Psychiatry Research 275 (2019) 221–227

Contents lists available at ScienceDirect

Psychiatry Research

journal homepage: www.elsevier.com/locate/psychres

The stigma of mental illness: Testing for the implicit bias in diagnostic labels

Björn Schlier*, Tania M. Lincoln

Universität Hamburg, Institute of Psychology, Clinical Psychology and Psychotherapy, Von-Melle-Park 5, 20146 Hamburg, Germany

ARTICLE INFO

Keywords:
Mental illness stigma
Implicit attitudes
Continuum beliefs
Labeling

ABSTRACT

Several models that attempt to explain stigmatization of people with mental illness emphasize the relevance of social categorization. However, research on illness-based social categories has been limited to explicit self-report measures. In this study, we explored whether implicit social categorization based on mental illness labels can be assessed with the "Who-Said-What" (WSW)-paradigm. In this paradigm, participants have to allocate a large number of statements to individuals who belong to different social groups (e.g., mentally ill vs. healthy). Due to the amount of statements, errors in allocation are to be expected. A disproportionate amount of erroneous allocation to individuals within a social group is interpreted to indicate implicit social categorization. In three studies, we (1) pilot-tested a WSW-paradigm for mental illness categorization ($n = 24$), (2) replicated these findings in an independent, larger sample ($n = 85$), and (3) aimed to rule out alternative explanations for the implicit categorization effects with a modified WSW-task and novel stimulus material ($n = 137$). We found consistent implicit categorization effects with overall medium to large effect sizes (range: $0.41 \leq d \leq 1.01$). Implicit components of mental illness based social categorization can be assessed with the WSW-paradigm. Future research needs to explore the interrelations of implicit categorization, implicit vs. explicit stigmatization and discriminating behavior.

1. Introduction

Stigmatization begins with a distinction between in-group and out-group. Mental illness labels highlight a difference between 'us' and 'them' (Corrigan et al., 2015). The psychological basis of groups (e.g., Tajfel, 1981; Tajfel and Turner, 1986) and the role of social categorization of oneself and others (Turner et al., 1987) for attitudes and intergroup behavior has been an important topic in psychology for decades. Accordingly, early theoretical frameworks of mental illness stigmatization acknowledged the role of identifying and labeling differences as a key component of stigma formation (Link and Phelan, 2001). Developmental theories explaining stigmatization propose that labeling differences plays a key role in the emergence of an in-group vs. out-group distinction that enables the attribution of negative characteristics to the respective out-groups (Bigler and Liben, 2006, 2007). Moreover, an in-group vs. out-group distinction often paves the way for a range of essentialist beliefs about social categories. Studies identified a pattern of essentialist beliefs for a variety of social categories - including mental illness – that comprised, for example, believing in the naturalness of the categories or in the uniformity of the group members (Haslam et al., 2002; Haslam, 2011). These beliefs correlated with stigmatizing attitudes towards the respective social groups.

Recent research on mental illness stigma revisited the topic of group perception, but focused on the degree to which a person agrees with the idea of a continuum of mental health vs. mental illness (Schomerus et al., 2013). In theory, such continuum beliefs constitute the opposite to a firm belief in a fixed entity "mental illness". Correspondingly, continuum beliefs about mental illness have been found to correlate with less stereotypes about mental illness and less discriminatory behavioral intentions (e.g., desire for social distance), with most pronounced findings for the mental illness label 'schizophrenia' (Schomerus et al., 2013; Wiesjahn et al., 2014; Corrigan et al., 2015; Schlier et al., 2016). However, experiments on promoting a categorical vs. continuum view of mental illness and mental health have yielded mixed results: Some studies found no results in terms of decreasing stigma (Thibodeau, 2017; Thibodeau and Peterson, 2018). Another study found stigmatizing attitudes only to decrease if the continuum beliefs intervention also involved contact to an individual with schizophrenia (Corrigan et al., 2017). Other studies found that categorical views increase stigmatizing attitudes, whereas promoting a continuum model decreases them (Schomerus et al., 2016; Wiesjahn et al., 2016; Thibodeau et al., 2018). Thus, there is strong evidence that categorization is an indicator of stigmatization and a potential opportunity to

* Corresponding author.
E-mail address: bjoern.schlier@uni-hamburg.de (B. Schlier).

https://doi.org/10.1016/j.psychres.2019.03.028
Received 9 November 2018; Received in revised form 16 March 2019; Accepted 17 March 2019
Available online 18 March 2019
0165-1781/

utilize information on categorization for anti-stigma interventions.

Despite the longstanding emphasis on the pivotal role of categorization in mental illness stigma, direct tests of the role of categorization only began in the wake of the research of continuum beliefs (e.g., in the form of control conditions to continuum beliefs anti-stigma interventions; Schomerus et al., 2016). However, this research has been limited to self-reported belief or disbelief in the veracity of mental illness based social categories (i.e., essentialist beliefs or continuum beliefs questionnaires). The implicit components of mental illness based categorization have not been investigated. Yet, evidence from adjacent fields of stigma research consistently shows that implicit attitudes constitute a stronger predictor for discriminatory behavior than explicit attitudes (Greenwald et al., 2009). Implicit attitudes are less prone to be influenced by social desirability and self-representation biases, because the evaluative nature of the paradigms is less obvious to the participants (Fazio and Olson, 2003). When compared to traditional paradigms like the implicit association test, measures of "neutral" implicit mental illness-related categories could also be less prone to confounding effects by self-presentational motives and thus likely to be more closely associated with discriminatory behavior in peoples' real-life.

The discipline of social psychology has extensively investigated a paradigm to test mental categorization processes. The statement allocation task or "Who-said-what?" (WSW) paradigm (Taylor et al., 1978) allows to quantify the prominence of social categories when processing information. In the WSW-paradigm, participants observe various members of two mutually exclusive groups providing statements about a predetermined topic. After participants have received a large amount of statements per member (e.g., 4 statements per member), they are instructed to allocate each statement to its originator. If statements by one member of a social group are more often erroneously allocated to other members of the same group (within-group error) than to members of the other group (between-group error), social categorization is interpreted to have contributed to information processing.

The WSW-paradigm has been used to investigate categorization processes by for example gender, ethnicity, sexual orientation, or attractiveness (Klauer and Wegener, 1998). To our knowledge, however, no study has tested whether it could be adapted to measure categorization due to mental illness labels and analyzed its association with subsequent stigmatization of mental illness. In this article we describe the development and validation of a WSW-paradigm over the course of three studies. In a pilot study (study 1), we tested a WSW-task on a small sample to screen for the amount and distribution of within-group errors and gain an impression of the effect sizes. In study 2, we aimed to replicate this finding with the same WSW-task in an independent and larger sample. In study 3, we tested the WSW-task with novel material to rule out influences of specific stimulus characteristics. Finally, we used study 3 to test how closely categorization measured by the WSW-task is associated with explicit mental illness stigma.

2. Study 1 – Pilot test of the WSW-task and estimating effect sizes

In the first study, we investigated whether the WSW-task can be adapted to categories based on mental illness labels. We explored which social category works best as a comparison group ("schizophrenia" vs. "no label" or "schizophrenia" vs. "depression"). Moreover, we aimed find the most promising within-group error based outcome variable for this WSW-task. Finally, we estimated the expected effect size for power calculation in the following studies.

2.1. Methods

2.1.1. Participants

Twenty-four participants were recruited via facebook-advertisement for this pilot online-study. Their mean age was 31.33 years (SD = 9.65) and 62.50% were female. Most of the participants (95.83%) indicated their education level to be the German university entrance certificate (Abitur). Half of the sample were university students (54.17%), 37.5% were employed and 8.33% were self-employed. A fifth of the sample (20.83%) reported to have had previous contact to people with schizophrenia.

2.1.2. Procedure

Upon starting the online-study, participants were told the cover-story that the study was investigating attention processes during group discussions. Participants then provided informed consent. Next, they were randomly assigned to and completed one of two WSW-tasks. Finally participants answered a brief demographic questionnaire and were debriefed about the exact purpose of the study. Participation took about 15–20 min. Participants received no monetary incentives.

2.1.3. WSW-task

Two experimental conditions of the WSW-task were used. In both conditions, participants were first shown 24 statements by eight different target persons. In order to avoid confounding categorization with recollection biases due to mental illness specific statement content, statements on an unrelated topic (gender roles in a relationship) were taken from a list of statements previously developed and tested in a WSW-task (Wegener, 2000; see Fig. 1, left column for two example statements). As pictures for the eight male target persons, a random selection of neutral, front-view photographs from the Caucasian subset of the Radboud faces repository (Langner et al., 2010) was used.

In the first experimental condition, the group consisted of four target persons with the label "diagnosed with schizophrenia" (experimental target) and four target persons without any mental illness label (comparison label targets 1). In the second condition, the group consisted of four target persons with the label "diagnosed with schizophrenia" (experimental label targets) and four target persons with the label "diagnosed with depression" (comparison label targets 2).

The WSW-task included two phases: the discussion phase and the statement allocation phase. In the discussion phase, participants saw a target person's picture (stimuli taken from Langner et al., 2010), name, age, and diagnostic label paired with one of 24 statements (see Fig. 1, left). Statements were shown in randomized order with each statement/target combination shown for 8 s. Every target person was shown with three statements during the discussion phase. Statements were allocated to targets in a balanced manner. Each target person was presented with at least one progressive and one conservative statement about the topic.

In the statement allocation phase, each statement was presented again and participants had to allocate the statement to the correct target person by selecting their picture. In order to increase the difficulty of the task, four new statements were mixed in and participants had the additional task to identify whether the statement presented was shown during the discussion (Fig. 1, right).

The full online-study (including the WSW-task) was programmed and presented to the participants with the EFS Survey web application (QuestBack GmbH, 2014).

2.1.4. Data-analysis

Spontaneous categorization was tested by analyzing the distribution of categorization errors during the statement allocation phase. Errors were classified as *within-group errors* when participants allocated a statement to a wrong target from the same label group (e.g., a statement made by person #5, "diagnosed with schizophrenia" allocated to person #7 who also had the label "diagnosed with schizophrenia") and as *between-group errors* when participants allocated a statement to a wrong target from the wrong label group (e.g., a statement made by person #5, "diagnosed with schizophrenia" allocated to person #1 who did not have the label "diagnosed with schizophrenia", see Fig. 1).

The sum-score of a participant's between-group errors was corrected by the factor 0.75, since there were three targets for a within-group error and four targets for a between-group error. Within-group and

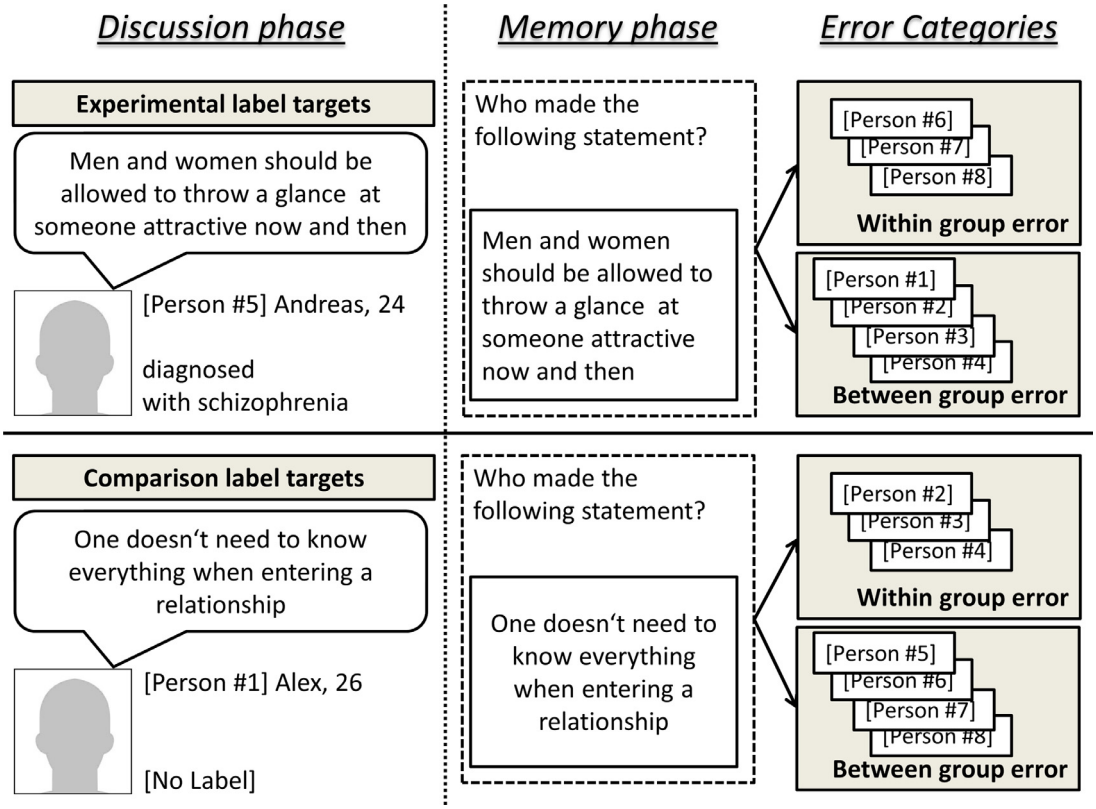

Fig. 1. Overview of the "Who-Said-What" paradigm and the error-types extracted. In this example, person #1-#4 received no label (comparison label targets 2), whereas person #5-#8 received the label "diagnosed with schizophrenia" (experimental targets).

corrected between-group errors were then entered as a within subject factor in a mixed-model ANOVA. A significant main effect of errors with significantly more within-group errors than corrected between-group errors (i.e., a WSW-effect) was interpreted to indicate the use of spontaneous categorization by mental illness labels during statement allocation. Thus, all effects of error-type with more within-group than between-group errors indicates constitute the expected effects of the WSW-paradigm.

In order to test for the optimal comparison group, we entered the experimental condition as a between subject factor into the ANOVA. A significant error-type x condition interaction effect would indicate that the WSW-effect is modulated by the type of comparison group.

Finally, we divided the within-group and between-group errors by label-type of the origin target (see Fig. 1, top vs. bottom row). We entered label-type (experimental label vs. comparison label) as another within subject factor. Interaction effects involving label-type (i.e., label-type x error-type and label-type x error-type x condition) would indicate that the WSW-effect may be limited to erroneous allocations of statements of either the "schizophrenia" and the "depression" groups (i.e., an outgroup categorization effect) or the "no label" group (i.e., ingroup categorization effect).

Because this was an exploratory pilot test to generate hypotheses on how a WSW-paradigm for mental illness labels works optimally, both statistical significance and the absolute effect size of the WSW-effect (i.e., Cohen's d based on the difference between within-group and between-group errors) were taken into account to interpret the findings.

2.2. Results

An ANOVA of errors (within-group vs. between-group) as the dependent variable, experimental condition (1: "schizophrenia" vs. "no label"; 2: "schizophrenia" vs. "depression") as between subject factor, and label type (1: "schizophrenia"; 2: "depression/no label") as well as error type (1: within-group errors; 2: between-group errors) as within subject factors yielded no significant main effects for error-type. The WSW-effect for the full sample was small ($d = 0.30$).

The largest non-significant effects in the ANOVA were observed for the label-type x error-type interaction effect ($F(1,22) = 2.02$, $p = 0.169$, $\eta^2_{partial} = 0.084$) and the condition x error-type interaction effect ($F(1,22) = 2.08$, $p = 0.163$, $\eta^2_{partial} = 0.086$). This indicated that both the comparison group as well as the label-type modulate the size of the WSW-effect. Correspondingly, the WSW-effect differed substantially between groups with a medium effect in the "schizophrenia vs. no diagnosis" condition ($d = 0.68$) and no WSW-effect in the "schizophrenia" vs. "depression" condition ($d = 0.13$). When taking the label-type within each group into account, there was a large WSW-effect for the experimental label (i.e., "schizophrenia") in the "schizophrenia" vs. "no label" condition ($d = 1.01$), while all other effects were small or in the opposite direction ($-0.36 < d < 0.28$; see Table 1).

2.3. Discussion

The results of this pilot study point toward an implicit categorization when people with and without mental illness labels are presented. No consistent results were found when people with different mental illness labels were juxtaposed. Thus, it seems that lay concepts for different diagnoses are not as elaborate as the general concept of a social group of "people with mental illness": In the "schizophrenia" vs. "depression"-condition, all targets were perceived as belonging to this one social group "people with mental illness". Consequently, the "schizophrenia" vs. "no label"-condition, in which the assumed lay social category "mental illness" is juxtaposed with a complementary group "healthy", seems to be the preferable setup for a mental illness WSW-task.

Surprisingly, within-errors were distributed unevenly between the two label groups. Within-group errors occurred more frequently among

Table 1
Average within- and corrected between-errors by condition and label type.

Label-group	Study 1 Within-errors (M; SD)	Betw.-errors (M; SD)	d	Study 2 Within-errors (M; SD)	Betw.-errors (M; SD)	d	Study 3 Within-errors (M; SD)	Betw.-errors (M; SD)	d
Schizophrenia vs. No label condition									
Schizophrenia	4.00; 1.48	2.59; 1.31	1.01	3.98; 1.63	2.91; 1.40	0.71	3.98; 1.63	2.91; 1.41	0.55
No diagnosis	3.45; 1.29	3.14; 0.94	0.28	3.05; 1.56	3.85; 1.36	−0.55	2.05; 1.56	3.85; 1.36	−0.41
Schizophrenia vs. Depression condition									
Schizophrenia	3.38; 1.98	3.17; 1.11	0.13	3.26; 1.73	3.84; 1.36	−0.36			
Depression	3.08; 1.55	3.58; 1.23	−0.36	3.67; 1.57	3.18; 1.44	0.33			
Depression vs. No diagnosis condition									
Depression							5.45; 2.22	4.62; 1.81	0.41
No diagnosis							3.97; 1.90	5.09; 1.64	−0.63

targets with a schizophrenia label than among targets with no label. Possibly, this is a behavioral manifestation of an essentialized category "schizophrenia" or "mentally ill" with no elaborated lay concept of a mutually social group of "normal" or "sane". In other words, people categorize others into the group "mentally ill" as a mean to organize information, whereas the social category 'healthy'/'sane' in itself appears to serve no informative value beyond a norm or 'default' mental representation of other people. Consequently, we would need to focus on the WSW-effect pertaining to the "schizophrenia" targets only if the aim is to examine spontaneous categorization based on mental illness labels.

Finally, based on the results of this pilot study, we determined the sample size for a replication study to be $N = 62$ in order to detect the condition x error-type and label type x error-type interaction with a test-power of 80% given and effect sizes of $\eta^2_{partial} = 0.084$ and $N = 102$ to increase test-power to 95%.

3. Study 2 – Test of label-dependent WSW-effects

In the second study, we aimed to replicate the results of study 1. Specifically, we hypothesized that the WSW-effect is (1) most pronounced in a "schizophrenia" vs. "no label" juxtaposition and (2) when the WSW-effect is calculated based on the erroneous allocation of statements originally shown by people with a "schizophrenia" label.

3.1. Methods

3.1.1. Participants

Based on the results of study 1, a sample between 60 and 110 participants was aimed for. Advertisement on various social networks yielded a new sample of 85 participants. The age of the participants ranged from 21 to 48 years, with a mean age of 25.58 years ($SD = 7.59$). About two third (62.4%) of the sample were female (23.5% male, and 14.1% other /no answer). About a quarter of the sample indicated their education level to be the German university entrance certificate (Abitur; 24.7%) or a Bachelor's degree (23.5%), respectively. Further 9.4% reported to hold their Master's degree and 12.9% did not report their education level. A fifth of the sample (20%) reported to have had previous contact with people with schizophrenia.

3.1.2. Procedure, material, and data-analysis

The procedure, material, and data-analyses were identical to study 1.

3.2. Results

An ANOVA of errors with experimental condition as between subject factor and label-type as well as error-type as within subject factors yielded a significant interaction of condition x label-type x error-type ($F(1,83) = 17.49$, $p < 0.001$, $\eta^2_{partial} = 0.174$). Overall, there was a very small, non-significant WSW-effect in the "schizophrenia" vs. "no label"-condition ($t(42) = 0.47$, $p = 0.638$, $d = 0.12$), while there was no WSW-effect in the "schizophrenia" vs. "depression" condition ($t(41) = -0.13$, $p = 0.893$, $d = -0.04$). Consistent with the results of study 1, a significant WSW-effect was found when only the within and between-errors in the label-type "schizophrenia" in the "schizophrenia" vs. "no label" condition ($t(42) = 2.77$, $p = 0.008$, $d = 0.71$) were analyzed, whereas there was a reverse effect with significantly more between than within-errors for the label-type "no diagnostic label" ($t(42) = -2.16$, $p = 0.037$, $d = -0.55$, see Table 1). No significant differences of within- and between-errors were found in the "schizophrenia" vs. "depression"-condition.

3.3. Discussion

In the second study, we replicated an out-group-specific WSW-effect in the "schizophrenia" vs. "no-label/healthy" condition. As expected, this effect was not observed in a WSW-task juxtaposing people with schizophrenia and people with depression. Thus, it appears that a WSW-paradigm for mental illness categorization works best with a healthy comparison group.

However, some alternative interpretations of the findings need to be considered. In the present version of the task, one group received a label ("schizophrenia"), whereas no label was presented for the other group (e.g., the other group was not explicitly labeled as "healthy" or "sane"). Thus, it is possible that participants memorized the statements in combination with the additional information "schizophrenia", leading to the within-errors in the schizophrenia group, while no such memorization cue was present for the comparator group. In this case, what we have interpreted as a WSW-effect in the "schizophrenia" vs. "healthy" group comparison would have resulted from a memorization technique rather than from mental categorization related to group perception. If this were the case, the memorization cue "schizophrenia" could be replaced by any other piece of information irrelevant to social groups or stigmatization (e.g., grouping four of the targets as "owning a pet" and providing no explicit descriptor for the other four).

However, if we follow this memorization-technique interpretation, the fact that no WSW-effect was observed in the "schizophrenia" vs. "depression"-condition becomes difficult to explain: If memorization was the driving force behind the effect, being presented with memory cues for both groups should yield similar results, unless the two written labels already divide the participants' attention enough to impede the effective use of this supposed memorization-technique. Thus, use of random memorization-techniques is a less likely interpretation than categorization based on salient social groups (or as the case may be, one social group "schizophrenia" vs. no salient social group) – yet both interpretations remain possible and need to be tested in further variations of the WSW-task.

4. Study 3 – Variation in experimental task characteristics and association with stigmatizing attitudes

In the third and final study, we aimed to further test for the aforementioned alternative explanations for the pattern of WSW-effects we found. We added an explicit label for the "no mental illness"-comparison targets, so that the "schizophrenia" vs. "no label/healthy" condition would equal the "schizophrenia" vs. "depression" condition of studies 1 and 2 in that all targets have a label to sort information by.

Furthermore, we wanted to test whether the "schizophrenia" vs "no label/healthy" WSW-paradigm can be altered to test for categorization based on other mental illness labels. Consequently, the experimental conditions for this study were (1) "schizophrenia" vs "healthy" and (2) "depression" vs. "healthy", in order to test whether the target-group WSW-effect emerges irrespective of a specific diagnosis.

Finally, we used a completely new set of target pictures and statements and added a brief rating of each target person to the end of the experiment in order to test for any association of the WSW-effect and the self-reported tendency to categorize people with mental illness as well as self-reported prejudice.

4.1. Methods

4.1.1. Participants

We recruited 137 participants. Of these, 45 participants were psychology students at the Universität Hamburg and 92 participants were recruited via social media advertisement. The mean age of the sample was 27.8 years (SD = 7.6) and 71.53% were female. Half of the participants (56.9%) indicated their education level to be the German university entrance certificate (Abitur), 29.9% held a university degree 9.5% completed vocational/ job training. Half of the sample (46%) reported to have had previous contact with people with schizophrenia.

4.1.2. Procedure and materials

At the beginning of the EFS-Survey based online-study, which was introduced as a study on first impressions about people (in order to justify later rating of the targets), participants provided informed consent and started the WSW-task.

The two experimental conditions included a variation of the previously used "schizophrenia" vs. "healthy" juxtaposition, in which one group consisted of four target persons with the label "diagnosed with schizophrenia" (experimental label targets 1) and four target persons with the label "healthy" (comparison label targets). In the second condition, one group consisted of four target persons with the label "diagnosed with depression" (experimental label targets 2) and four target persons with the label "healthy".

In both conditions, participants were first shown 32 statements by eight different male target persons (i.e., four statements per person). A new set of pictures from the Caucasian subset of the Radboud faces repository (Langner et al., 2010) was used as visual representation of the 8 male target persons. We selected four pairs of front-view pictures depicting a person with a neutral facial expression. Within each pair, the pictures were matched for hair-color as well as valence of the picture and genuineness of the expression based on existing validation data (Langner et al., 2010). Twenty-four positive (e.g., "keeping good relations to my friends is important to me") and 24 negative self-descriptory statements (e.g., "there are few things I'm really interested in") were generated for this study. A test-sample (n = 5) rated for each of the 48 statements how they would perceive a person after hearing only this statement on three six-point semantic differential scales (i.e., "positive" – "negative", "competent" – "incompetent", and "likeable" – "unlikeable"). The 16 most consistently positively and negatively rated items were extracted for the final list of statements. Each target person was paired with two positive and two negative statements. All other parameters of the WSW-task and its procedure were similar to studies 1 and 2.

Following the WSW-task, participants were shown the eight pictures of the target persons in randomized order for a final time. As a self-report approximation of categorization tendencies, we adapted a brief assessment of continuum beliefs (Corrigan et al., 2015; Schomerus et al., 2012) to the task. Participants rated how much they perceived each target person to be similar to themselves (two items per target, e.g., "my first impression is that this person is similar to me"; mental illness label targets: α = 0.82, healthy label targets: α = 0.82) and similar to all other people (two items per target, e.g., "this person basically like everyone else", mental illness label targets: α = 0.83, healthy label targets: α = 0.84).

Also, in order to test for an association of WSW-effects and prejudice, participants were asked to „describe [their] initial impression of this person based on the statements that person has made", using three seven-point semantic differentials ("likeable-unlikeable", "pleasant-unpleasant", "friendly-unfriendly") for each target person. Internal consistencies of the mean likeability-rating (4 × 3 items per target group) were α = 0.86 in the mental illness label groups and α = 0.85 for the "healthy" label group.

4.1.3. Data-analysis

In a first step, the overall WSW-effects were tested using mixed model ANOVAs as in study 1 and 2. We tested for label-type x error-type interaction effects to see whether we could replicate the WSW-effect for the mental illness target groups. Furthermore, we tested for experimental condition x label-type x error-type interaction effects to explore whether the changes we made to the labels (i.e., explicitly labeling the comparison group as well) changed the pattern of WSW-effects.

Next, the individual target group WSW-scores for (1) the mental illness label targets and (2) the healthy label targets were calculated by subtracting a participant's corrected between-group errors from the within group errors. Then, the WSW-scores the mental illness label targets were entered into three multiple regression models that predicted the average similarity of mental illness targets to themselves and to everyone else (i.e., a test of convergent validity between the WSW-effect and self-reported categorization), and the averaged likeability of the mental illness targets (as a test for the association between categorization and prejudice/stigmatization). If any of these models yielded a significant effect involving the WSW-score, a corresponding regression model was calculated with the WSW-score pertaining to healthy targets to test whether the effect is specific to the pre-established outcome.

In all regression models, the respective WSW-score, experimental condition and the interaction of WSW-score and condition were entered as independent variable, and age, gender, being a psychology student (yes/no) as well as the respective averaged rating (of similarity/likeability) of the healthy targets were entered as covariates.

4.2. Results

4.2.1. WSW effect

An ANOVA of errors with experimental condition (1: "Schizophrenia"; 2: "Depression") as between subject factor and label-type (1: Mental illness label; 2: Healthy) as well as error-type as within subject factors yielded no significant main or interaction effect with experimental condition. There was, however, a label-type x error-type interaction ($F(1135)$ = 30.54, p < 0.001, $\eta^2_{partial}$ = 0.184). As can be seen in Fig. 2, participants in both conditions showed a WSW-effect pertaining to stimulus group with mental illness labels, but there was no WSW-effect for the healthy group.

4.2.2. Association of WSW-effect and self-reported stigma tendency

WSW-scores neither predicted the perceived similarity of target persons with mental illness label to others (b = −0.04, T = −1.63, p = 0.107) nor the perceived similarity of target persons with mental

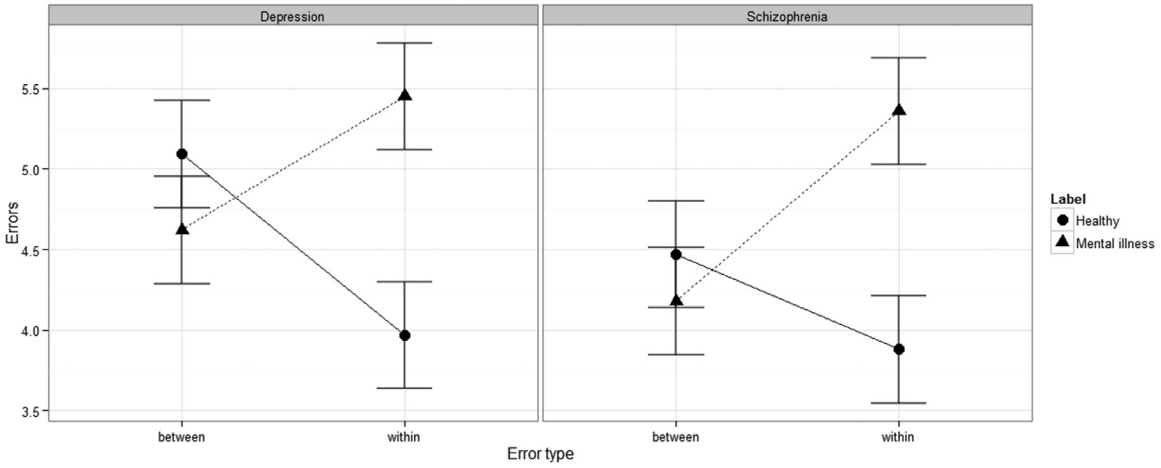

Fig. 2. Mean and 95% confidence interval of errors by error-type, label type, and experimental group (study 3).

illness label to oneself ($b = -0.03$, $T = -1.01$, $p = 0.314$).

However, the WSW-score did significantly predicted the average self-reported likeability of target persons with a mental illness label (see Table 2). Higher WSW-scores were predictive of negative evaluations of people with mental illness ($b = 0.06$, $T = 2.13$, $p = 0.035$) This association was not found in a similar regression analysis with the WSW-score based on the healthy targets replacing the mental illness label target WSW-score ($b = 0.02$, $T = 0.35$, $p = 0.525$).

5. General discussion

In this study, we tested whether mental categorization in social groups based on mental health status can be assessed using the "Who-Said-What"-paradigm (Taylor et al., 1978). A medium to large WSW-effect for people with mental illness labels (schizophrenia, depression) vs. healthy people was found and replicated in three studies. Alternative explanations such as type of stimuli used or the way the comparison group was labeled could be ruled out: The effect was present whether or not the comparison group received the label "healthy". However, irrespective of whether the comparison group was labeled or not, the within-between error difference indicative of spontaneous categorization was only observed for the target groups with a mental illness label. Previous research that tested the WSW-effect based on minimal groups (i.e., group allocation based on the color of the targets' shirts) found the magnitude of the effect to vary based on the perceived meaningfulness of groups (i.e., the effect was more pronounced when an extended explanation was given that the colors represented opposite groups of people who tend to over- or underestimation in perceptual estimation tasks vs. no explanation for the colors; Brewer et al., 1995). Moreover, if there is meaningfulness attributed to group status, the perceived majority/minority status of the depicted groups in society and the salience of a participant's own membership to either group can lead to "one-sided" modulation of the WSW-effect (Brewer et al., 1995). We thus cannot exclude the possibility that any combination of these factors could be responsible for the group-specific effect we found in our studies. For example, "healthy" constitutes a less meaningful social group category that is considered broader, more diffuse, and less rich in information than a specific diagnostic label, such as "schizophrenia". This could facilitate a "one-sided" social categorization. Future research is needed that experimentally manipulates the aforementioned factors (e.g., by providing mock-information on the prevalence of diagnoses that changes participants' perception of the minority status of the mental illness and healthy category, respectively). Nevertheless, the effect we found arguably corresponds to how spontaneous categorization based on a mental illness label emerges under natural circumstances. Possibly, this one-sided outgroup categorization even amplifies subsequent stigmatization and discriminating behavior.

Of importance, the failure to find a WSW-effect when two mental illness groups were presented together may point towards the tendencies of the lay population in the way they categorize mental disorders: At least as far as implicit-behavioral parameters are concerned, the general population may tend to broadly categorize people into people with mental illness and people without mental illness. The automatic use of specific diagnostic categories is far less likely. Possibly, the automated concept of mental illness is not that regularly used in everyday-live resulting in more global categories. Future studies could verify this theory with external criteria by testing the link between the WSW-scores that result from paradigms with different comparison groups and other stigma related variables.

Interestingly, we found no association between a person's WSW-score and self-reported indicators of categorization (i.e., a brief assessment of continuum beliefs, Corrigan et al., 2015; Schomerus et al., 2012), and only a small association between WSW-scores and self-reported prejudice towards people with mental illness. This finding is in line with a repeatedly found divergence between self-reported and implicitly assessed attitudes, which is often a result of motivation to respond without prejudice in self-report assessments (Hofmann et al., 2005). It seems reasonable to assume that this divergence extends to mental illness stigmatization. Nevertheless, future research needs to test the association of implicit and explicit categorization and implicit as well as explicit prejudice, including measures of implicit prejudice to compare to WSW-scores (e.g., an affective priming paradigm; Degner and Wentura, 2010).

Table 2
. Regression of the mean positive-negative semantic differential evaluation of targets with a mental illness label.

Predictor	Beta	Estimate	SE	*T*	*p*
Demographic and control variables					
Age	0.06	0.01	0.01	0.89	0.374
female gender (no vs. yes)	−0.15	−0.27	0.13	−2.02	0.046
Psychology student (no vs. yes)	0.01	0.09	0.12	0.77	0.441
Evaluation of healthy target persons	0.57	0.53	0.07	7.35	<0.001
R^2		0.32			
Variables of interest					
Mental illness label condition (depression vs. schizophrenia)	0.09	0.15	0.13	1.19	0.238
WSW-score(mental Illness targets)	0.22	0.06	0.03	2.13	0.035
Mental illness label condition × WSW-score	−0.05	−0.07	0.04	1.97	0.052
R^2		0.35			
ΔR^2		0.03			

Mapping diverging paths from implicit categorization to implicit stigmatization and from explicit categorization to explicit stigmatization could provide a deeper understanding of mental illness stigma that could ultimately inform anti-stigma interventions. For example, a recent study (Thibodeau et al., 2018) found that experimentally manipulating participants to endorse continuum or category beliefs by means of an educational text affects self-reported attitudes and self-reported behavioral intentions to distance oneself, but does not change the actual behavior (i.e., the distance in the seating position to a person with a mental illness). Correspondingly, some findings in social psychology indicate spontaneous behavior to be more closely linked to implicit rather than explicit attitudes (e.g., McConnell and Leibold, 2001). Finding valid and reliable indicators of implicit aspects of mental illness stigma and implementing them in the evaluation of future intervention studies may thus provide us with a more balanced view of the effects of interventions and prevent overly optimistic conclusions based on self-report measures.

5.1. Strengths and limitations

The replication of our results in different samples and with different stimulus materials can be considered a strength of our study. It needs to be noted though that all studies used ad-hoc samples with above average education status (more than half of all samples with college entrance degrees or higher levels of education) and a majority of female participants. This is in line with previous studies using social media based recruitment (e.g., Topolovec-Vranic and Natarajan, 2016). Nevertheless, studies on more representative samples would be needed to guarantee the generalizability of the WSW-effect found in this study. Moreover, the WSW-task employed could be further refined to simultaneously increase the cognitive load of the task, incite participants to engage in the task, and include a more ecologically valid setting. For example, the discussion phase could be filmed as an actual video discussion, which would allow for more information to be included. Finally, it has to be noted that all studies were online-studies, so we were unable to filter out participants that did not pay attention to the WSW-discussion phase, resulting in an answer pattern unrelated to any categorization effect and overall underestimated WSW-effect sizes. Nevertheless, we found consistent results regarding the WSW-effect across all online-studies, which speaks for the WSW-task as a cost-effective and easily implemented test of implicit categorization.

5.2. Conclusion

Recent research on mental illness stigma has seen a shift towards assessing a perceived differentness of people with mental illness as a crucial component and starting point of stigmatization (Corrigan et al., 2015). We hope to have provided this strand of research with an elaborate and cost-effective way to assess this “perceived differentness” in an indirect manner to complement self-reported categorization or continuum beliefs. We found promising results with respect to the stability of the presented assessment paradigm. With this, we hope to lay the ground-stone for a more comprehensive understanding of the perceived difference between people with and without a diagnosis of a mental illness that will ultimately contribute to reduce the stigma of mental illness.

Conflict of interest

None.

Acknowledgments

The authors would like to thank Rachel Theiß and Katrin Scheffler for their help with programming the first variation of the “Who-Said-What”-paradigm, Vivien Breckwoldt for her assistance with participant recruitment for study 2 and Celina Fiß for her help in revising the “Who-Said-What”-paradigm and with participant recruitment for study 3.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.psychres.2019.03.028.

References

Bigler, R.S., Liben, L.S., 2006. A developmental intergroup theory of social stereotypes and prejudice. Adv. Child Dev. Behav. 34, 39–89.
Bigler, R.S., Liben, L.S., 2007. Developmental intergroup theory: explaining and reducing children's social stereotyping and prejudice. Curr. Dir. Psychol. Sci. 16, 162–166.
Brewer, M.B., Weber, J.G., Carini, B., 1995. Person memory in intergroup contexts: categorization versus individuation. J. Pers. Soc. Psychol. 69, 29–40.
Corrigan, P.W., Bink, A.B., Fokuo, J.K., Schmidt, A., 2015. The public stigma of mental illness means a difference between you and me. Psychiatry Res. 226, 186–191.
Corrigan, P.W., Schmidt, A., Bink, A.B., Nieweglowski, K., Al-Khouja, M.A., Qin, S., Discont, S., 2017. Changing public stigma with continuum beliefs. J. Ment. Health 26, 411–418.
Degner, J., Wentura, D., 2010. Automatic prejudice in childhood and early adolescence. J. Pers. Soc. Psychol. 98, 356–374.
Fazio, R.H., Olson, M.A., 2003. Implicit measures in social cognition research: their meaning and use. Annu. Rev. Psychol. 54, 297–327.
Greenwald, A.G., Poehlman, T.A., Uhlmann, E.L., Banaji, M.R., 2009. Understanding and using the implicit association test: III. Meta-analysis of predictive validity. J. Pers. Soc. Psychol. 97, 17–41.
Haslam, N., 2011. Genetic essentialism, neuroessentialism, and stigma: commentary on Dar-Nimrod and Heine (2011). Psychol. Bull. 137, 819–824.
Haslam, N., Rothschild, L., Ernst, D., 2002. Are essentialist beliefs associated with prejudice? Br. J. Soc. Psychol. 41, 87–100.
Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., Schmitt, M., 2005. A meta-analysis on the correlation between the implicit association test and explicit self-report measures. Pers. Soc. Psychol. Bull. 31, 1369–1385.
Klauer, K.C., Wegener, I., 1998. Unraveling social categorization in the “Who said what?” paradigm. J. Pers. Soc. Psychol. 75, 1155–1178.
Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D.H.J., Hawk, S.T., van Knippenberg, A., 2010. Presentation and validation of the Radboud Faces Database. Cogn. Emot. 24, 1377–1388.
Link, B.G., Phelan, J.C., 2001. Conceptualizing stigma. Annu. Rev. Sociol. 27, 363–385.
McConnell, A.R., Leibold, J.M., 2001. Relations among the implicit association test, discriminatory behavior, and explicit measures of racial attitudes. J. Exp. Soc. Psychol. 37, 435–442.
QuestBack GmbH, 2014. Questback EFS.
Schlier, B., Scheunemann, J., Lincoln, T.M., 2016. Continuum beliefs about psychotic symptoms are a valid, unidimensional construct: construction and validation of a revised continuum beliefs questionnaire. Psychiatry Res. 241, 147–153.
Schomerus, G., Angermeyer, M.C., Baumeister, S.E., Stolzenburg, S., Link, B.G., 2016. An online intervention using information on the mental health-mental illness continuum to reduce stigma. Eur. Psychiatry 32, 21–27.
Schomerus, G., Matschinger, H., Angermeyer, M.C., 2013. Continuum beliefs and stigmatizing attitudes towards persons with schizophrenia, depression and alcohol dependence. Psychiatry Res. 209, 665–669.
Schomerus, G., Schwahn, C., Holzinger, A., Corrigan, P.W., Grabe, H.J., Carta, M.G., Angermeyer, M.C., 2012. Evolution of public attitudes about mental illness: a systematic review and meta-analysis. Acta Psychiatr. Scand. 125, 440–452.
Tajfel, H., 1981. Human Groups and Social Categories: Studies in Social Psychology. Cambridge University Press, Cambridge.
Tajfel, H., Turner, J.C., 1986. The social identity theory of intergroup behaviour. Psychology of Intergroup Relations. Nelson, Chicago, pp. 7–24.
Taylor, S.E., Fiske, S.T., Etcoff, N.L., Ruderman, A.J., 1978. Categorical and contextual bases of person memory and stereotyping. J. Pers. Soc. Psychol. 36, 778–793.
Thibodeau, R., 2017. Continuum beliefs and schizophrenia stigma: correlational and experimental evidence. Stigma Health 2, 266–270.
Thibodeau, R., Peterson, K.M., 2018. On continuum beliefs and psychiatric stigma: similarity to a person with schizophrenia can feel too close for comfort. Psychiatry Res. 270, 731–737.
Thibodeau, R., Shanks, L.N., Smith, B.P., 2018. Do continuum beliefs reduce schizophrenia stigma? Effects of a laboratory intervention on behavioral and self-reported stigma. J. Behav. Ther. Exp. Psychiatry 58, 29–35.
Topolovec-Vranic, J., Natarajan, K., 2016. The use of social media in recruitment for medical research studies: A scoping review. J. Med. Internet Res. 18, 1–13.
Turner, J.C., Hogg, M.A., Oakes, P.J., Reicher, S.D., Wetherell, M.S., 1987. Rediscovering the Social group: A self-Categorization Theory. Blackwell, Oxford.
Wegener, I., 2000. [Social Categories in the Situative Context: Cognitive Flexibility in Person Perception]. Universität Bonn.
Wiesjahn, M., Brabban, A., Jung, E., Gebauer, U.B., Lincoln, T.M., 2014. Are continuum beliefs about psychotic symptoms associated with stereotypes about schizophrenia? Psychosis 6, 50–60.
Wiesjahn, M., Jung, E., Kremser, J.D., Rief, W., Lincoln, T.M., 2016. The potential of continuum versus biogenetic beliefs in reducing stigmatization against persons with schizophrenia: an experimental study. J. Behav. Ther. Exp. Psychiatry 50, 231–237.