances of conclusions across studies. For example, acceptance rates for 48 syllogisms reported by Bajšanski, Žauhar, and Valerjev (2018) correlated highly with acceptance rates reported by Evans, Handley, Harper, and Johnson-Laird (1999). Therefore, syllogisms reliably vary in item consensus, and in related proportions of consensual and nonconsensual responses. From the perspective of SCM, similar patterns of relationships between confidence, consensuality and consistency are expected for different contents, as long as the tasks include two response options and reliably vary in item consensus.

Second, in our previous study (Bajšanski et al., 2018, Experiment 1) we showed that the basic principles of the SCM are applicable to the data about accuracy, confidence and response times obtained with conclusion evaluation task in syllogistic reasoning. Three sets of items were constructed. Consensually correct (CC) items included easy syllogisms with high accuracy, consensually wrong (CW) items included problems with low accuracy, and nonconsensual items (NC) items included problems with accuracy about 50%. It was demonstrated that CC and CW syllogisms were evaluated with higher confidence than NC syllogisms and that answers endorsed by the majority of participants were given higher confidence ratings than answers endorsed by the minority of participants. Therefore, in the set of CC items, correct responses were endorsed with higher confidence than incorrect responses, whereas in the set of CW items, incorrect responses were endorsed with higher confidence than correct responses.

Third, it can be assumed that changes of responses in two-response paradigm could, in part, reflect random fluctuations in representation sampling on two presentations of an item and, as a consequence, random fluctuations in decisions about the options. Patterns of responses at the trial level can be analysed with respect to within-person consistency, with 11 and 00 being the consistent responses, and 10 and 01 inconsistent responses. An important difference between two-response paradigm and Koriat's studies that examined within-person consistency is that two-response paradigm includes two sets of instructions, one set for the initial response, and the other for the final response. Participants are expected to provide the first response that comes to mind as their initial response, and to carefully rethink two options before they provide the final response. Accordingly, initial and final responses differ in response times. However, as long as we assume that responses to an item are based on the sampling from the common pool of representations, that the choices are based on the strength of evidence favouring each of the two response options, and that confidence correlates with consensuality and consistency, the probability of answer change should track consensuality of initial response.

The expected relationship between consensuality and consistency should affect probability of answer change of initial responses. We expected a higher probability of change of the initial response when it is nonconsensual than when it is consensual. As a consequence, for CC items we expected a larger proportion of change for incorrect initial responses than for correct initial responses. For CW items we expected the opposite, larger proportion of change for correct initial responses than for incorrect initial responses. For NC items similar proportions of changes of correct and incorrect initial responses were expected.

Furthermore, similar effects should emerge at the item level. Items with higher consensus (higher proportion of consensual response) are expected to have a lower probability of answer change of consensual responses than items with lower consensus. On the other hand, probability of change of initial nonconsensual responses should increase with item consensus.

Several hypotheses about metacognitive judgments and response times can also be derived.

First, higher FORs and FJCs should be given to consistent responses (categories 11 and 00) than to inconsistent responses (categories 10 and 01). This effect was demonstrated in previous studies using two-response paradigm (Bago & De Neys, 2017). Second, metacognitive judgments should differ between two subsets of consistent responses. More precisely, both FORs and FJCs should be higher for consensual consistent responses that for nonconsensual consistent responses. Third, differences in FJC given to final inconsistent responses should also depend on the response consensuality. When the initial nonconsensual response is changed to consensual response, FJC should be higher than when the initial consensual response is changed to nonconsensual response. Fourth, response times of the initial and final responses should mimic the effects of consensuality and consistency on metacognitive judgments. In short, faster response times are expected for consensual and consistent responses than for nonconsensual and inconsistent responses.

To summarize, we examined the patterns of answer change with respect to the variables previously found to affect confidence in syllogistic reasoning: item consensus and response consensuality. Two-response procedure was used, and participants evaluated 24 syllogistic problems previously found to differ in consensuality, including CC, NC, and CW items (Bajšanski et al., 2018).

Method

Participants

Seventy undergraduate psychology students (63 female) from the University of Rijeka, Croatia, participated in the experiment in exchange for course credits.

Materials

Syllogisms. The participants evaluated 24 syllogisms used in our previous study (Bajšanski et al., 2018, Experiment 1). The content of the syllogisms consisted of professions (e.g., engineer) and pastimes (e.g., mountaineer). Each pair of premises contained a unique combination of terms. All conclusions were particular-negative conclusions, including *Some a are not c* and *Some c are not a* conclusions. Three sets of consensus categories of syllogisms were used (CC, NC, CW). Half of the syllogisms in each category were valid, and half were invalid. Consensus categories systematically differed in accuracy, with CC category including syllogisms with accuracy above 60%, NC category including syllogisms with accuracy about 50%, and CW category including syllogisms with accuracy below 40% (for the detailed description of stimuli see Bajšanski et al., 2018).

Feeling-of-rightness (FOR). After providing the initial response to each problem, the participants made a FOR judgment on a 6-point scale ranging from 50% (guessing) to 100% (fully confident).

Final judgment of confidence (FJC). After providing the final response to each problem, the participants made a FJC on a 6-point scale ranging from 50% (guessing) to 100% (fully confident).

Procedure

The participants were tested individually. The stimuli presentation and data collection were controlled using the E-prime 2.0 software (Psychology Software Tools, Inc., Pittsburgh, PA, USA) running on a personal computer. At the beginning of the test, the instructions were presented on a computer screen. The procedure was explained to the participants, and they were told how to give their answers and judgments for each task. The two-response procedure was used and the participants were informed that they will evaluate each problem two times. In each trial, two premises and a conclusion were presented on the screen. The participants were asked to assume that all information in the premises was true and to evaluate each conclusion. They were instructed to give the initial response, that is, the first response that comes to mind, by pressing the YES key if they thought the conclusion followed from the premises or the NO key otherwise. After providing the first response, they were asked to provide FOR judgment on a six-point scale appearing on the screen and ranging from 50% to 100% by pressing the appropriate key (50%, 60%, 70%, 80%, 90%, 100%). The FOR judgment was followed by the second presentation of the same problem and participants were instructed to carefully rethink their response without time limit and to provide their final response by pressing the YES or NO key. They were additionally instructed that they could change the given initial response. After providing the second or final response, they were asked to provide FJC on a six-point scale ranging from 50% to 100% by pressing the appropriate key. Participants were given as much time as they needed to comprehend the instructions. After the instructions, they were given two practice problems. After the two practice problems, 24 problems were presented in a random order. The response times were collected.

Results

Direction of Change Categories: Descriptive Data

We analysed the frequencies of direction of change categories. Following Bago and DeNeys (2017) we labelled these categories as 00, 01, 10, and 11, where the first number indicates the accuracy of the initial response, and the second number indicates the accuracy of the final response. Accurate responses included acceptances of valid conclusions and rejections of invalid conclusions. Table 1 presents the frequencies of direction of change categories for three consensus categories (CC, NC, CW). Additionally, for each cell two percentages are reported. The first refers to the percentages within consensus categories. The second refers to the percentages within each of the first response categories (correct, incorrect). Finally, total frequencies of each of the four direction of change categories are presented.

Table 1

Frequencies of Trials within Consensus Categories and Direction of Change Categories

		Correct first response		Incorrect first response	
Consensus category		No change (11)	Change (10)	No change (00)	Change (01)
CC	*N*	350	51	86	73
	% within CC	62.5%	9.1%	15.4%	13.0%
	% within first response	87.3%	12.7%	54.1%	45.9%
NC	*N*	202	79	193	86
	% within NC	36.1%	14.1%	34.5%	15.4%
	% within first response	71.9%	29.1%	69.2%	30.8%
CW	*N*	80	97	351	32
	% within CW	14.3%	17.3%	62.7%	5.7%
	% within first response	45.2%	54.8%	91.6%	8.4%
Total	*N*	632	227	630	191
	% within total	37.6%	13.5%	37.5%	11.4%

As can be seen, 75.1% of all initial responses were not changed (responses in 11 and 00 categories) and 24.9% of responses were changed. This is a standard finding, that participants are reluctant to change their initial responses (Bago & De Neys, 2017). Furthermore, responses in 11 category are most frequent for CC items, while responses in 00 category are most common for CW items, as expected.

We examined the probability of answer change with respect to the accuracy of the first response and the consensus categories. We hypothesized that nonconsensual responses would be more likely to be changed than consensual responses. As can be seen from Table 1, the percentages of changes of consensual responses in CC and CW categories (that is, initial correct responses for CC items and initial incorrect responses for CW items) are 12.7% and 8.4%, respectively. The percentages of changes of nonconsensual responses in CC and CW categories (initial incorrect responses for CC items and initial correct responses for CW items) are 45.9% and 54.8%, respectively. The percentages of changes of correct and incorrect initial responses in the NC category are 29.1% and 30.8%, respectively. In the next section, we present the analysis of proportions of answer change in relation to response consensuality, consensus categories, and item consensus.

Direction of Change and Consensuality

For each participant, we calculated the proportion of consensual and nonconsensual initial responses that were changed. Two items were excluded from the analysis because exactly 50% of participants provided each of two possible responses. Participants were more likely to change their nonconsensual initial responses (M = 0.47, SD = 0.26) than their consensual responses (M = 0.16, SD = 0.12), $t(69)$ = 10.17, p < .001. The obtained results provide support for the main hypothesis, that is, nonconsensual responses were more likely to be changed than consensual responses.

If participants are more likely to change their nonconsensual than their consensual responses, the proportions of change of initial correct and incorrect responses should vary between consensus categories. For each participant we calculated the proportion of change of accurate and inaccurate initial responses in each consensus category. Fourteen participants were excluded from the analysis because they failed to produce at least one incorrect response to CC items, and/or at least one correct response to CW items. In order to examine the effect of consensus categories (CC, CW, NC) and the accuracy of initial response (correct, incorrect) on the proportion of answer changes, we performed two-way repeated measures ANOVA. The main effects of consensus category, $F(2,110) = 1.90$, $p > .05$, and accuracy of the first response, $F(1,55) = 1.28$, $p > .05$, were not significant. However, their interaction was significant, $F(2,110) = 50.04$, $p < .001$, $\eta_p^2 = .48$. As expected, the analysis of simple effects revealed that for CC items, the proportion of answer changes was higher for incorrect first responses ($M = 0.50$, $SD = 0.42$) than for correct first responses ($M = 0.15$, $SD = 0.15$), $F(1,55) = 37.59$, $p < .001$, $\eta_p^2 = .41$. For CW items, the proportion of answer changes was higher for correct first responses ($M = 0.59$, $SD = 0.33$) than for incorrect first responses ($M = 0.13$, $SD = 0.16$), $F(1,55) = 75.08$, $p < .001$, $\eta_p^2 = .58$. For NC items there were no differences in the proportion of answer changes between correct ($M = 0.28$, $SD = 0.27$) and incorrect first responses ($M = 0.31$, $SD = 0.26$), $F < 1$.

The observed differences in the proportion of answer changes between correct and incorrect initial responses affected overall accuracy of final responses in three consensus categories. Differences in accuracy between three consensus categories and response stage (initial and final responses) were analysed by two-way repeated measures ANOVA. The interaction between consensus categories and response stage was significant, $F(2,138) = 14.27$, $p < .001$, $\eta_p^2 = .17$. For CW items, final responses provided after rethinking had lower accuracy ($M = 0.20$, $SD = 0.18$) than initial responses ($M = 0.32$, $SD = 0.19$), $F(1,69) = 45.95$, $p < .001$, $\eta_p^2 = .40$. For CC items, accuracy of final responses ($M = 0.76$, $SD = 0.19$) was not significantly higher than accuracy of initial responses ($M = 0.72$, $SD = 0.19$), $F(1,69) = 3.04$, $p = .09$, $\eta_p^2 = .04$. In the set of NC items there were no differences in accuracy between initial ($M = 0.50$, $SD = 0.17$) and final responses ($M = 0.51$, $SD = 0.16$), $F < 1$.

In the next analysis, we focused on the relationship between item consensus and probability of change of consensual and nonconsensual responses. For each of the 24 items, we calculated item consensus (the proportion of participants who endorsed the consensual initial response), the proportion of change of initial consensual responses, the proportion of change of initial nonconsensual responses, and the proportion of change of all responses. Two items were excluded from the analysis because exactly 50% of participants provided each of two possible responses. Figure 1 shows the percentages of change of the first response, separately for consensual and nonconsensual responses, and overall. For the sake of clarity, items were classified into four categories. First category included items with consensus between 51% and 60%, the second category included items with consensus between 61% and 70%, the third category included items with consensus between 71% and 80%, and the fourth category included items with consensus higher than 81%.

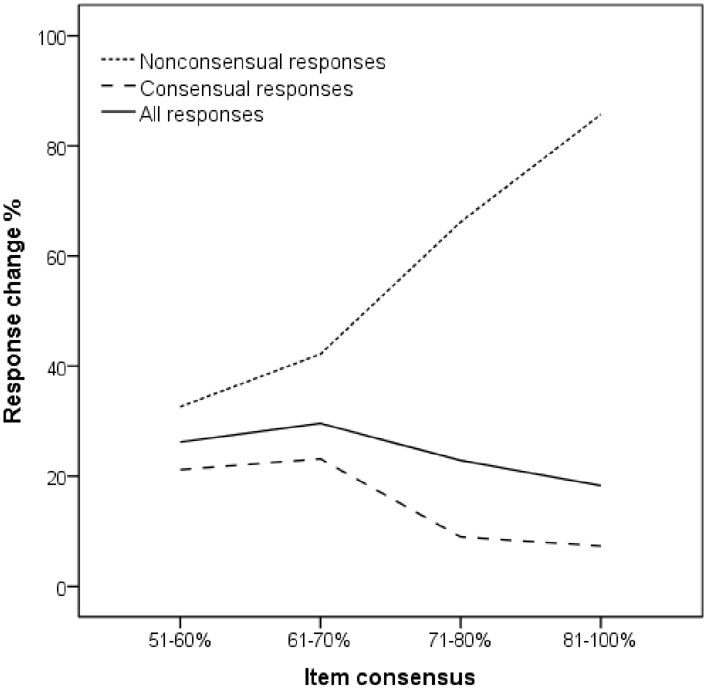

Figure 1. Percentage of answer changes with respect to item consensus.

The correlational analysis by items revealed that item consensus was negatively correlated with the overall proportion of change, r = -.56, p < .01. Items with higher consensus had lower probability of answer change. The probability of change of initial consensual responses also decreased with item consensus, r = -.56, p < .01. The proportion of answer changes of nonconsensual responses was highly correlated with item consensus, r = .88, p < .001. As the item consensus increased, the participants were more likely to change their nonconsensual responses.

We also calculated differences in accuracy between final and initial responses for each item. Difference score ranged between -.19 and +.10. This difference correlated positively with accuracy of initial responses, r = .44, p < .05. Therefore, items with low initial accuracy, or CW items, tended to have lower accuracy of final responses compared to initial responses. Items with high initial accuracy, or CC items, tended to have higher accuracy of final responses.

The Analyses of Metacognitive Judgments and Response Times

The final set of analyses was focused on metacognitive judgments and response times. We examined the role of initial response consensuality and response consistency in metacognitive judgments and response times. We classified each initial response with respect to consensuality (consensual, nonconsensual) and consistency (consistent, inconsistent). It should be noticed that consistent response included both 11 and 00 responses, and inconsistent responses included both 01 and 10 responses. For each participant we computed mean metacognitive judgments (FOR and FJC) and mean response times for the first and the second response, for each of four combinations of consensuality and consistency. Nineteen participants were excluded from the analyses because they failed to provide at least one response in each of the four combinations of consensuality and consistency. Descriptive data are presented in Table 2.

Table 2

Mean Metacognitive Judgments and Response Times (SDs are Given in Parentheses)

	Consensual first response		Nonconsensual first response	
	Consistent	Inconsistent	Consistent	Inconsistent
FOR	70.92 (12.09)	67.09 (12.16)	68.62 (12.61)	62.68 (11.58)
FJC	86.38 (9.81)	77.67 (11.20)	81.03 (12.03)	82.97 (9.57)
RT1(sec)	10.02 (3.89)	10.48 (5.99)	10.15 (3.52)	10.80 (4.89)
RT2(sec)	15.70 (6.23)	22.30 (11.16)	18.39 (8.58)	19.35 (7.99)

Note. Non-transformed response times are presented.

The data were analysed with two-way repeated measures ANOVAs. To analyse response times, $\log_{10}$ transformation was used.

FOR was higher for consensual (*M* = 69.01, *SD* = 11.08) than for nonconsensual responses (*M* = 65.65, *SD* = 10.89), *F*(1,50) = 16.34, $p < .001$, $\eta_p^2 = .25$, and it was higher for consistent (*M* = 69.77, *SD* = 11.84) than for inconsistent responses (*M* = 64.89, *SD* = 10.83), *F*(1,50) = 18.00, $p < .001$, $\eta_p^2 = .27$. The interaction between consensuality and consistency was not significant, *F*(1,50) = 1.52, $p > .05$.

For the response times of initial responses, neither the main effects of consensuality and consistency nor their interaction approached significance (all *ps* > .12). However, it should be noted that initial response times were negatively correlated with FOR in the analysis by items, r = -.70, $p < .001$, indicating the effect of fluency of processing on FOR.

FJC was higher for consistent responses (*M* = 83.70, *SD* = 9.91) than for inconsistent responses (*M* = 80.32, *SD* = 8.81), *F*(1,50) = 11.97, $p < .01$, $\eta_p^2 = .19$. Although the main effect of consensuality of initial response on FJC was not significant, $F < 1$, the interaction between consistency and consensuality was significant, *F*(1,50) = 20.47, $p < .001$, $\eta_p^2 = .29$. Similar pattern of results was obtained for response times of final responses. Final consistent responses were given more quickly (M = 17.05, SD = 6.46) than final inconsistent responses (M = 20.83, SD = 8.21), F(1,50) = 26.53, p < .001, η_p^2 = .35. Again, although the main effect of consensuality of initial response was not significant, F < 1, the interaction between consistency and consensuality was significant, F(1,50) = 13.37, p < .01, η_p^2 = .21. The analysis of simple effects confirmed that consensual consistent responses had higher FJC than nonconsensual consistent responses, F(1,50) = 16.34, p < .001, η_p^2 = .25, and were given more quickly than nonconsensual responses, F(1,50) = 5.40, p < .05, η_p^2 = .10. Final inconsistent consensual responses (that is, final responses in trials in which initial nonconsensual responses were changed) were given more quickly than final inconsistent nonconsensual responses, F(1, 50) = 6.29, p < .05, η_p^2 = .11, and they were given higher FJC ratings, F(1,50) = 11.61, p < .01, η_p^2 = .19.

Discussion

In this study, we analysed the effects of response consensuality on patterns of answer change by using the two-response paradigm with syllogistic problems. We observed that the probability of change of initial response is related to the consensuality of that response. More precisely, there are three key findings.

First, initial consensual responses, that is, responses endorsed by a larger proportion of participants, were more likely to be repeated as final responses. Initial nonconsensual responses were more likely to be changed. Furthermore, the probability of answer change correlated with item consensus. As the item consensus increases, there is a larger proportion of consensual responses, and the probability of change of those responses decreases. On the contrary, as the proportion of nonconsensual responses decreases, the probability of change of those responses increases.

Second, when all items were taken into the analysis, normative accuracy of initial responses was not related to the probability of answer change, or, to response consistency. Participants were equally likely to change their initial correct and incorrect responses across the whole set of items. However, the proportion of responses in inconsistent 10 and 01 categories differed between three consensus categories. For NC items, participants were equally likely to change their initial correct and incorrect responses. For CC and CW items participants were more likely to change their nonconsensual responses than their consensual responses. As a consequence, there is a higher probability of change of nonconsensual initial incorrect responses than consensual initial correct responses given to CC items, and the opposite pattern holds for CW items. We also observed that the accuracy of final responses tended to change toward consensual responses, or, to increase for CC items and to decrease for CW items, although this change was small. Therefore, observed changes of responses were substantially associated with consensuality of initial responses, regardless of their normative accuracy.

Third, the findings about metacognitive judgments (FOR and FJC) and response times largely replicate the results of previous studies (Bago & De Neys, 2017; Shynkaruk & Thompson, 2006; Thompson & Johnson, 2014; Thompson et al., 2011), providing further evidence for the robustness of these results. In short, FOR judgments predicted the probability of answer change and rethinking times. Both types of metacognitive judgments were higher for consistent than for inconsistent responses. Although we did not find the correlation between fluency and the probability of answer change, fluency was correlated with FOR. In addition, our study showed that metacognitive judgments and response times of final responses also track consensuality of initial responses. Namely, both types of metacognitive judgments were lower for nonconsensual than for consensual responses. When initial nonconsensual responses were changed to consensual responses, rethinking was associated with higher confidence and it was faster than when initial consensual responses were changed to nonconsensual responses.

All observed effects clearly correspond to the findings reported by Koriat in his studies that included multiple presentations of identical items (Koriat, 2008, 2011, 2013; Koriat & Adiv, 2011, 2012; Koriat & Sorka, 2015). In these studies consensual responses were more likely to be consistent, that is, to be repeated on the following encounters with the item. Consistent responding was also associated with higher confidence and shorter response latencies. It can be stressed again that these findings are very reliable across domains as different as general knowledge, perceptual judgments, personal preferences, judgments of category membership, attitudinal judgments, and social beliefs. An interpretation of these results within the SCM framework is based on the hypothetical process of representation sampling from the shared pool of representations associated with an item. Larger the proportion of representations that favours the consensual response, the process of representation sampling will result in a larger proportion of participants endorsing the consensual response, and in higher consistency of responding at the intraindividual level.

We previously demonstrated that confidence judgments in syllogistic reasoning systematically varied with respect to item consensus and response consensuality (Bajšanski et al., 2018). The results of the present study showed that the probability of repeating (and changing) the response provided at the first response stage, is related to consensuality in a similar way, just as it can be expected on the basis of the SCM. These findings provide further support for the applicability of the SCM in the domain of reasoning, and for the generality and robustness of the model.

Why people sometimes change their initial response when given an opportunity to solve the problem for the second time? According to the interpretation by Thompson et al. (2011; see also Thompson, 2009; Thompson & Johnson, 2014), answer change is primarily determined by the intervention of Type 2 processes. Low FOR has a crucial role in the willingness to engage deliberate rethinking. As a consequence, more time is spent on rethinking, and, occasionally, different final response is given. There is strong evidence that supports this view: low FOR is related to longer rethinking times, and to higher probability of answer change.

Our results offer an additional possible route to answer change: random fluctuations in representation sampling, or generating evidence for each of the two response options. Different response opportunities may lead to different responses as a consequence of different outcomes of the sampling process, and resulting evidence favouring each of the two response options. If this is the case, answer change will occur even without the intervention of qualitatively different Type 2 processes.

A negative correlation between FOR and rethinking times may indicate engagement of Type 2 processes. On the other hand, this correlation may also reflect the effects of self-consistency. Koriat (2012) suggested that the duration of the sampling process is determined by the consistency of sampling, and, as a consequence, that more consistent sampling terminates faster. For example, when engaging items with low self-consistency, initial responses should be given low FOR and they should have longer response latencies. During the second response opportunity, these items should also elicit longer response times. Therefore, the relation between FOR and rethinking times does not have to be a causal one. As a further piece of evidence, when our participants were changing their responses, they were faster when their second response was consensual than when it was nonconsensual.

Following the same logic, the relationship between FOR and answer change may also be a consequence of fluctuations in the sampling process. Items with low self-consistency should have lower item consensus, they should be assigned lower FOR, and they should have a higher probability of response change. This speculation is supported by the strong negative correlation between item consensus and the probability of answer change of consensual responses, and a strong positive correlation between item consensus and probability of answer change of nonconsensual responses.

To summarize, we speculate that there are two possible routes to answer change. First route is via FOR and analytic engagement. Responses that are generated without strong evidence will be accompanied by the feeling of metacognitive uncertainty (Quayle & Ball, 2000; Thompson, 2009) which will trigger more careful rethinking and will engage analytic Type 2 processes, with rationalization and decoupling as potential outcomes (Evans & Stanovich, 2013; Pennycook, Fugelsang, & Koehler, 2015). Second route is via random fluctuations in representation sampling. Responses that are generated without strong evidence will often be nonconsensual responses, responses given to items with low item consensus, and they will be followed by low FOR. Second presentation of the item will restart the sampling process, with all the consequences described before. However, it should be also noticed that generating the second response is surely not independent from the initial response, since participants can base their second decision on the remembered initial response.

To connect the results of this study to the literature on dual-process theory (Evans & Frankish, 2009; Kahneman, 2003; Sloman, 1996; Stanovich, 1999), the obtained results can be related to current dual process models that propose multiple Type 1 processes (Bago & De Neys, 2017; De Neys, 2017; Pennycook et al., 2015). According to these models, when participants attempt to solve conflict problems, Type 1 processes can cue several responses or give rise to two different intuitions, one logical, and one heuristic. Comparative strength of initial intuitions determines the initial response, as well as the engagement of Type 2 processes. In particular, according to Bago and DeNeys (2017), stronger intuition determines which of two response options will be endorsed. The relative strength of two intuitions determine the probability of answer change: larger the difference in strength between two intuitions, smaller the probability of answer change. Therefore, normative responses are not reachable only by Type 2 processes, but by quick intuitions as well. Similarly, in a recent study, Newman, Gibb, and Thompson (2017) showed that quick responses were affected by the rules of probability, and slow responses by belief-based information, contrary to common assumption that belief-based reasoning is fast and characteristic for Type 1 processes, and that rule-based reasoning is slow and requires Type 2 processes. There is a growing body of research with similar findings (De Neys, 2014; De Neys & Glumicic, 2008; Dujmović & Valerjev, 2018; Handley, Newstead, & Trippas, 2011).

These findings and theoretical accounts can be related to the view that initial responses are determined by the strength of evidence favouring each response option. Importantly, we speculate that this evidence is obtained by sampling from the shared pool of representations, both for initial and final responses. The hypothetical pool of representations comprises all possible pieces of evidence that may affect response choice. It can be further argued that task demands and individual differences may affect the availability of different pieces of evidence. For example, instructions to answer quickly may increase availability of belief-based information (Evans & Curtis-Holmes, 2005), working memory load may affect the availability of rule-based information (De Neys, 2006), and individual differences in cognitive capacity and disposition to think analytically may affect the sensitivity to normative information (Stanovich & West, 1998; Thompson & Johnson, 2014; Toplak, West, & Stanovich, 2011; West, Toplak, & Stanovich, 2008).

As a final consideration, the results of this study have methodological implications. The obtained results indicate self-consistency as a factor that affects answer change. Inconsistent responding may not be a consequence of the engagement of Type 2 processes and careful rethinking, but a consequence of random fluctuations in generating evidence during two encounters with an item. Therefore, additional controls are needed to differentiate the potential effects of Type 2 processes from random fluctuations in decisions about the options. For example, results obtained with two-response paradigm could be compared to the consistency of responding on multiple opportunities to give intuitive responses. Another methodological implication concerns properties of items used. Across different domains, items with different consensus elicit different patterns of responding, including differences in metacognitive judgments, consistency of responding, and response times. Therefore, careful examination of properties of items used, including distributions of consensual and nonconsensual responses and their accuracy should be an important initial step in studying metacognitive judgments and consistency of responding in the domain of reasoning.

To conclude, the results of this study provide further support for the relevance of the SCM in a domain of reasoning. Patterns of answer change, as well as metacognitive judgments and response times, were clearly related to item consensus and response consensuality. However, our findings are of limited generality, since we used only one type of problems, syllogistic reasoning problems, to test our predictions. Further studies should examine the generalizability of obtained results, using different tasks.

References

Ackerman, R., & Thompson, V. A. (2015). Meta-reasoning: What can we learn from meta-memory? In A. Feeney & V. A. Thompson (Eds.), *Reasoning as memory* (pp. 164-178). Sussex, UK: Psychology Press.

Ackerman, R., & Thompson, V. A. (2017). Meta-reasoning: Monitoring and control of thinking and reasoning. *Trends in Cognitive Sciences, 21*(8), 607-617. doi:10.1016/j.tics.2017.05.004

Bago, B., & De Neys, W. (2017). Fast logic?: Examining the time course assumption of dual process theory. *Cognition, 158,* 90-109. doi:10.1016/j.cognition.2016.10.014

Bajšanski, I., Žauhar, V., & Valerjev, P. (2018, in press). Confidence judgments in syllogistic reasoning: The role of consistency and response cardinality. *Thinking and Reasoning*, 1-34. doi:10.1080/13546783.2018.1464506

De Neys, W. (2006). Dual processing in reasoning: Two systems but one reasoner. *Psychological Science, 17,* 428-433. doi:10.1111/j.1467-9280.2006.01723.x

De Neys, W. (2014). Conflict detection, dual processes, and logical intuitions: Some clarifications. *Thinking & Reasoning, 20,* 169-187. doi:10.1080/13546783.2013.854725

De Neys, W. (2017). Bias, conflict, and fast logic: Towards a hybrid dual process future? In W. De Neys (Ed.), *Dual process theory 2.0* (pp. 47-65). Oxon, UK: Routledge.

De Neys, W., & Glumicic, T. (2008). Conflict monitoring in dual process theories of thinking. *Cognition, 106,* 1248-1299. doi:10.1016/j.cognition.2007.06.002

Dujmović, M., & Valerjev, P. (2018). The influence of conflict monitoring on meta-reasoning and response times in a base rate task. *Quarterly Journal of Experimental Psychology, 71*(12), 2548-2561. doi:10.1177/1747021817746924.

Evans, J. St. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences, 7*(10), 454-459. doi:10.1016/j.tics.2003.08.012

Evans, J. St. B. T., & Curtis-Holmes, J. (2005). Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning. *Thinking & Reasoning, 11,* 382-389. doi:10.1080/13546780542000005

Evans, J. St. B. T., & Frankish, K. (2009). *In two minds: Dual processes and beyond.* Oxford: Oxford University Press.

Evans, J. St. B. T., Handley, S. J., Harper, C. N. J., & Johnson-Laird, P. N. (1999). Reasoning about necessity and possibility: A test of the mental model theory of deduction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(6), 1495-1513. doi:10.1037/0278-7393.25.6.1495

Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science, 8,* 223-241. doi:10.1177/1745691612460685

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58,* 697-720. doi:10.1037/0003-066X.58.9.697

Handley, S. J., Newstead, S. E., & Trippas, D. (2011). Logic, beliefs, and instruction: A test of the default interventionist account of belief bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 28-43. doi:10.1037/a0021098

Koriat, A. (2008). Subjective confidence in one's answers: The consensuality principle. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 945-959. doi:10.1037/0278-7393.34.4.945

Koriat, A. (2011). Subjective confidence in perceptual judgments: A test of the self-consistency model. *Journal of Experimental Psychology: General, 140*, 117-139. doi:10.1037/a0022171

Koriat, A. (2012). The self-consistency model of subjective confidence. *Psychological Review, 119,* 80-113. doi:10.1037/a0025648

Koriat, A. (2013). Confidence in personal preferences. *Journal of Behavioral Decision Making, 26*, 247-259. doi:10.1002/bdm.1758

Koriat, A., & Adiv, S. (2011). The construction of attitudinal judgments: Evidence from attitude certainty and response latency. *Social Cognition*, *29,* 577-611. doi:10.1521/soco.2011.29.5.577

Koriat, A., & Adiv, S. (2012). Confidence in one's social beliefs: Implications for belief justification. *Consciousness and Cognition*, *21,* 1599-1616. doi:10.1016/j.concog.2012.08.008

Koriat, A., & Adiv, S. (2016). The self-consistency theory of subjective confidence. In J. Dunlosky & S. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 127-147). New York: Oxford.

Koriat, A., & Sorka, H. (2015). The construction of categorization judgments: Using subjective confidence and response latency to test a distributed model. *Cognition*, *134*, 21-38. doi:10.1016/j.cognition.2014.09.009

Newman, I. R., Gibb, M., & Thompson, V. A. (2017). Rule-based reasoning is fast and belief-based reasoning can be slow: Challenging current explanations of belief-bias and base-rate neglect. *Journal of Experimental Psychology: Learning, Memory & Cognition, 43*(7), 1154-1170. doi:10.1037/xlm0000372

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015). What makes us think? A three-stage dual-process model of analytic engagement. *Cognitive Psychology, 80*, 34-72. doi:10.1016/j.cogpsych.2015.05.001

Pennnycook, G., & Thompson, V. A. (2012). Reasoning with base-rates is routine, relatively effortless and context-dependent. *Psychonomic Bulletin & Review*, *19*(3), 528-534. doi:10.3758/s13423-012-0249-3

Quayle, J. D., & Ball, L. J. (2000). Working memory, metacognitive uncertainty and belief bias in syllogistic reasoning. *Quarterly Journal of Experimental Psychology, 53A,* 1202-1223. doi:10.1080/02724980050156362

Shynkaruk, J. M., & Thompson, V. A. (2006). Confidence and accuracy in deductive reasoning. *Memory & Cognition, 34*, 619-632. doi:10.3758/BF03193584

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*(1), 3-22. doi:10.1037/0033-2909.119.1.3

Stanovich, K. E. (1999). *Who is rational?: Studies of individual differences in reasoning.* Mahwah, NJ: Lawrence Erlbaum Associates.

Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General, 127*(2), 161-188. doi:10.1037/0096-3445.127.2.161

Thompson, V. A. (2009). Dual process theories: A metacognitive perspective. In J. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 171-195). Oxford, UK: Oxford University Press.

Thompson, V. A., & Johnson, S. J. (2014). Conflict, metacognition, and analytic thinking. *Thinking & Reasoning, 20*(2), 215-244. doi:10.1080/13546783.2013.869763

Thompson, V. A., Evans, J. St. B. T., & Campbell, J. I. C. (2013). Matching bias on the selection task: It's fast and it feels good. *Thinking & Reasoning, 19*(3-4), 431-452. doi:10.1080/13546783.2013.820220

Thompson, V. A., Prowse Turner, J. A., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology, 63,* 107-140. doi:10.1016/j.cogpsych.2011.06.001

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition, 39,* 1275-1289. doi:10.3758/s13421-011-0104-1

West, R. F., Toplak, M. E., & Stanovich, K. E. (2008). Heuristics and biases as measures of critical thinking: Associations with cognitive ability and thinking dispositions. *Journal of Educational Psychology, 100,* 930-941. doi:10.1037/a0012842

Received: February 11, 2019