Feedback Influences Syllogistic Strategy: An Analysis based on Joint Nonnegative Matrix Factorization

Nicolas Riesterer* **(riestern@cs.uni-freiburg.de)**
Daniel Brand* **(daniel.brand@cognition.uni-freiburg.de)**
Marco Ragni (ragni@cs.uni-freiburg.de)
Cognitive Computation Lab, University of Freiburg
Georges-Koehler-Allee 52, 79110 Freiburg, Germany

Abstract

Feedback for drawn inferences can lead to an adaption of future responses and underlying cognitive mechanisms. This article presents a reanalysis of recent hypothesis-driven experiments in syllogistic reasoning in which participants were presented with different feedback conditions (no feedback, 1s, 10s). We extend the original analysis, which only focused on no feedback vs. 1s feedback, by including the additional 10s condition. For our analysis, we rely on the data-driven theory- and hypothesis-agnostic Joint Nonnegative Matrix Factorization which allows us to contrast datasets based on the extraction of response patterns reflecting common and distinct response behavior. Our results support the previous claims that feedback does not generally boost logical reasoning ability but reduces the influence of biases against the response indicating that nothing logically follows from the premises.

Keywords: syllogistic reasoning; feedback; joint nonnegative matrix factorization; nvc bias

Introduction

It is a well-established fact that individuals do not only differ with respect to the strategies they employ but are also capable to adapting them to the problems they are confronted with (e.g., Bucciarelli & Johnson-Laird, 1999; Roberts et al., 2001). One consequence of this manifests, for instance, in terms of the effects different instructions may have on participants' performances (Dickstein, 1975).

Since the general goal of investigating reasoning is to understand the processes underlying human inference, exploring these adaption capabilities is an important paradigm of research. One approach for this goal is via feedback. A recent study (Dames et al., in press) investigated the effects of feedback in the domain of human syllogistic reasoning, which is concerned with inferences based on quantified relations (e.g., Khemlani & Johnson-Laird, 2012). In the study, participants were presented with feedback about the correctness of their conclusions after each task. The results suggest that feedback helps to boost logical correctness of responses and leads to post-error adaption effects with respect to reaction times. However, the authors also note that large parts of the improved correctness could be attributed to a substantial increase of the response "No Valid Conclusion" indicating that no quantified conclusion can logically be inferred from the premises. This casts doubt on the possible interpretation that feedback benefits logical thinking in its literal meaning.

The study provides insight into the impact of feedback on syllogistic reasoning ability on a statistical level by relying on a hypothesis-driven analysis. However, in doing so, the authors essentially apply a hypothesis-based filter to their data which could lead to additional results being left in the dark. In particular, the question remains if the reported results conclusively reflect the effects of feedback or if further influence on syllogistic response behavior remains. Investigating precisely this influence on the level of response patterns is crucial for cognitive modeling, because gaining insight into how manipulations affect human behavior on the response level could provide the information necessary to develop improved models, both on the level of predictive accuracy and explainability.

The goal of this article is therefore to obtain insight into the effects of feedback on the level of response patterns in the domain of syllogistic reasoning. To adopt a data-analytic perspective that is unbiased with respect to theoretical assumptions about syllogistic reasoning ability, we rely on an analysis using *Joint Nonnegative Matrix Factorization* (JNMF, Kim et al., 2015), a general approach for contrasting datasets that was originally introduced in the field of information systems but has since been transferred to the domain of reasoning (Brand et al., in press). By simultaneously solving a matrix factorization problem for two data matrices, JNMF allows to directly extract patterns which are common or distinct to the two input datasets. As such, the results of JNMF, in terms of interpretability and expressiveness, go beyond what single-dataset factorization methods can offer (Kim et al., 2015).

The remainder of this article is structured into four parts. First, we introduce the relevant background literature about syllogistic reasoning and the application of JNMF. Second, we present the methodology of our analysis. Third, we present the results, i.e., the extracted response patterns and interpret them in terms of the effects of feedback. A discussion of the implications of our analysis concludes the article.

Related Work

Syllogistic reasoning is one of the central domains investigated in research of human deductive reasoning ability (e.g., Johnson-Laird & Byrne, 1991). A syllogism consists of two premises which specify quantified relationships (using quantifiers *All*, *Some*, *No*, *Some ... not*) between three categorical terms (e.g., A, B, C):

No A are B
Some B are not C

What, if anything, follows?

The goal in syllogistic reasoning is to use the middle term, B, which occurs in both premises to infer information about the remaining two terms, A and C (end terms). In total, the syllogistic reasoning domain consists of 64 distinct problems which are obtained from the 16 possible combinations of premise quantifiers

*Both authors contributed equally to this manuscript.

and the four possible arrangement of terms in the premise (the so-called *figures*; see Khemlani & Johnson-Laird, 2012). Each syllogistic problem has nine possible propositional conclusions, eight of which relate the end terms (in either direction) using one of the four quantifiers, and the conclusion "No Valid Conclusion" (NVC) to indicate that, in accordance to first-order logic, no quantified conclusion follows from the premises.

A core result of syllogistic reasoning research is that human inferences deviate substantially from what classical logic would predict (e.g., Khemlani & Johnson-Laird, 2012). Because of this, research in the domain has focused in large parts on the development of high-level cognitive theories and corresponding model implementations (for a review, see Khemlani & Johnson-Laird, 2012). Recently, the interest in the effects of interindividual differences has been rekindled by studies focusing on different subgroups of reasoners (Khemlani & Johnson-Laird, 2016) and analyses of the shortcomings of current models when subjected to individual response data (Riesterer et al., 2019) being published.

Analyzing individual human reasoning behavior, it could be shown that human performance in syllogistic reasoning tasks is far from robust. If participants respond to syllogisms in two sessions one week apart from each other, logical correctness improves even though participants are not provided with feedback to their responses (Johnson-Laird & Steedman, 1978; Ragni et al., 2018). Similarly, within the sequence of 64 syllogistic problems, it can be observed that the likelihood of giving NVC conclusions rises as a function of presented problems, causing logical correctness to rise (Ragni et al., 2019).

To investigate the susceptibility to changes in syllogistic response behavior, a recent study adopted a paradigm in which reasoners were presented with immediate feedback for different durations indicating the correctness of their responses (Dames et al., in press). Analyzing the resulting response data, it could be shown that participants who were not provided with feedback tended to give less logically correct conclusions than participants in the feedback group. Additionally, it could be shown that feedback induced post-error adaption effects causing reaction times to slow down. The results suggest that participants are capable of adapting their response behavior in light of feedback. However, the authors also note that participants who received feedback responded substantially more often with NVC conclusions. They argue that this effect could be due to a bias or aversion against NVC responses which is overcome by providing feedback. This is a hypothesis that has been discussed, albeit inconclusively, in the literature of syllogistic reasoning before (e.g., Revlis, 1975; Roberts et al., 2001).

In the following analyses, we want to push the statistical work of Dames et al. (in press) one step further to obtain results on the level of behavioral patterns which could lead to information useful for improving models of human syllogistic reasoning. To this end, we base our analysis on contrasting.

Joint Nonnegative Matrix Factorization

Contrasting refers to the problem of finding iconic distinction factors that best describe the differences between datasets. Trivially, computing differences is one way of performing contrasting. However, given structurally rich data that are potentially noisy, the results of trivial contrasting is lacking with respect to their potential for interpretation. More sophisticated contrasting is based on the results of factor analyses. For example, by performing *Principal Component Analysis* (PCA; e.g., Murphy, 2012), the dimensionality of data can effectively be reduced to a smaller number of k latent features which can then serve as the basis for dataset comparison. However, for two independent applications of PCA such as for two separate datasets in a contrasting scenario, there is no guarantee that the resulting factorizations are related and support comparison. Especially if differences between datasets are expected to be small, it is important to factor out the commonalities in order to expose the crucial distinctions.

Motivated by this problem, work in the field of information systems has developed *Joint Nonnegative Matrix Factorization* (JNMF; Kim et al., 2015), an approach for contrasting two datasets via matrix decomposition. To achieve this, JNMF extends regular *Nonnegative Matrix Factorization* (NMF; see Pauca et al., 2004) by simultaneously searching for factors representing commonalities and distinctions between both datasets.

Formally, NMF is the problem of finding a decomposition of a single data matrix $X \in \mathbb{R}^{m \times n}$ where m is the dimensionality of the data and n denotes the number of objects in the dataset into a basis matrix $W \in \mathbb{R}^{m \times k}$ and coefficient matrix $H \in \mathbb{R}^{n \times k}$ where $k < \min\{m,n\}$ is the number of patterns, or factors, to decompose the data into, such that

$$X \approx WH^T \tag{1}$$

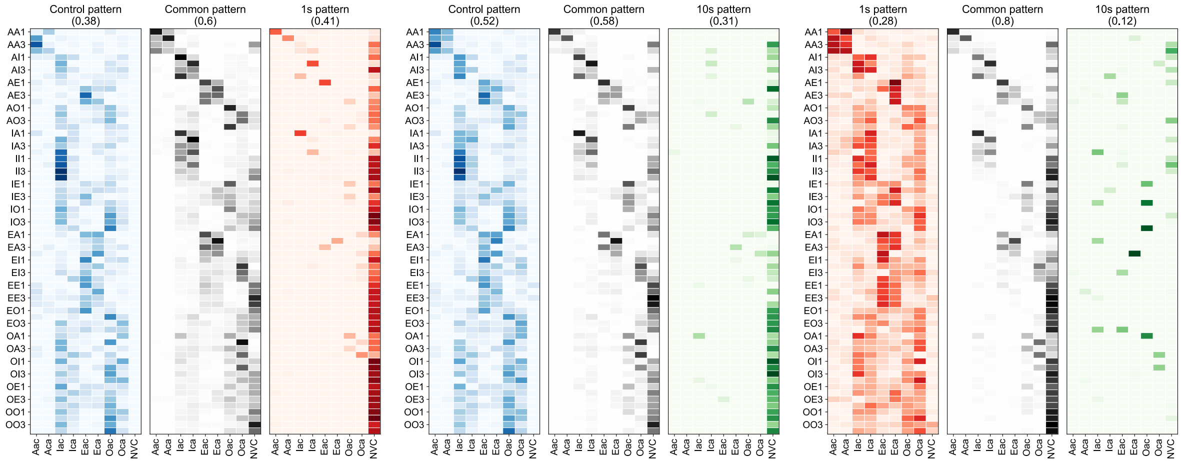
JNMF refers to the problem of finding a decomposition of two data matrices $X_1 \in \mathbb{R}^{m \times n_1}$ and $X_1 \in \mathbb{R}^{m \times n_2}$ into a basis matrix $W_i \in \mathbb{R}^{m \times k}$ and a coefficient matrix $H_i \in \mathbb{R}^{n_i \times k}$ for $i = 1,2$ (Kim et al., 2015). Crucially, $k = k_c + k_d$ refers to the total number of patterns consisting of k_c common and k_d distinct patterns which means that $W_i = [W_{i,c}, W_{i,d}]$ is composed of columns referring to k_c common ($W_{i,c}$) and k_d distinct patterns ($W_{i,d}$). The goal of JNMF is to find matrices W_1, H_1 and W_2, H_2 solving Equation (1) for both data matrices X_1 and X_2 under constraints regularizing the identified W and H matrices to ensure that the distances between common and distinct patterns are minimized and maximized, respectively (Kim et al., 2015).

An important advantage of NMF (or JNMF for that matter) for its application in the context of cognitive science is its focus on nonnegativity. Since data obtained in behavioral experimentation is usually nonnegative as well (response choices, reaction times, etc.), NMF operates directly and natively on the expected range of values which allows for better interpretability of the results (Pauca et al., 2004).

Stemming originally from the field of information systems, JNMF has recently been applied successfully to syllogistic reasoning data (Brand et al., in press). In this analysis, the authors investigated the influence of personality factors on syllogistic reasoning behavior. Using JNMF, they managed to extract the patterns distinctly representing the response behavior of participants with varying scores on personality traits. In the following analyses we apply this method to feedback data.

Method

Our goal is to investigate the effects of feedback on the responses given to syllogistic reasoning problems. To this end, we employ



(a) Control vs. 1s. (b) Control vs. 10s. (c) 1s vs. 10s.

Figure 1: Typical response patterns obtained from JNMF application and arranged as matrices of 64 rows (syllogistic problems) and 9 columns (possible conclusions). Common patterns represent the average common pattern from W_1 and W_2. The values in parentheses denote the relative importances of the patterns for the reconstruction of the dataset. These values are derived from the respective columns of the H-matrices.

JNMF in order to decompose the data into common and distinct patterns which can then be interpreted directly. Relying on JNMF allows us to adopt a theory- and hypothesis-agnostic perspective which, in turn, allows us to obtain comprehensive results which could potentially go beyond the findings of Dames et al. (in press).

Dataset

For our analysis, we rely on a dataset that was published recently as part of a study of feedback effects in syllogistic reasoning (Dames et al., in press). The focus of the study was to investigate the influence of feedback on participants' propensities to give logically valid conclusions (Dames et al., in press, Experiment 1) and on their reaction times (Dames et al., in press, Experiment 2). To this end, the authors conducted a series of experiments via Amazon Mechanical Turk in which participants were instructed to give conclusions to all 64 syllogistic problems. In total, the dataset comprises three conditions: a control group which received no feedback ($n = 39$), a group which was presented with short feedback (1s, $n = 102$), and a group which was presented with extended feedback (10s, $n = 29$). In their analysis, the authors focused on the control and 1s conditions. The effects of extended feedback have not been published up till now.

Data Preparation

To make the data accessible to JNMF analysis, we first transform the response data for each condition into matrix representations. This is achieved by onehot-encoding individual responses as zero-vectors of dimensionality 9 in which a single 1 indicates the corresponding response. As examples, for the syllogistic response "All A are C" this leads to the onehot-encoded vector $(1,0,0,0,0,0,0,0,0)$, and for "No Valid Conclusion" to $(0,0,0,0,0,0,0,0,1)$. Concatenating the 64 onehot-encoded responses of a single participant results in a 576-dimensional vector which reflects their individual response pattern, or *reasoner profile* (Riesterer et al., 2019; Brand et al., in press). By arranging all reasoner profiles as column vectors in a matrix, we obtain $m \times n$ matrices, where m is the number of features, i.e., the dimensionality of our reasoner profiles ($m = 576$) and n denotes the number of participants in our datasets ($n_1 = 39$, $n_2 = 102$, and $n_3 = 29$ for no feedback, short feedback, and extended feedback, respectively). The following analyses are computed directly on these data matrices. The data and analysis scripts are publicly available on GitHub[1].

Results

Our analyses are based on JNMF application for pairwise contrasting of the three datasets. For each pair of datasets, JNMF produces W and H matrices containing the common and distinct response patterns and their weightings for reconstructing each individual from the input data, respectively.

Pattern Analysis

The general results of the JNMF application are depicted in Figure 1. The heatmaps visualize the patterns extracted from the W-matrices in pairwise contrasting applications for the three feedback conditions: control vs. 1s (Figure 1a), control vs. 10s (Figure 1b), and 1s vs. 10s (Figure 1c). In each subplot, the patterns are presented as heatmaps with distinct patterns located left and right, and the common pattern being depicted as the mean of the common vectors from W_1 and W_2 in the middle. The shading of cells indicates the weighting of individual responses for the patterns in accordance to the values in W. The values in parentheses next to the titles denote the importances of the patterns for the reconstruction of the original dataset which were calculated from the sum of the

[1] https://github.com/nriesterer/iccm-nmffeedback

Table 1: Proportion to reassignment of individuals from source groups (rows) to target groups (columns) based on response pattern similarities. Ties were resolved by ignoring participants which causes percentages to not sum up to 1.

	control	1s	10s
control		77%	15%
1s	32%		59%
10s	14%	83%	

respective columns in the corresponding H matrix normalized by the total sum of the H matrix. These values show that, overall, commonalities are more important than distinctions which is to be expected in a complex task such as syllogistic reasoning.

Both, the contrasting between control and 1s (Figure 1a), and between control and 10s (Figure 1b) show that the distinct differences between the datasets manifest in terms of the dominance of NVC responses in the feedback groups. Contrasting both feedback groups (Figure 1c), a different picture emerges. Here, the 10s group yields a pattern that looks sparse and scattered in comparison to the other patterns. When taking the importance of the pattern into consideration, it becomes apparent that the JNMF assigned only low importance to the 10s pattern suggesting that the common pattern suffices to reconstruct the original data.

The results suggest that the 1s group reflects a mixture between the control and 10s group. Contrasted with control, it appears similar to the 10s pattern. However, when contrasted with 10s, it appears similar to the control pattern. Evidence supporting this assumption can be obtained from the similarities between the response behaviors of individuals which is represented in Table 1. For a participant from one of the conditions (denoted as source group), we checked which of the other conditions (target groups) contained the individual most similar to them. The values therefore represent proportions to which individuals from one group prefer another in terms of similar response behavior (ties were resolved by ignoring the participants which is why values do not sum to 100%). The results illustrate the special role of the 1s group. While both control and 10s clearly favor 1s over another, 1s is much more evenly distributed which indicates that it consists both of participants showing feedback behavior and participants who are still unaffected by it. In case of 10s, the proportion of individuals performing similar to controls is reduced substantially.

Put together, the contrastings provide evidence for the NVC aversion hypothesis (e.g., Revlis, 1975). Contrasted against control, the distinct patterns of both, 1s and 10s focus chiefly on the NVC response. However, when comparing the feedback patterns resulting from contrasting with control, it appears as if the NVC dominance is stronger for 1s than for 10s which could hint at a time-dependent effect of feedback. As previously concluded by Dames et al. (in press), given short (1s) feedback, a lot of participants seem to quickly grasp the importance of NVC. Combined with the evidence obtained from the extended (10s) condition, it appears as if the lack of time to reflect the meaning of NVC results in participants overestimating the frequency of invalid syllogisms. There still remain some individuals, though, who are unaffected by feedback resulting in 1s representing a mixture between reasoning patterns related to the control and 10s groups. In case of extended feedback, participants are given time to reflect their use of NVC which may lead to a more deliberative and careful reliance on this response affecting most of the individuals in the data. Consequently, when contrasting between both feedback conditions, due to the similarities of their NVC reliance, JNMF mainly uses the common pattern to capture the feedback reasoning behavior (including NVC) and uses the distinct patterns to capture residual responses.

Prediction Analysis

While the previous analysis illustrated the general structure and properties of the response patterns for the different conditions of feedback, their quality still remains obscure. To evaluate this, we now interpret the obtained patterns as predictive models and subject them to an analysis in which the accuracy of JNMF patterns in accounting for individual reasoners' responses is assessed (for the predictive task, see Riesterer et al., 2019). In this analysis, we expect patterns to perform best with respect to predicting responses on their respective conditions (i.e., the control pattern on the control dataset and so on). Simultaneously, we expect patterns to perform in proportion to their importances, i.e., to the number of individuals for which the pattern is crucial when reconstructing the response data.

Figure 2 depicts the results of this analysis. Datasets on which the patterns are evaluated are depicted on the x-axis while the y-axis denotes the predictive accuracies they achieve. Each line reflects a pattern with colors indicating their respective conditions (grey, blue, red, and green representing common, control, 1s, and 10s, respectively).

On a high-level, comparing the patterns shows that, overall, the common patterns achieve the highest predictive accuracies which is in line with the findings above. Since the common patterns are most important for reconstructing the data, the distinct patterns are not expected to be good predictors of reasoning behavior on their own, while the commonalities reflecting general reasoning behavior are. Because of this, the quality of the distinct patterns should not be assessed based on absolute accuracy values but based on their ability to provide suitable predictors for their respective data conditions.

A prime example for this is the trend of the blue lines corresponding to the no-feedback control patterns. The fact that accuracy drops substantially on the feedback data indicates that the patterns are highly descriptive for the control data only and bear little meaning for the feedback groups. Additionally, the fact that control's accuracies on the feedback data is similar for 1s and 10s suggests that the differences between the two are minor. Considering the feedback patterns obtained from contrasting with control (dark red for 1s, dark green for 10s), a different picture emerges. Here, as expected, the patterns perform much better in accounting for the feedback data than for the control data. Put together, the blue, dark red, and dark green lines illustrate the clear distinction between the control and feedback groups. Distinct patterns obtained for either condition are suitable predictors for their own data but offer severely limited applicability to the other. Additionally, the high similarity between accuracies on both feedback groups suggests that JNMF application found patterns distinct to general feedback behavior regardless of the underlying feedback duration. The lines also show that 10s elicits the most consistent feedback behavior. While 10s patterns as

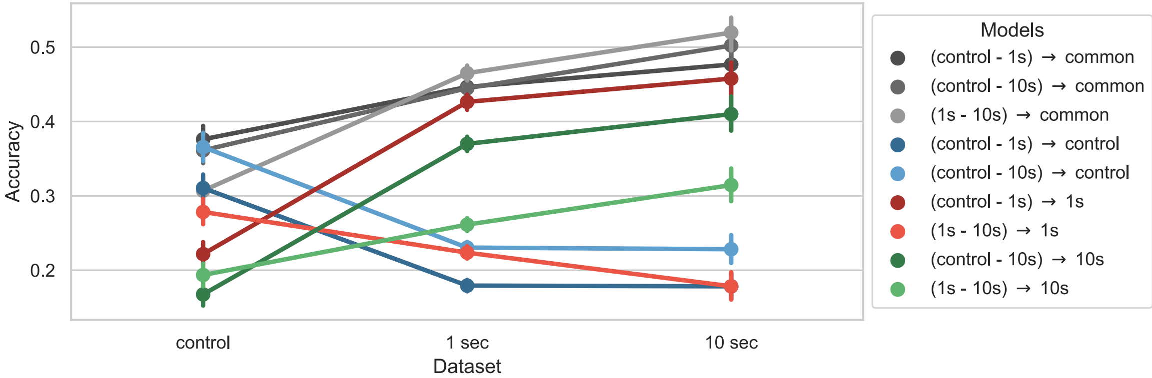


Figure 2: Accuracies of the response patterns resulting from the different contrastings (e.g., "control - 1s") interpreted as predictive models. Error bars denote 95% confidence intervals.

expected always perform best on the 10s data, this is also the case for 1s (dark red). This indicates that contrasting control with 1s does not only yield particular feedback patterns accounting for 1s, but general feedback patterns which perform even better on the 10s data.

Considering the contrasting of both feedback conditions (light red for 1s, light green for 10s), peculiar patterns emerge. The pattern for 10s (light green) performs best on the 10s data and drops substantially for the other groups thereby indicating that it captures distinct properties of the extended 10s feedback group (despite its scattered appearance in Figure 1c). The pattern for 1s (light red), however, fails to capture feedback behavior scoring higher on control data than on feedback data. Again, this indicates that 1s represents a mixed pattern. When compared to control, it clearly reflects feedback behavior. However, compared to extended feedback its distinct patterns correspond more to the control group.

The prediction analysis supports the interpretation of the results so far. The observations can be explained by assuming a time-dependent influence of feedback on reasoning behavior. Given naive reasoners, short feedback allows them to acknowledge the importance of the NVC response causing big parts of them to radically adapt their behavior in its favor (see Figure 1a) while leaving the behavior of others unchanged. Extending feedback increases the effects with diminishing returns. It appears as if the effects of extended feedback manifest in terms of a more differentiated or more deliberative use of NVC, which, at its core, is only a slight deviation from the distinct effects observable in the short feedback group. As a result, contrasting the feedback conditions leads to an overestimation of their respective differences causing the 1s pattern to be pushed towards the naive control state of reasoning behavior and the 10s pattern to focus on the very few distinct differences extended feedback results in.

Comparison of Task Performances

In a final analysis, we investigated the congruency of participant responses with formal logic in order to gain insight into whether the effects of feedback exclusively affect NVC responses as the patterns might suggest at a first glance (Figure 1) or if the effects manifest in terms of general logical correctness.

Table 2: Investigation in terms of logical correctness. The values refer to average proportions of logical correct responses and their corresponding standard errors.

Condition	Total	Valid Syllogisms	Invalid Syllogisms
Control	$(33\pm3)\%$	$(49\pm2)\%$	$(20\pm4)\%$
1s	$(46\pm2)\%$	$(44\pm1)\%$	$(47\pm2)\%$
10s	$(50\pm4)\%$	$(48\pm4)\%$	$(52\pm6)\%$

The results of this analysis are summarized in Table 2. The total correctness shows that logical performance increases with prolonged feedback. When considering invalid syllogisms, i.e., the problems which do not have a propositional conclusion, the dominance of NVC responses in the feedback condition become apparent. For the valid syllogisms, however, the changes are more complex. In line with our interpretation, participants seem to overestimate the relevance of the NVC response when presented with short feedback (1s) which results in a decrease in logical correctness. With extended feedback (10s), participants seem to be able to handle NVC responses better resulting in a performance on valid syllogisms that is similar to the control condition.

In conclusion, we agree with Dames et al. (in press) that while feedback appears to boost logical correctness at first glance, it is highly unlikely that this is due to an improvement of logical reasoning ability. Additionally, based on our pattern analysis, we conclude that the observed effects can be attributed solely to a different handling of NVC responses.

General Discussion

We investigated the effects of feedback on human reasoning behavior in the domain of syllogistic reasoning using Nonnegative Matrix Factorization (JNMF; Kim et al., 2015). We were able to replicate the findings of Dames et al. (in press) who showed that feedback improves logical correctness of participants' responses mostly due to an increase in No Valid Conclusion (NVC) response frequency on invalid syllogisms. By relying on a data-driven, theory- and hypothesis-agnostic approach and by including an extended feedback condition (10s; the authors of the original study based their analyses on the 1s condition alone), we pushed the analysis of feedback in terms of its influence on response patterns one step further.

Our results suggest that the impact of feedback depends on the duration of its presentation. Short feedback (1s) does not allow participants to properly reflect about their reasoning strategies. Instead, as Dames et al. (in press) already suggested, it only teaches them the relative importance of the NVC response which is logically correct in 37 of the 64 syllogisms (58%) and a response for which the existence of biases causing participants to reject it have frequently been assumed (e.g., Revlis, 1975). Strong evidence for this claim is found by considering the proportion of logically correct conclusions which overall increases for the short feedback condition but decreases on valid syllogisms for which NVC is an incorrect conclusion. If feedback is extended (10s), an increase of NVC responses is still the dominating distinction when compared to the control condition receiving no feedback. The logical correctness of responses to valid syllogisms remains similar, however. This suggests that the extended feedback duration allows participants to properly reflect over their reasoning strategies. As a result, the effects of increased logical correctness are in terms of invalid syllogisms only. Since performance on valid syllogisms is not affected positively by feedback (at least not significantly), in similar spirit to Dames et al. (in press), we conclude that feedback does not necessarily help to boost logical thinking in general. A more likely explanation, especially considering the differences between 1s and 10s, is, that at first feedback helps to combat an NVC aversion bias (e.g., Roberts et al., 2001) which leads to an overestimation of the relevance of NVC. Given more time to reflect, a more deliberate NVC handling with overall improved logical correctness is adopted by most reasoners.

For the general field of cognitive modeling, our findings bear relevance for two reasons. First, it is crucial to keep in mind that effects such as the lacking performance of reasoners or the corresponding increase for the feedback condition are not necessarily explained by fundamental cognitive processes. As can be shown here, rejections of particular response options must be considered by experimenters and modelers alike to ensure the proper interpretation of data and the proper development of corresponding models. Second, our analysis which was devoid of theoretical assumptions about potential inferential processes allowed us to draw comprehensive and unbiased conclusions about the available data. We strongly believe that maintaining a balance between theoretical and theory-agnostic exploration of cognition is key to ensuring a steady and uninterrupted progression of the field.

Acknowledgements

This paper was supported by DFG grants RA 1934/2-1, RA 1934/3-1 and RA 1934/4-1 to MR.

References

Brand, D., Riesterer, N., Dames, H., & Ragni, M. (in press). Analyzing the differences in human reasoning via joint nonnegative matrix factorization. In *Proceedings of the 42nd Annual Conference of the Cognitive Science Society.*

Bucciarelli, M., & Johnson-Laird, P. (1999). Strategies in syllogistic reasoning. *Cognitive Science*, *23*(3), 247–303.

Dames, H., Schiebel, C., & Ragni, M. (in press). The role of feedback and post-error adaptations in reasoning. In *Proceedings of the 42nd Annual Conference of the Cognitive Science Society.*

Dickstein, L. S. (1975). Effects of instructions and premise order on errors in syllogistic reasoning. *Journal of Experimental Psychology: Human Learning & Memory*, *1*(4), 376–384.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hillsdale, NJ: Erlbaum.

Johnson-Laird, P. N., & Steedman, M. (1978). The psychology of syllogisms. *Cognitive Psychology*, *10*(1), 64–99.

Khemlani, S., & Johnson-Laird, P. N. (2012). Theories of the syllogism: A meta-analysis. *Psychological Bulletin*, *138*(3), 427–457.

Khemlani, S., & Johnson-Laird, P. N. (2016). How people differ in syllogistic reasoning. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th annual conference of the cognitive science society* (pp. 2165–2170). Austin, TX: Cognitive Science Society.

Kim, H., Choo, J., Kim, J., Reddy, C. K., & Park, H. (2015). Simultaneous discovery of common and discriminative topics via joint nonnegative matrix factorization. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining - KDD '15.* ACM Press.

Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. Cambridge, MA: MIT Press.

Pauca, V. P., Shahnaz, F., Berry, M. W., & Plemmons, R. J. (2004). Text mining using non-negative matrix factorizations. In M. W. Berry, U. Dayal, C. Kamath, & D. Skillicorn (Eds.), *Proceedings of the 2004 SIAM International Conference on Data Mining* (pp. 452–456). Philadelphia, PA: Society for Industrial and Applied Mathematics.

Ragni, M., Dames, H., Brand, D., & Riesterer, N. (2019). When does a reasoner respond: Nothing follows? In A. K. Goel, C. M. Seifert, & C. Freksa (Eds.), *Proceedings of the 41st Annual Conference of the Cognitive Science Society* (pp. 2640–2646). Montreal, QB: Cognitive Science Society.

Ragni, M., Riesterer, N., Khemlani, S., & Johnson-Laird, P. (2018). Individuals become more logical without feedback. In T. Rogers, M. Rau, J. Zhu, & C. Kalish (Eds.), *Proceedings of the 40th Annual Conference of the Cognitive Science Society* (pp. 1584–1589). Austin, TX: Cognitive Science Society.

Revlis, R. (1975). Syllogistic reasoning: Logical decisions from a complex data base. In R. Falmagne (Ed.), *Reasoning: Representation and process.* Hillsdale, NJ: Erlbaum.

Riesterer, N., Brand, D., & Ragni, M. (2019). Predictive modeling of individual human cognition: Upper bounds and a new perspective on performance. In T. Stewart (Ed.), *Proceedings of the 17th International Conference on Cognitive Modeling* (pp. 178–183). Waterloo, Canada: University of Waterloo.

Roberts, M. J., Newstead, S. E., & Griggs, R. A. (2001). Quantifier interpretation and syllogistic reasoning. *Thinking & Reasoning*, *7*(2), 173–204.